

JGR Solid Earth



RESEARCH ARTICLE

10.1029/2022JB025952

Key Points:

- Relative associations of geothermal wells with seismicity rate are evaluated using regression analysis and time series feature engineering
- Reservoir operational changes such as injection transfers between wells can affect the association between wells and microseismicity
- There is a stronger correlation between injection and long-term seismicity rate than short-term seismicity rate at Rotokawa and Hellisheiði site

Supporting Information:

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

Correspondence to:

D. Dempsey,
david.dempsey@canterbury.ac.nz

Citation:

Yu, P., Dempsey, D., Rinaldi, A. P., Calibugan, A., Ritz, V. A., & Archer, R. (2023). Association between injection and microseismicity in geothermal fields with multiple wells: Data-driven modeling of Rotokawa, New Zealand, and Húsmúli, Iceland. *Journal of Geophysical Research: Solid Earth*, 128, e2022JB025952. <https://doi.org/10.1029/2022JB025952>

Received 31 OCT 2022
Accepted 11 MAR 2023

Author Contributions:

Conceptualization: Pengliang Yu, David Dempsey
Data curation: Antonio Pio Rinaldi, Aimee Calibugan, Vanille A. Ritz
Formal analysis: Pengliang Yu
Funding acquisition: Rosalind Archer
Investigation: Pengliang Yu, David Dempsey, Vanille A. Ritz
Methodology: Pengliang Yu, David Dempsey

© 2023. 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.

Association Between Injection and Microseismicity in Geothermal Fields With Multiple Wells: Data-Driven Modeling of Rotokawa, New Zealand, and Húsmúli, Iceland

Pengliang Yu^{1,2,3} , David Dempsey⁴ , Antonio Pio Rinaldi⁵ , Aimee Calibugan⁶, Vanille A. Ritz⁵ , and Rosalind Archer⁷

¹Department of Engineering Science, University of Auckland, Auckland, New Zealand, ²Geothermal Institute, University of Auckland, Auckland, New Zealand, ³Now at Department of Energy and Mineral Engineering, EMS Energy Institute, The Pennsylvania State University, University of Park, PA, USA, ⁴Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand, ⁵Swiss Seismological Service, Swiss Federal Institute of Technology, ETHZ, Zurich, Switzerland, ⁶Mercury NZ, Rotorua, New Zealand, ⁷School of Engineering and Built Environment, Griffith University, Nathan, QLD, Australia

Abstract Understanding injection-induced microseismicity in geothermal systems can provide insight into reservoir connectedness. However, fault and reservoir complexity are difficult to represent in simple analytical models, which makes it difficult to discern clear relationships from incidental associations. Here, we have used data-driven models to study how fluid injection and microseismicity are related in the Rotokawa (New Zealand) and Hellisheiði (Iceland) geothermal fields. We tested two classes of model: (a) lagged linear regression of seismicity rate as a function of well injection rates; and (b) systematic extraction of injection time series features that are then evaluated for associations with the seismicity. These models allowed us to determine which wells had the greatest correlation with microseismicity and to explain this association in a reservoir context. Finally, exploring different data types and transformations, we were unable to establish a link between rapid changes in injection rate and seismicity spikes, as suggested by some theoretical models.

Plain Language Summary To better understand the small earthquakes triggered in geothermal fields, we used two kinds of data-driven models: time series feature engineering (a machine learning method) and the non-negative linear regression method. We applied these techniques to microearthquakes recorded in the Rotokawa geothermal field, New Zealand, and the Húsmúli reinjection area in the Hellisheiði field, Iceland. We searched for associations and patterns hidden within the noisy injection and seismicity data. These patterns are further complicated by the complex injection schedules that involve several wells that switch off and on over time. We found a stronger association between injection and long-term microseismicity rate as opposed to short-term fluctuations driven by noise. We showed that particular injection wells are strongly associated with the microseismicity, which we interpret in terms of reservoir connectedness and stress changes in the rock. We found that injection rate and volume are the most useful data for understanding microseismicity, whereas there was no evidence that rapid injection rate changes triggered seismicity at these two locations.

1. Introduction

Microseismicity is a common occurrence in geothermal fields, arising from natural and anthropogenic stresses to the hydrothermal system. At the Wairakei and Rotokawa fields in New Zealand, microseismicity clusters at shallow depths (<3 km) around wells that reinject condensate and brine back into the system (Hopp et al., 2020; McNamara et al., 2016; Sepulveda et al., 2014). It also occurs in some production areas (Sepulveda et al., 2014) and can be associated with the deep fluid upflow (Sherburn, Bannister, et al., 2015; Sherburn, Sewell, et al., 2015).

In Iceland's Hellisheiði geothermal area, injection-induced microseismicity was first associated with drilling and well testing operations at the Húsmúli reinjection site (Gunnarsson, 2013). Since commissioning of the Húsmúli reinjection area in September 2011, elevated seismicity has been associated with varying injection temperature as well as injection rates of water and non-condensable gases (CO₂ and H₂S). The CO₂ and H₂S are dissolved into pressurized water and then the gas-charged water is reinjected into subsurface basalts for permanent storage (Matter et al., 2016). The seismicity is assumed to be controlled by fluid flow through complex structural features (Gunnarsson, 2013; Mahzari et al., 2021; Ratouis et al., 2022). Other well-studied examples of geothermal

Resources: David Dempsey, Antonio Pio Rinaldi, Aimee Calibugan, Rosalind Archer
Software: Pengliang Yu, David Dempsey
Supervision: David Dempsey, Antonio Pio Rinaldi, Rosalind Archer
Writing – original draft: Pengliang Yu
Writing – review & editing: David Dempsey, Antonio Pio Rinaldi, Aimee Calibugan, Vanille A. Ritz, Rosalind Archer

microseismicity include the Coso and Geysers geothermal fields in California. At Coso, both natural and induced seismicity has been distinguished on the basis of space-time-magnitude characteristics (Schoenball et al., 2015) with induced events showing greater clustering in space and time. At the Geysers, associations are noted between fluctuating injection and seismicity rates (Majer & Peterson, 2007), although quantifying the relationship has proven challenging (Kwiitek et al., 2015).

Because of their small magnitude (often $M_w < 2$), sensitive seismic networks are usually required to detect and locate microearthquakes (Rintamäki et al., 2021; Warpinski, 2009). These events provide a rich source of information about subsurface structure and fluids. For example, the relative proportion of small to large magnitude events, parameterized as the Gutenberg-Richter b -value, has been linked to fluid pressure distributions (Bachmann et al., 2012; Dempsey & Suckale, 2016; Hopp et al., 2020; Ritz et al., 2022). The locations and rates of events have been used to delineate fluid conductive structures (Sherburn, Bannister, et al., 2015; Sherburn, Sewell, et al., 2015), pore pressure distributions (Shapiro et al., 2005) and permeability structure (Riffault et al., 2018).

In hydrocarbon extraction industries, microseismicity has been used to track engineered fracture propagation in near real-time (Chen et al., 2018; Cipolla et al., 2011; Sigmundsson et al., 2014). Adapting such techniques for geothermal settings could unlock design of fracture systems with optimal heat extraction properties (e.g., Jung, 2013; Yu et al., 2021). However, a persistent challenge when extracting information from microseismic catalogs is the inherent stochasticity of the earthquake process. This apparent randomness contributes noise that may obscure signals of interest. With this limitation, it can be difficult to access the full potential of microseismicity data.

In addition to microseismicity, larger-magnitude earthquakes can also occur in geothermal fields (Majer & Peterson, 2007). However, it can be difficult to attribute individual events to natural or anthropogenic causes because most geothermal fields are located in tectonically active regions with high background seismicity. Felt seismicity is sometimes associated with well stimulation operations to create Enhanced geothermal systems (EGS; Deichmann & Giardini, 2009; Ellsworth et al., 2019). Moderate magnitude events have also occurred in geothermal reservoirs, including numerous events with $M > 4$ in the Geysers since 2003 (Martínez-Garzón et al., 2014) and two $M > 4$ at Húsúli in October 2011.

In this study, we have focused on small-magnitude microseismicity and what it can tell us about the geothermal reservoir. Our approach has been to use data-driven models to extract broad patterns from noisy data, and to demonstrate these using case studies in New Zealand (Rotokawa) and Iceland (Húsúli).

1.1. Physics-Based Models of Induced Seismicity

Fluid injection that causes pressure build-up on critically stressed faults is a widely recognized triggering mechanism for induced earthquakes (Cardiff et al., 2018; Ellsworth, 2013; Healy et al., 1968; Hubbert King & Rubey, 1959; Raleigh et al., 1976). Other mechanisms have also been suggested, including poroelastic stressing that mediates long-range stress changes (Deng et al., 2020; Mossop & Segall, 1999) or short-term rate fluctuations (Segall & Lu, 2015), and thermoelastic stresses related to cooling around injection wells (Candela et al., 2018; Sherburn, Bannister, et al., 2015; Sherburn, Sewell, et al., 2015).

These theoretical mechanisms have been used to underpin physics-based models that link injection and seismicity. For example, Segall and Lu (2015) used a model of the poroelastic stress surge that follows a rapid drop of injection rate to propose that well shut-in could trigger seismicity. Catalli et al. (2016) modeled the static stress transfers from induced earthquakes and showed that cascading events could be triggered outside the fluid pressurized region. Riffault et al. (2019, 2018) used pressurization-rate models to invert for the evolving permeability distributions around EGS projects in Australia. Various approaches have been taken to address stochastic randomness, including seeding random earthquake sources (Rinaldi & Nespoli, 2017), modeling stochastic fracture networks (Norbeck et al., 2018), and introducing fractal stress distributions (Dempsey & Suckale, 2016).

Despite progress in understanding the mechanisms of induced seismicity, causative modeling is still challenging on a case-by-case basis. Operational parameters like injection rate, wellhead pressure, total injection volume, well location, and geological and geomechanical characteristics are known to influence induced seismicity rates (He et al., 2020; Hofmann et al., 2018). However, extracting relationships is difficult when data sets are large and noisy. This is often not a shortcoming of the monitoring systems, but rather an aggregate outcome of

uncharacterized heterogeneity in the system. For instance, earthquake generating fractures are not uniformly distributed or oriented, and nor are their resolved stress conditions homogeneous. This makes it hard to separate out the randomness and to extract useful information using conventional analytical or numerical methods (He et al., 2020).

1.2. Data-Driven Models of Injection-Induced Seismicity

Empirical models that are motivated, parameterized, and calibrated with respect to data are an alternative tool for interpreting geophysical observations. Within this paradigm, models that link injection and seismicity are not justified on the basis of hypothesized physics, but rather by their goodness-of-fit to the data. Thus, simpler classes of models can be used, such as regression-fitted polynomials, which are often more computationally efficient. For example, Brodsky and Lajoie (2013) found evidence for a linear relationship between injection, net production and seismicity rate during some periods of operation at the Salton Sea geothermal system, California. A cross-correlation analysis revealed no credible lags between operation and seismicity longer than 1 month. Taking a similar approach, Trugman et al. (2016) observed a strong correlation between declustered seismicity and injection and production volumes at the Geysers geothermal system, although not at Coso and Salton Sea fields.

Many models of induced seismicity use injection or production rate as a proxy variable to drive seismicity rate (e.g., Langenbruch et al., 2018; Shapiro et al., 2013; Zöller & Holschneider, 2016). Such models implicitly link to pressure and stress changes that trigger the seismicity, although they may have difficulties representing nonlinear relationships such as threshold triggering pressures or Kaiser effect (Dempsey & Riffault, 2019). Nevertheless, these models are powerful from an operational forecasting perspective because they are computationally inexpensive and have few tunable parameters. For example, Broccardo et al. (2017) have demonstrated progressive tuning of a nonhomogeneous Poisson process model for fluid-induced seismicity conditioned on prescribed fluid injection rate and a provision for well shut-in. This formulation had impressive performance on several induced seismicity case studies and shows how such systems could be used for real-time short-term forecasting and operational decision making.

Machine learning (ML) models also fall under the data-driven classification. Their application has proliferated in recent years, in part due to their ability to identify obscured patterns in data (Bergen et al., 2019; Kong et al., 2019). Recent relevant applications include identification and forecasting of laboratory (Lubbers et al., 2018; Rouet-Leduc et al., 2017) and field earthquakes (e.g., DeVries et al., 2018; Mignan & Broccardo, 2019; Rouet-Leduc et al., 2020), phase picking (Di Stefano et al., 2006; Mousavi et al., 2020; Ross et al., 2019), classification of remote sensing images (Bialas et al., 2016; Frank et al., 2017), and identifying changes in faulting processes (e.g., Corbi et al., 2020; Holtzman et al., 2018; Hulbert et al., 2020).

Several studies have used ML techniques to understand the interplay between geological and operational factors that control induced seismicity (Mehrabifard & Eberhardt, 2021). For example, Pawley et al. (2018) used binary classification to investigate the relative importance of geological features as a controlling factor on the seismic activation potential of induced seismicity in central Alberta. Wozniakowska and Eaton (2020a, 2020b) used a similar approach to show that tectonic setting, stratigraphy, distance to deformation front, and injection depth are all factors that influence induced seismicity susceptibility during hydraulic fracturing operations in the Montney Formation. Amini and Eberhardt (2021) applied several supervised ML algorithms to test whether geological or operational parameters had the larger effect on induced seismicity. Although they showed geological features to be more important, operational features continue to be of interest as these have the potential to be controlled or manipulated.

There have been a few studies that used ML methods to investigate correlations between induced seismicity and injection well data, mainly focused on moderate magnitude seismicity in Oklahoma. For example, Amirlatifi et al. (2021) applied a hybrid genetic-K-means algorithm to study the relationship between induced seismicity and injection wells in Oklahoma. They showed that cumulative injection volume and bottom-hole pressures were the most significant variables associated with induced seismicity increases. Larson et al. (2020) modeled earthquake magnitude using various classification and regression methods, and found that earthquake frequency correlated with wastewater disposal location, fault location, injection volumes, and injection pressure. Sinha et al. (2018) used the hierarchical and K-means clustering method to group wells into clusters that conform with

seismicity event density in Oklahoma. Amirlatifi et al. (2021) also demonstrated that increased seismicity rates are correlated with a specific cohort of wells based on an analysis and co-visualization of clustered data with earthquake maps.

However, the above-described studies did not evaluate the significance of individual wells within specific well groups. Doing so could help to identify problematic wells or develop conceptual models about relative connect- edness of wells to the larger flow system. This is a principal motivation for our study, which seeks to understand the relative sensitivity of different injection wells to the aggregate induced microseismicity rate.

The data-driven modeling studies described above are generally focused on the hazard associated with moder- ate magnitude induced seismicity. There has been comparatively less attention on the characterization of small-magnitude microseismicity, where the priority is to extract information useful for reservoir management. Furthermore, we are not aware of any studies that focus on geothermal settings. These can be more complex because the well field often comprises a complex portfolio of clustered production and injection wells, and flow rates may vary across hour-to month-long time scales in accordance with operational and economic constraints. Although physics-driven models have had good success over the years (O'Sullivan et al., 2001; Rutqvist et al., 2016b), there is an opportunity for data-driven models to identify latent relationships and highlight avenues for further investigation.

In this study, we have evaluated two classes of data-driven model for quantifying links between geothermal rein- jection and microearthquakes at Rotokawa geothermal field and the Húsmúli reinjection site at the Hellisheidi geothermal field. The first method uses linear regression to reconstruct seismicity rates using weighted combina- tions of lagged injection time series. The second method uses systematic time series feature engineering to search for latent patterns that are statistically significant for classifying periods of high and low seismicity. We also eval- uated the relative importance of different injection well operation data (injection rate, wellhead pressure, hydrau- lic work rate, injection volume). These models allowed us to rank the relative contribution of individual wells for generating microseismicity, and to contextualize this in terms of subsurface structure and reservoir connectivity.

1.3. Rotokawa Geothermal Field, New Zealand

Rotokawa geothermal field is one of two locations that we analyze in this article (Figure 1b). The field is one of more than 20 geothermal systems located in the central Taupō Volcanic Zone (TVZ) of New Zealand. The TVZ is a region of active rifting with extensive rhyolite and andesite volcanism, high rates of background seismicity, and widespread hydrothermal circulation (Bibby et al., 1995; Rowland et al., 2010). Natural seismicity in the TVZ usually occurs above depths of 6–8 km, which defines the depth of the brittle crust (Bryan et al., 1999). Much of the seismicity occurs in the Taupō Fault Belt, a central region of high-density faulting; however, swarm activity is noted elsewhere, often associated with calderas (Bannister et al., 2016; Illsley-Kemp et al., 2021; Webb et al., 2012).

Within the geothermal fields of the TVZ, natural and anthropogenic seismicity associated with geothermal operations has been well documented (Clarke et al., 2009; Hopp et al., 2019; Sepulveda et al., 2016; Sherburn et al., 2013; Sherburn, Bannister, et al., 2015; Sherburn, Sewell, et al., 2015). The hydrogeological response to reinjection at Rotokawa geothermal field is influenced by two NE-SW striking faults (Figure 1c): the Injection Field Fault (IFF) and the Central Field Fault (CFF; Wallis et al., 2013). Slow or non-existent tracer returns from injection to production regions indicate that the CFF acts as a buffer, slowing down the migration of injection fluids from the injection compartment to the production compartments (Sewell, Addison, et al., 2015). Addison et al. (2017), Sherburn, Bannister, et al. (2015), and Sherburn, Sewell, et al. (2015) have suggested that the CFF may act as a conduit for the injection fluid along the strike direction of the fault. However, it is still unclear what role the IFF plays in reservoir behavior (Hopp et al., 2020).

The Rotokawa geothermal field has sustained geothermal energy production for more than 20 yr (McNamara et al., 2016). Previous studies of microseismicity (Hopp et al., 2020; Sherburn et al., 2013) describe its localiza- tion in a region south of the Waikato River and near to a cluster of injection wells (Figure 1c). In 2010, reinjec- tion volumes into this area increased by about a factor of four. Around the same time, the microseismicity rate increased by a factor of 3.7, which suggests a relationship between the two (Sherburn et al., 2013).

Previous studies (Sewell, Cumming, et al., 2015; Sherburn, Bannister, et al., 2015; Sherburn, Sewell, et al., 2015) of the seismicity at Rotokawa were conducted in the period 2008–2012. These revealed that seismicity occurred

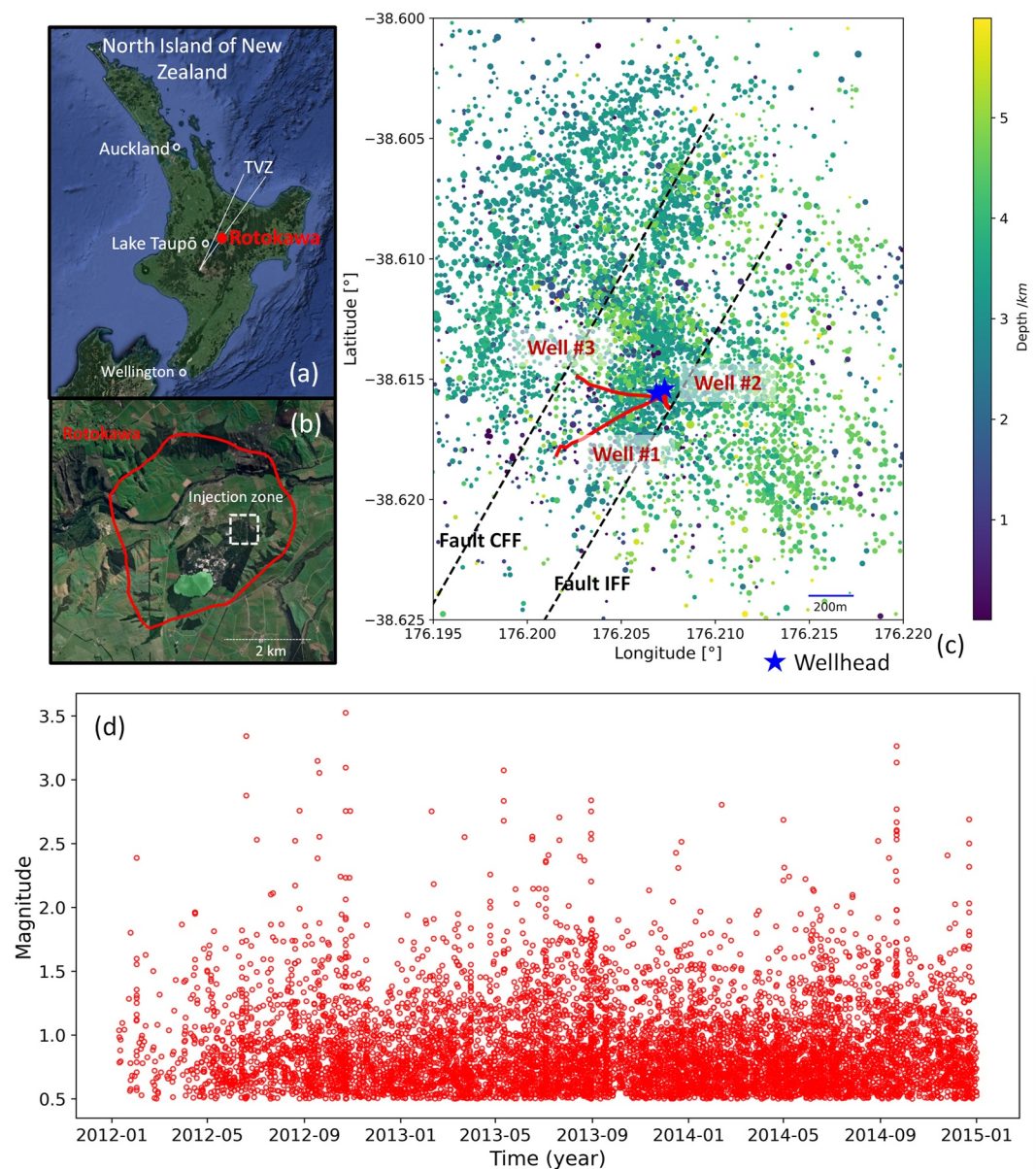


Figure 1. Overview of the injection zone at Rotokawa geothermal field. (a) Location of the field in Taupō Volcanic Zone (TVZ) of the North Island of New Zealand (red circle). (b) Field boundary with the injection zone shown as a white dashed square. (c) Location of microseismicity proximal to the reinjection zone for the years of 2012–2015 colored by depth and scaled by magnitude. Well traces are shown as red lines with wellheads indicated by blue stars. (d) Magnitude-time evolution of events between 2012 and 2015.

at the approximate depth of the permeable zones surrounding three wells, denoted numerically here #1, #2, and #3. This cluster was inferred to be bounded to the northwest by the CFF during the study period (Sherburn, Bannister, et al., 2015; Sherburn, Sewell, et al., 2015). Based on a more recent microseismicity cloud, obtained during the period 2012–2015 (Figure 1c), more events have occurred adjacent to the IFF. This indicates the cluster may be bounded by both the CFF and IFF (Hopp et al., 2020).

Two microseismicity triggering mechanisms have been proposed at Rotokawa: localized pressure rise and reservoir cooling due to injection (Sewell, Cumming, et al., 2015; Sherburn, Bannister, et al., 2015; Sherburn, Sewell, et al., 2015). These effects trigger microearthquakes by destabilizing faults or fractures that are close to shear failure (Sherburn, Bannister, et al., 2015; Sherburn, Sewell, et al., 2015). Hopp et al. (2020) have recently studied the rate and location of seismicity between 2012 and 2016 in the context of injection strategy. They found the rate

to be insensitive to major changes in well injectivity, which is the ratio of the injection rate to wellhead pressure. They also found that seismicity rate was insensitive to individual well injection volumes.

The Rotokawa data set analyzed in this work spans a 3 yr period from 2012 to 2015. We used the microseismic catalog published by Hopp et al. (2020) and filtered it to include only events in the area (Longitude: [176.19°, 176.22°]; Latitude: [−38.60°, −38.625°]) associated with injection wells #1, #2, #3. Thus, induced microseismicity located close to other injection and production wells has not been considered. We only considered events larger than the completeness magnitude of 0.5, which further reduced the catalog to 10,507 events in the 3 yr period. The allocation of fluid injection into three wells at the site are shown in Figure 1c (a subset of these data are also reported in Hopp et al., 2020).

By mid-2013, the largest volume injection well, Well #3, had suffered an injectivity decline due to unknown causes. To avoid a negative impact on power plant operation, excess fluid from Well #3 was shifted to Well #2 (Figure 2). At the same time, there appeared to be an approximate stabilization of the microseismicity rate, which had been increasing in 2012 and early 2013. In addition, prior to this operational change, Well #2 was injected with plant condensate (steam, separated from geofluid, exiting as turbine exhaust) at 40°C. Afterward, it was injected with 130°C brine (separated liquid geofluid not dispatched to the turbine), which meant its injection temperature was now similar to Well #1 and Well #3. Given the possible presence of two separate regimes, we also performed separate analyses for the two different operation periods, denoted pre-July 2013 and post-September 2013.

Production rate data were not available for use in this study. For this reason, we have limited our study to seismicity located near to the injection wells, which also happens to be the densest concentration of events (Hopp et al., 2020; Figure 3). It is reasonable to assume that, due to their proximity, these events will be predominantly controlled by reinjection activities. Furthermore, the two likeliest mechanisms for geothermal microseismicity—pressurization and thermal stressing—are conceptually associated with reinjection activities.

1.4. Húsmúli Reinjection Area of Hellisheidi Geothermal Reservoir, Iceland

The Hellisheidi geothermal field is located in the South of the Hengill volcanic area in South-West Iceland. The Hengill volcanic area is at the triple junction of three tectonic features, which results in a complex subsurface flow regime and varied induced seismicity mechanisms (Juncu et al., 2020; Mahzari et al., 2021). The field is crisscrossed by three sets of faults: NE-SW normal faults associated with the Reykjanes Peninsula Rift system; N-S strike-slip faults associated with the South Iceland Seismic Zone; and NW-SE faults reactivated in the late 1990s during a volcanic uplift event in the East of the field (Khodayar et al., 2015; Ratouis et al., 2022; Tómasdóttir, 2018). The geology comprises hyaloclastites with alternating inter-glacial basaltic lava flows and intrusions resting on an altered crystalline basement, which forms the base of the Hengill system (Franzson et al., 2005; Snaebjörnsdóttir et al., 2018).

There are two main reinjection areas in this field: Gráuhnúkar, which started in 2007, and Húsmúli, which started in 2011. In the Húsmúli reinjection area, five wells are actively reinjecting wastewater, with HN-16 injecting a combination of water and CO₂ since 2014. The locations of these five wells are shown in Figure 3c, and together they account for about 60% of the total reinjection in the geothermal field (Ratouis et al., 2019).

Previous studies of microseismicity at Húsmúli have highlighted an association between elevated seismicity and reinjection (Ritz et al., 2021; Tómasdóttir, 2018). Earlier seismic activity at Húsmúli has also been associated with repeated circulation losses during drilling (Ágústsson et al., 2015) and pump testing of the reinjection wells (Gunnarsson, 2013). When full-scale injection started at Húsmúli in September 2011, a strong increase in the rate and magnitude of events was observed, including two M_w 4 earthquakes on 15 October 2011. Both the increased pore pressure and fault slip associated with seismic events are thought to have contributed to 2 cm of observed surface deformation (Juncu et al., 2020). Other studies have focused on characterizing fluid flow paths in the context of complex geological and structural features (Mahzari et al., 2021; Ratouis et al., 2022; Snaebjörnsdóttir et al., 2018; Tómasdóttir, 2018), injection-induced surface deformation (Juncu et al., 2020), and microseismicity monitoring and management (Gunnarsson, 2013; Ritz et al., 2020, 2021).

Annual reinjection at Húsmúli is about 12 million tonnes of geothermal brine and condensate (~60°C–80°C) at depths between 1,200 and 2,100 m. Average rates into these wells have similar magnitudes over the study period (Figure 4), although there is considerable individual variability over short and medium time scales due to operational requirements. Several seismicity spikes occur during this period, which are likely due to seismic swarms

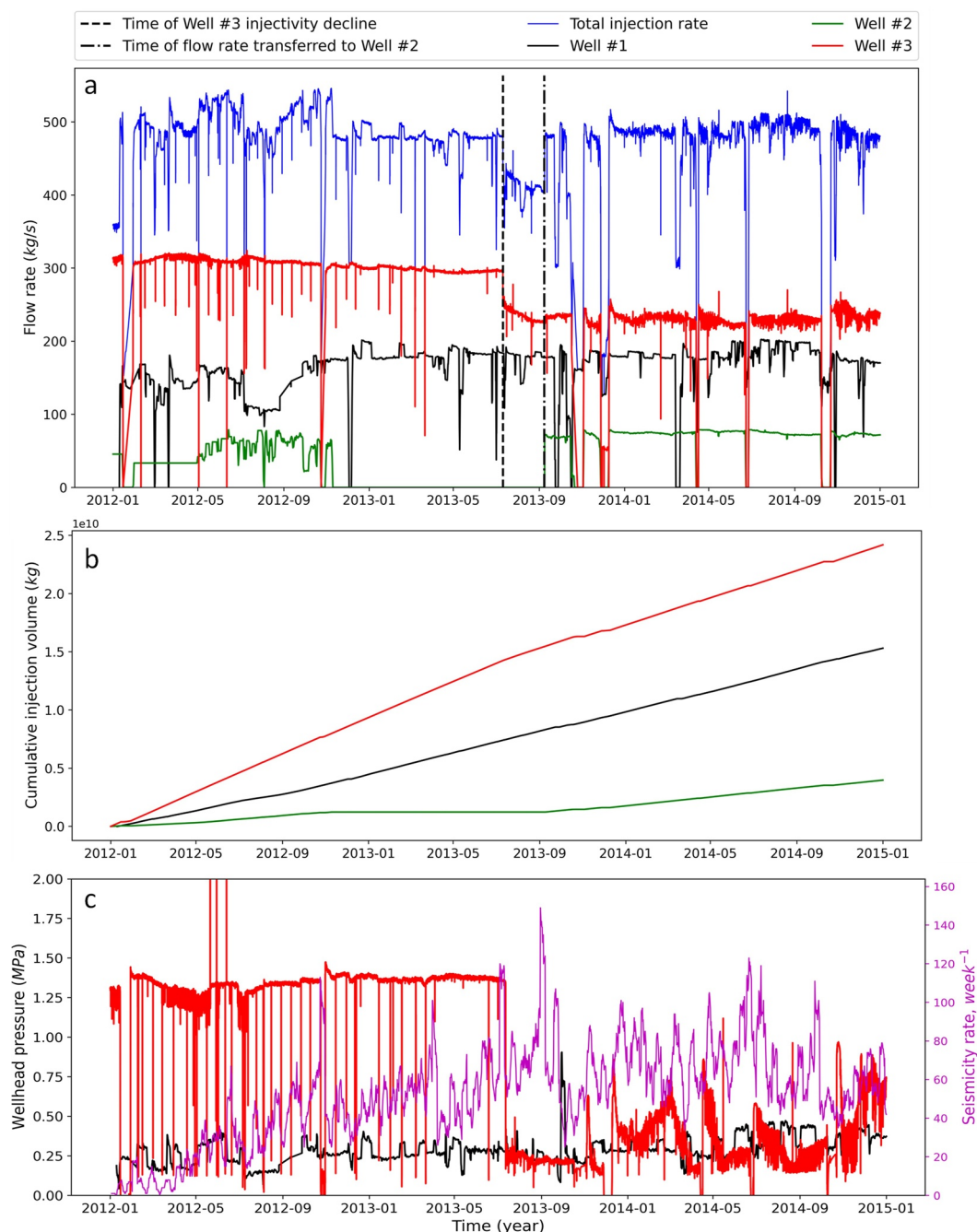


Figure 2. (a) Injection rate, (b) cumulative injection volume, and (c) wellhead pressure of Well #1 (black), Well #2 (green), Well #3 (red), and total injection rate (blue) during the period of 2012–2015 in Rotokawa geothermal field. Vertical dashed and dash-dot lines separate the pre-July 2013 and post-September 2013 operational periods. (c) also shows weekly seismicity rate changes during period of 2012–2015.

associated with injection activities, but could be a mix of natural and triggered seismicity (Gunnarsson et al., 2015; Juncu et al., 2020). While well HN-14 injects into low permeability hyaloclastite formation, HN-09 is injecting into a fairly permeable region, and HN-12 and HN-16 are drilled into a very permeable region (Gunnarsson, 2013). Injectivity is dependent on the temperature of the injected water in all wells, which could be explained by a thermomechanical effect that promotes fracture opening in the fractured reservoir (Gunnarsson, 2011). There was an injectivity increase in HN-14 from September to November of 2012, which is thought to be connected to changes in fracture connectivity and linked to deformation in the region (Gunnarsson, 2013).

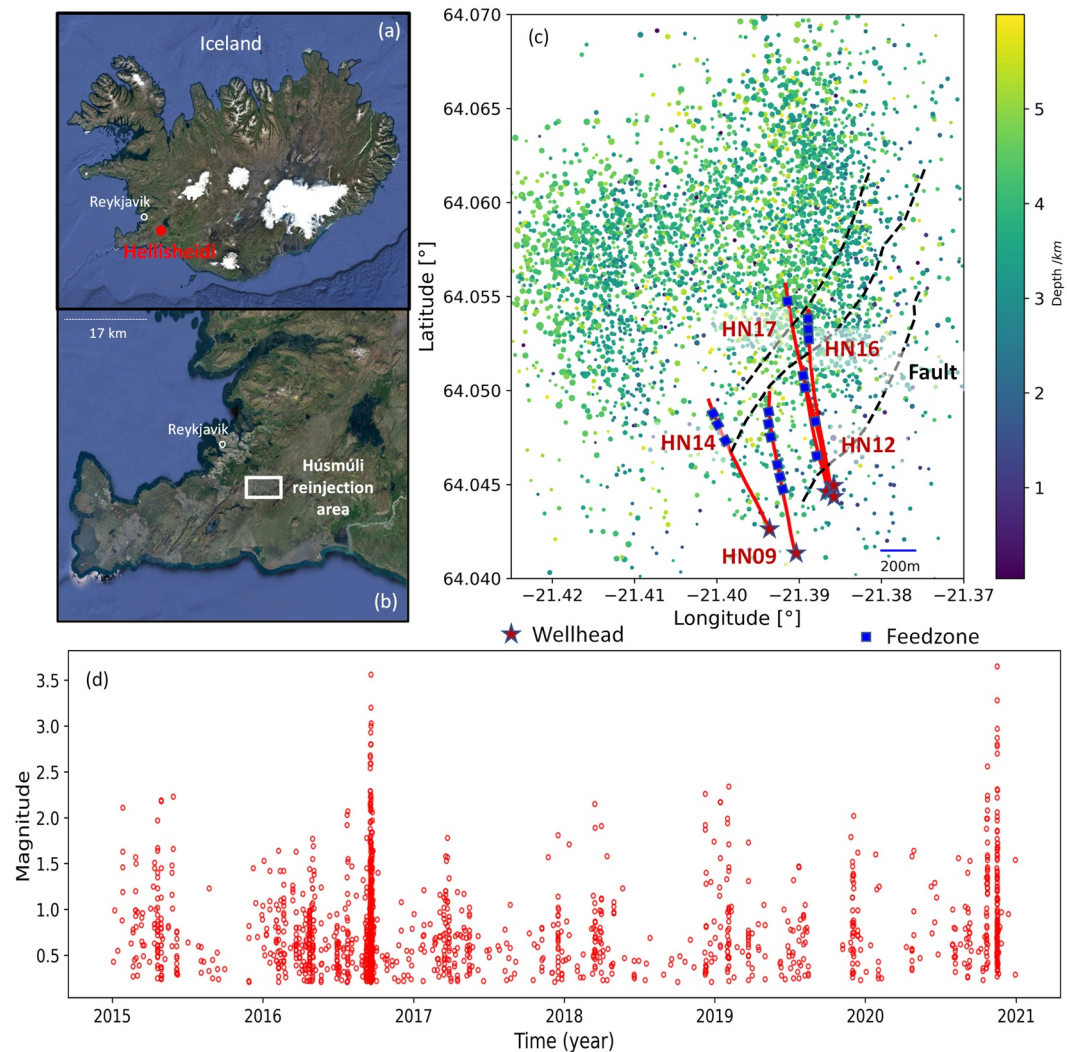


Figure 3. Overview of the Húsmúli reinjection zone at Hellisheidi geothermal field. (a) shows the location of the Hellisheidi geothermal field in the southwest of Iceland (red circle). (b) shows the Húsmúli reinjection area shown as the white box. (c) shows the microseismicity locations for the years of 2015–2020 colored by depth and scaled by magnitude. Well traces are shown by red lines with magenta star identifying the wellhead, and blue squares showing the location of feedzones. (d) Magnitude-time evolution of microseismicity between 2015 and 2020.

2. Regression Modeling

This section introduces our method for constructing linear regression models. The goal of this modeling is to find a weighted combination of injection rate and volume time series, lagged if necessary, that approximately reconstructs the microseismicity rate. The weights, lags, and their changes over time can then be interpreted in the context of geothermal reservoir management.

2.1. Model Development

Raw injection rate data for injection well i were first averaged to obtain hourly injection rate data, $q_i(t)$ for further use in this study. Missing data were imputed using linear interpolation. The cumulative injection volume for each well was calculated as the integral of the injection rate time series, $Q_i(t) = \int_0^t q_i dt$, approximated as the cumulative sum of the time series multiplied by the sample time step. To separate the effect of an individual well from the volume of fluid injected into it, we also calculated a z -score normalized injection rate, $\tilde{q}_i(t)$ and normalized cumulative injection volume $\tilde{Q}_i$ for Rotokawa and Húsmúli for their respective study periods (Rotokawa: from

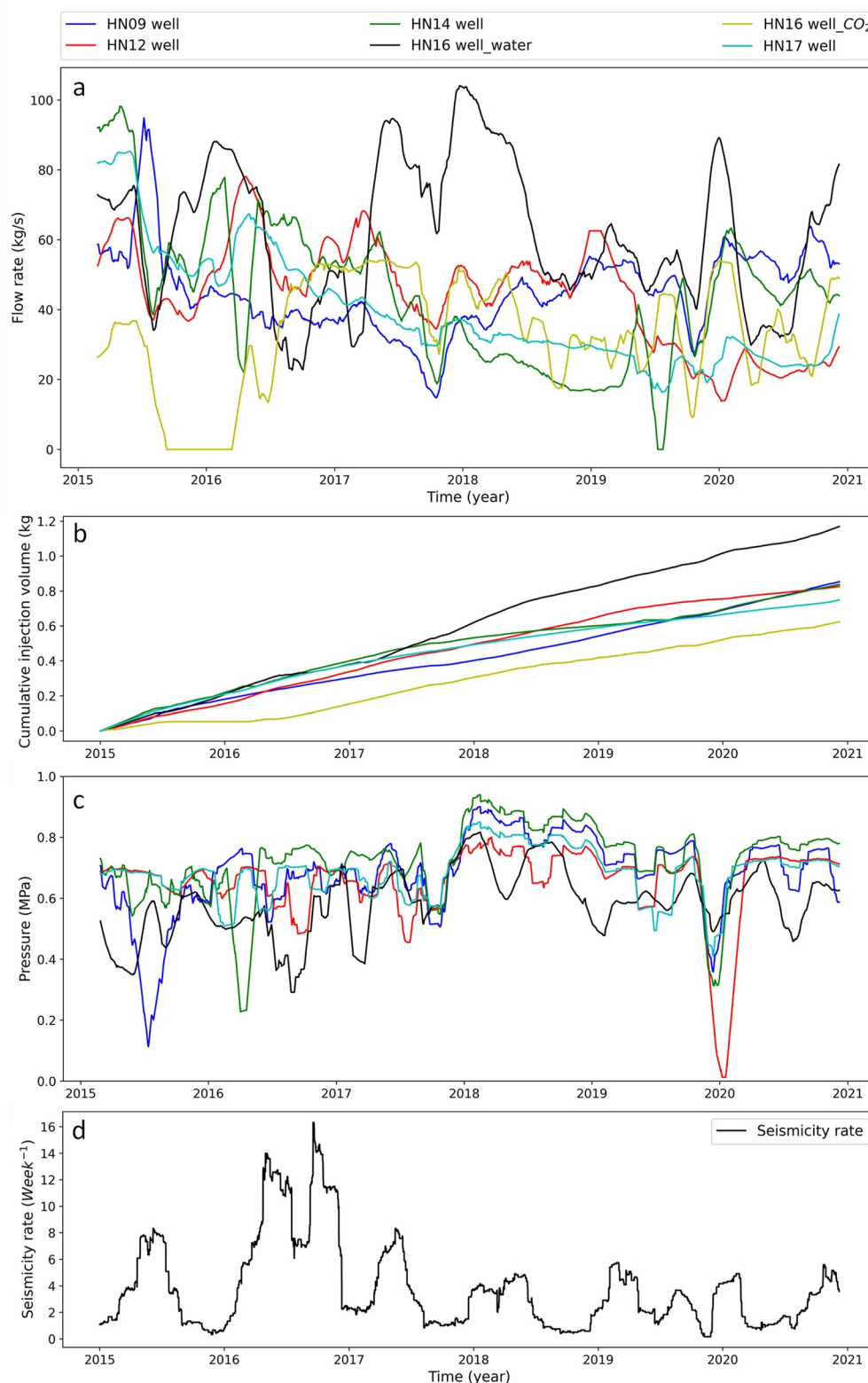


Figure 4. (a) Injection rate, (b) cumulative injection volume, (c) wellhead pressure for HN-09 (blue), HN-12 (red), HN-14 (green), HN-16_water (black), HN-16_CO₂ (yellow), HN-17 (cyan) wells between 2015 and 2020 in Húsúli reinjection area. The injection and wellhead pressure data were smoothed for illustration purposes using a moving 2 month averaging window. (d) Weekly seismicity rate changes between 2015 and 2020.

2012 to 2014; Húmulí: from 2015 to 2020). This normalization was performed by subtracting the mean then dividing by the standard deviation with both statistics computed over entire study period.

The observed microseismicity rate, λ , at time, t_j , is calculated by counting the number of events in the preceding interval $[t_j - \Delta t, t_j]$, and then dividing by the interval length, Δt , to obtain the seismicity rate. This averaging approach, using period Δt , imposes a controllable degree of smoothing on the stochastic earthquake process.

The modeled microseismicity rate was calculated as a weighted sum of q_i and Q_i (or, alternatively, the normalized $\tilde{q}_i$ and $\tilde{Q}_i$), with respective weights α_i and β_i , and an intercept term, γ . The weights were constrained to be positive, which embeds an assumption that injection and microseismicity are positively correlated. Although we cannot rule out idiosyncratic scenarios where microseismicity and injection are negatively correlated, we judged this physically informed constraint on model behavior to be reasonable.

As pressure and thermal responses are not instantaneously mediated throughout a reservoir, it is possible that seismicity may respond to injection with a delay. Therefore, we allowed for a time lag, θ , imposed uniformly on all injection data. Thus, the modeled seismicity rate for N wells can be expressed:

$$\bar{\lambda}(t) = \gamma + \sum_i^N \alpha_i q_i(t - \theta) + \sum_i^N \beta_i Q_i(t - \theta) \quad (1)$$

or an equivalent expression using $\tilde{q}_i$ and $\tilde{Q}_i$.

The best-fit model was obtained by minimizing the sum of squared differences between measured and modeled seismicity rate for different values of θ . The estimated parameters α_i , β_i , and γ were obtained by the least squares linear regression method from the scikit-learn (Pedregosa et al., 2011) Python package. Confidence intervals for the seismicity rate were constructed using the chi-square distribution (Garwood, 1936). A p -value is also given for the null hypothesis that the coefficient is zero. We use this to interpret whether a particular term, and hence the well it belongs to, has a significant effect on the seismicity rate. If it is not likely different from 0 at the 95% confidence limit, then this term is dropped from the rate model, implying the investigated well has little to no association with microseismicity.

There are alternative methods for inferring an underlying seismicity rate model that assume a non-homogeneous Poisson process (e.g., Dempsey & Suckale, 2017). When fitting, this method assigns a greater weight to periods with many microearthquakes, such as aftershock clusters that would not generally be captured by injection-driven models. Bias to these clusters degraded the fit to other parts of the catalog, even after attempts to decluster the catalog itself. In contrast, we found that the regression models were less sensitive to the brief periods of high seismicity rate that were contaminated by aftershocks.

2.2. Application to Rotokawa Geothermal Field

For Rotokawa, we constructed five regression models to test different subsets and transformations of the injection data:

1. a model for the full analysis period, 2012 to end of 2014, using the unnormalized injection data;
2. two separate models for the (2a) pre-July 2013 and (2b) post-September 2013 periods using unnormalized injection data;
3. two separate models for the (3a) pre-July 2013 and (3b) post-September 2013 periods using normalized data.

Fitted model coefficients are reported in Table 1. The 95% confidence intervals of regression model coefficients are given in Table S1 in Supporting Information S1. The results of model (3) are reported in Figure 5, and the goodness of fit of the pre-July 2013 model (3a) across different time lags is shown in Figure 6.

Of the five models, the greatest mismatch with the data is associated with Model (1), which attempts to fit the full period of seismicity with a single set of coefficients. The root mean squared error (RMSE) for this model (Table 2) is larger than for separate models (2a) and (2b).

Arguably, the improvement in model fit by considering pre-July and post-September 2013 periods separately, can be attributed by introducing additional parameters (more complexity) that allows the model to overfit to noise. A simple check is to compute the Akaike Information Criterion (AIC; Akaike, 1974) for models (1) versus (2a)

Table 1

Regression Model Coefficients and *p*-Values (Shown Beneath in Brackets) for Different Rotokawa Models

Terms	Model (1): 2012–2014 unnormalized regression model coefficients (<i>p</i> -value)	Model (2a): pre-July 2013 unnormalized regression model coefficients (<i>p</i> -value)	Model (2b): post-September 2013 unnormalized regression model coefficients (<i>p</i> -value)	Model (3a): pre-July 2013 normalized regression model coefficients (<i>p</i> -value)	Model (3b): post-September 2013 normalized regression model coefficients (<i>p</i> -value)
Intercept	1.23 (1.23 × 10 ⁻¹)	7.10 × 10 ⁻² (9.48 × 10 ⁻¹)	28.4 (5.22 × 10 ⁻¹)	120 (2.05 × 10 ⁻²⁶)	62.2 (7.07 × 10 ⁻¹⁷)
<i>q</i> ₁	3.61 × 10⁻² (1.35 × 10 ⁻²⁶⁸)	0.00 (1.00)	4.71 × 10⁻² (1.58 × 10 ⁻³⁰⁹)	2.97 (2.04 × 10 ⁻⁹⁴)	5.37 (1.58 × 10 ⁻³⁰⁹)
<i>Q</i> ₁	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)
<i>q</i> ₂	0.00 (1.00)	5.71 × 10⁻² (1.21 × 10 ⁻¹⁵⁶)	3.45 × 10⁻² (9.42 × 10 ⁻⁹)	5.71 (2.35 × 10 ⁻¹⁷⁵)	3.42 (9.42 × 10 ⁻⁹)
<i>Q</i> ₂	0.00 (1.00)	6.00 × 10⁻⁶ (2.45 × 10 ⁻²)	0.00 (1.00)	28.7 (3.99 × 10 ⁻⁶)	0.00 (1.00)
<i>q</i> ₃	2.42 × 10⁻³ (1.31 × 10 ⁻⁴)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)
<i>Q</i> ₃	2.00 × 10⁻⁶ (1.08 × 10 ⁻²⁷)	4.00 × 10⁻⁶ (2.09 × 10 ⁻²)	0.00 (1.00)	35.60 (6.01 × 10 ⁻²)	0.00 (1.00)
RMSE	18.3	13.7	16.1	13.3	16.1
Lag	9 days	11 days	5 days	11 days	5 days
AIC	2.28 × 10 ⁵	0.83 × 10 ⁵		0.83 × 10 ⁵	

Note. Significant coefficients are highlighted with bold type. The coefficients units of intercept, *Q*_{*i*}, and *q*_{*i*}, are [T⁻¹], [M⁻¹], and [M⁻¹T T⁻¹], for example, week⁻¹, kg⁻¹, kg⁻¹ s. week⁻¹. The latter unit encodes a unit conversion from flow rate (kilogram per second) to seismicity rate (per week). Coefficients of the normalized model are all [T⁻¹], that is, week⁻¹.

and (2b) together. The AIC evaluates the relative quality of each model for a given set of data by estimating the amount of information that the model loses when it is fit to the data. Smaller values of the AIC are preferred as these indicate less information loss. The AIC was calculated as follows:

$$AIC = 2k - 2LLK \quad (2)$$

$$LLK = -\frac{1}{2} \sum_{i=1}^M \frac{(\bar{\lambda}_i - \lambda_i)^2}{\sigma_i^2} - \frac{1}{2} \sum_{i=1}^M \ln(2\pi\sigma_i^2) \quad (3)$$

where *k* is the number of model parameters, LLK is the log-likelihood for a normal distribution, *M* is the number of time points, *λ_i* is a seismicity rate observation, $\bar{\lambda}_i$ is the equivalent model prediction, and *σ_i²* is the distribution variance. For a Poisson process, the variance can be approximated by the mean, and therefore we set *σ_i²* = *λ_i*. We dropped the second term in Equation 3 because it is the same for all models.

Model AIC values are given in Table 1. The value of model 2 (2a + 2b) is much smaller than that of model (1). This indicates that performance improvements due to treating injection periods separately likely reflect our capturing of additional non-noise processes.

During the pre-July 2013 period, microseismicity builds slowly over time (Figure 5a) and is linked with the rate of injection into Well #2, and volumes of injection into wells #2 and #3 (Table 1, column 2, non-zero coefficients). After the post-September 2013 change in injection strategy, microseismicity rates were more stable. The corresponding fitted model has a larger intercept (constant rate value) and now emphasizes the rate of injection into wells #1 and #2. The reduced sensitivity to Well #3 corresponds with reduced injection into this well post-September 2013. This might suggest that the average magnitude of *q*₃ has dropped below a threshold that is sensitive to microseismicity.

If injection volumes and rates had a simple linear relationship with seismicity, we should not expect there to be major changes in the model coefficients or goodness-of-fit when subdividing into pre-July and post-September 2013 periods. That we do see considerable improvement in the model, and a change of the dominant coefficients, suggests that there are nonlinear effects associated with injection into individual wells. This nonlinearity is approximated in the model through step changes in coefficients.

Within a single model, coefficient values can be difficult to compare because they are influenced by the magnitude of the injection data on which they are calibrated. Taking the pre-July 2013 model (2a) as an example,

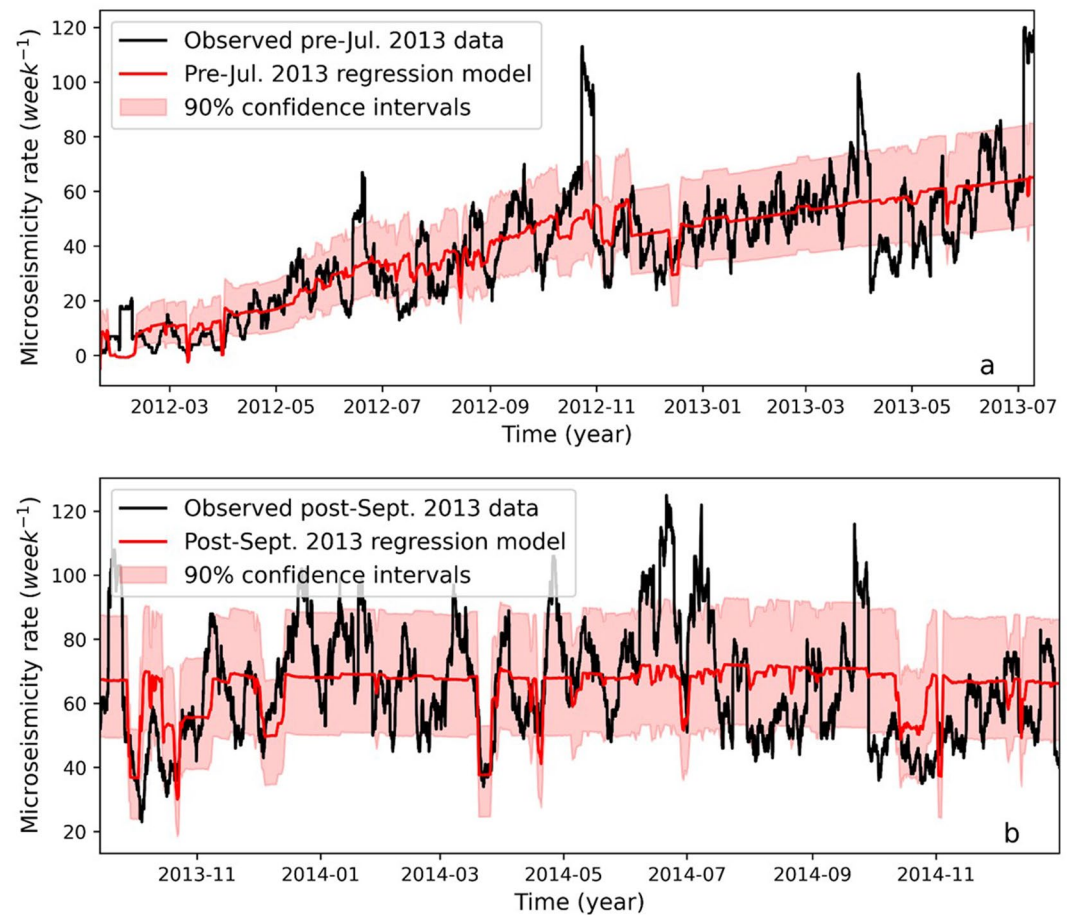


Figure 5. Comparison between observed and modeled seismicity rates for models (a) pre-July 2013 and (b) post-September 2013 (models 3a and 3b, respectively). Models are fitted using a non-negative linear regression model with normalized Rotokawa injection data. An 11-day lag was used for pre-July 2013, and 5-day lag was used for post-September 2013 model. One week Δt is used to calculate the seismicity rate.

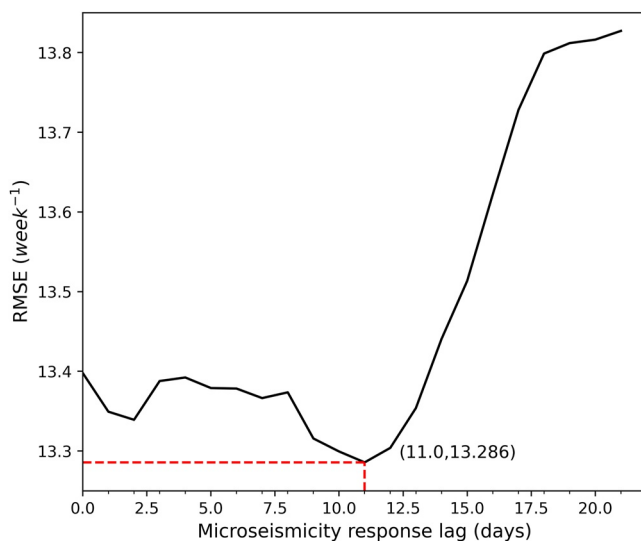


Figure 6. Root mean squared error (RMSE) of regression model (3a) under different lag values, for models fit to the pre-July 2013 period.

coefficients associated with volumes are orders of magnitude smaller than those attached to rates (Table 1 column 2). It is arguable though that the volume terms should be more important because, as integral measures, they introduce an accumulating trend that correlates well with the increasing microseismicity rate. To remove the effect of data magnitude on the coefficients, we refitted the models using the normalized injection data (for pre-July (3a) and post-September 2013 (3b) separately). This makes it easier to separate the relative importance of individual wells, as well as the whole-sale effect of the change in operational strategy.

Using normalized data to fit a model for the pre-July 2013 period, we found contributions from all wells on either injection rate or volume terms (Table 1, column 4), with the largest coefficients assigned to monotonically increasing terms Q_2 and Q_3 . This suggests that seismicity during the period can be characterized by an increasing sensitivity to injection—that is, more and more microearthquakes are triggered per unit total injection rate—and that this is being driven by the fluid pressure changes that increase in size and extent with the cumulative injection volume. Fluctuations in the individual rates into wells #1 and #2 may be driving seismicity rate fluctuations about a first-order increasing trend. The best-fitting time lag for pre-July 2013

Table 2
Regression Model Coefficients and p -Values (Shown Beneath in Brackets)
for Different Húsmúli Models With 3-Month Averaging Δt and No Lag

Terms	Unnormalized regression model coefficients (p -value)	Normalized regression model coefficients (p -value)
Intercept	-0.42 (1.54×10^{-4})	0.42 (0.00)
q_{HN09}	0.00 (1.00)	0.00 (1.00)
q_{HN12}	9.92×10^{-4} (2.67×10^{-130})	6.10×10^{-2} (2.67×10^{-130})
q_{HN14}	3.01×10^{-4} (1.70×10^{-31})	2.64×10^{-2} (1.70×10^{-31})
q_{HN16_w}	0.00 (1.00)	0.00 (1.00)
$q_{HN16_{CO_2}}$	1.33×10^{-3} (0.00)	9.45×10^{-2} (0.00)
q_{HN17}	1.75×10^{-3} (2.09×10^{-151})	0.11 (2.09×10^{-151})
Q_{HN09}	0.00 (1.00)	0.00 (1.00)
Q_{HN12}	0.00 (1.00)	0.00 (1.00)
Q_{HN14}	0.00 (1.00)	0.00 (1.00)
Q_{HN16_w}	0.00 (1.00)	0.00 (1.00)
$Q_{HN16_{CO_2}}$	0.00 (1.00)	0.00 (1.00)
Q_{HN17}	1.39×10^{-8} (7.39×10^{-1})	2.83×10^{-2} (7.39×10^{-1})
RMSE	3.13	3.13

Note. Significant coefficients are highlighted with bold type. The coefficients units of the intercept, Q_i , and q_i are $[T^{-1}]$, $[M^{-1}]$, and $[M^{-1}T^{-1}]$, for example, $week^{-1}$, kg^{-1} , $kg^{-1} s$, $week^{-1}$. The latter unit encodes a unit conversion from flow rate (kilogram per second) to seismicity rate (per week). Coefficients of the normalized model are all $[T^{-1}]$, that is, $week^{-1}$.

The seismicity rate was calculated using a 3-month averaging window, which helped reduce the effect of short-term spikes that we could not capture during regression modeling. We settled on this window after testing values between one and 3 months, and finding that 3 months returned the lowest RMSE for both normalized and unnormalized injection data (Table S2, Table S3). We also tried to determine a best fit lag period (θ) but found this returned unrealistic values of 2–3 months (Tables S2, Table S3). This suggests lags may not be applicable for the Húsmúli reinjection area and, therefore, we decided to set $\theta = 0$ in subsequent regression models.

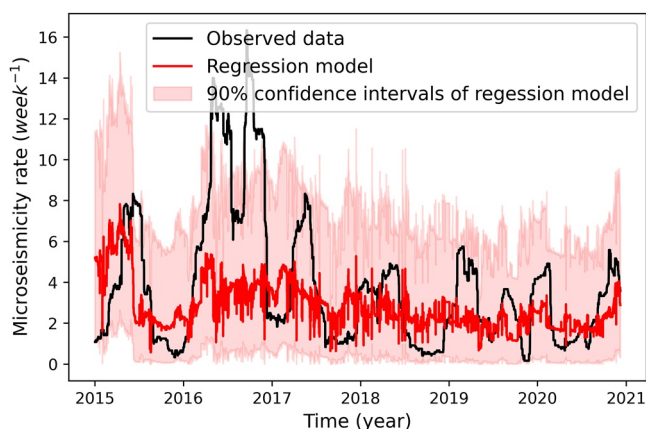


Figure 7. Comparison between normalized linear regression model (red) and observed seismicity rate (black) at the Húsmúli reinjection area.

period appears to be 11 days, although we note that relatively low misfits are returned for all lags less than 2 weeks (Figure 6).

For the post-September 2013 period, there appears to be a change in reservoir response. Although total injection rate remains largely the same, microseismicity stabilizes. Model coefficients associated with cumulative injected volume (β_i) are dropped from the best-fit model in this period, and only coefficients associated with rate injection into wells #1 and #2 are retained. This continues to suggest seismicity rate fluctuations around a background average may be controlled by these two wells. We note with interest a change in their relative coefficient weights: Well #2 had a stronger influence than Well #1 in the pre-July period, but weaker in the post-September period. This is in spite of relatively more fluid injection into Well #2 post-September.

2.3. Application to Húsmúli Reinjection Area

We used data from five active wells in the Húsmúli reinjection area, covering 6 yr of continuous injection between 2015 and 2020. Since 2014, there was injection of both wastewater and condensate CO_2 into HN-16. We treated these two fluids as different injection terms in the regression model, in case there were individual effects that could be attributed. The injection rate and volume data were preprocessed by applying z-score normalization, similar to model (3) of Rotokawa.

To reduce the effect of aftershock clusters, the catalog was declustered by removing all events in a 1-week window following events larger than M_L 2.5. Unlike with Rotokawa, this did improve aspects of the model fit, although it does introduce a degree of arbitrariness in the application of declustering.

We also had some difficulty with model fits that returned negative seismicity rates. To avoid this unphysical outcome, we fitted the well data to the logarithm of seismicity rate, and then applied the inverse transform when computing the model output.

Coefficients and p -values for individual wells used to model Húsmúli seismicity are shown in Table 2. The regression model results along with RMSE obtained using unnormalized data were analyzed to compare the normalized regression model results. We note that the relative magnitude of the coefficients is unchanged between the two models, which could be because the mean and variance of the injection rates of the individual wells are similar (Figure 4; Table S4 in Supporting Information S1).

The best-fit normalized regression model results are compared with the observed seismicity data in Figure 7. Most of the raw seismicity data fall in the 90% confidence interval of the regression model, although there is clearly systematic over and under prediction in places where it is unable to capture periods of elevated seismicity, like those observed in 2016. The coefficients of each regression term for the normalized best-fit regression model are illustrated in Table 2.

The HN-17 and HN-16 condensate-CO₂ injection rate terms have the largest coefficients. This suggests that microseismicity may be sensitive to these two wells, and particularly the introduction of CO₂ injection. HN-12 and HN-14 wells also have some association with microseismicity rates as indicated by non-zero coefficients (albeit, smaller than HN-16 and HN-17). The microseismicity does not appear to be sensitive to the HN-09 and HN-16 water injection feedzones, although given the relatively poor fit of the model to the data, it is difficult to be conclusive. None of the injection volume terms were selected as important, which probably reflects long-term stability in the seismicity rate (discounting medium-term departures from the average). This long-term stability of the seismicity rate is consistent with continuous injection since September 2011 when the seismicity quickly ramped up before stabilizing (Bessason et al., 2012).

3. Time Series Feature Engineering

This section introduces the second data-driven modeling technique. In contrast to the regression models used in Section 2, the goal here is search for latent patterns in the injection time series and to see whether these can be linked with the microseismicity rate. Given the systematic and voluminous nature of such pattern mining, we also evaluate whether the patterns are statistically significant while building in safeguards to avoid false discovery. We then seek physical interpretations for individual patterns, or particular data types that produce a larger number of significant patterns.

The time series feature engineering techniques used in this paper were initially developed for anomaly detection in industrial steel casting (Christ et al., 2016). They have since been used to study seismic tremor in subduction zones (Rouet-Leduc et al., 2018) and volcanic tremor prior to eruptions (Ardid et al., 2022; Dempsey et al., 2020). Here, we are exploring whether they may be useful in an induced microseismicity setting. The Python package tsfresh was used to automatically perform time series feature extraction and selection on the basis of the FRESH algorithm (Christ et al., 2016). The result is over 700 extracted features from the raw data (e.g., injection rate, pressure) and its transforms (e.g., hydraulic work rate). These features are evaluated for statistical significance using a labeled quantity of interest, in this case, the microseismicity rate.

3.1. Feature Extraction and Selection

3.1.1. Data Compilation

We used the same earthquake catalogs for Rotokawa and Húsmúli as used in the regression modeling described in previous section. For injection rate, q_i , and wellhead pressure data, P_i , we used hourly records for the three wells at Rotokawa over the 3 yr period 2012–2014. For Húsmúli, we converted the original data set (at 10 min intervals) to hourly records of injection rate into each of the five wells between 2015 and 2020. Wellhead pressure data were not available for the Húsmúli data. Gaps were imputed using linear interpolation. Injection rate was also summed for all wells, q_{total} , or sometimes sub-groups of wells.

We also performed feature extraction for several transformations of the data. These included the hydraulic work rate $E_i = q_i P_i$, which is the product of injection rate and wellhead pressure. We chose this quantity because it potentially quantifies both the magnitude and growth of the pressurized volume. The former characteristic it achieves via the factor P_i , which sets a wellbore boundary condition that controls the largest value of pressure increase. At the same time, the factor q_i , controls the rate that fronts of constant pressure diffuse away from the injection well and is thus a proxy for growth rate. Strictly speaking, hydraulic work rate should be calculated using the downhole pressure, because wellhead pressure can be contaminated by wellbore processes (friction, phase change) that do not affect the reservoir. Unfortunately, these data were not available. In addition to E_i , q_i , and P_i , we also extracted features from their derivatives, $\dot{E}_i$, $\dot{q}_i$, and $\dot{P}_i$.

In total, there were 39 data streams: nine each for Well #1 and #3 at Rotokawa (P_i , q_i , E_i , their time derivatives, and integrals), three for Well #2 at Rotokawa (q_i , its time derivative and integral), three for each of the six wells at Húsmúli (q_i , its time derivative, and integral).

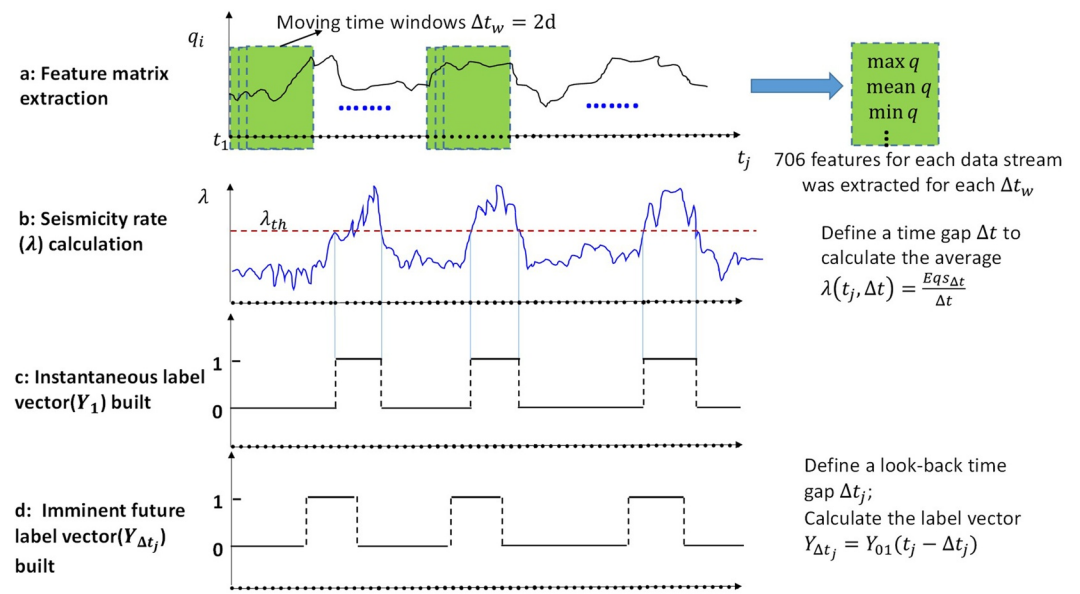


Figure 8. Schematic diagram of feature matrix and label vector constructed using the time series feature engineering method. (a) Feature time series are extracted from moving window calculation on the original data. (b) Seismicity rate, λ , is judged above or below a threshold, λ_{th} , which produces (c) a binary label vector, Y_1 , that characterizes the instantaneous seismicity as high or low: $1 = \lambda \geq \lambda_{th}$, $0 = \lambda < \lambda_{th}$. (d) A look-back binary label vector, $Y_{\Delta t_j}$, which selects all time windows that occur prior to (within a period Δt_j) elevated seismicity, $Y_1 = 1$. This second label vector characterizes when a period of elevated seismicity will soon begin.

3.1.2. Feature Extraction From Injection Data

Automatic time series feature extraction was performed as follows:

1. Each time series was subdivided into overlapping windows of length $\Delta t_w = 2$ days. Each window overlapped the previous by 75% (Figure 8a). Respectively, this yielded about 2,200–3,600 windows for the 3 yr of Rotokawa data and 6 yr of Húsmlí data.
2. For each window, the Python package tsfresh (Christ et al., 2016) was used to calculate 706 time series features. Examples of time series features include distribution parameters (e.g., mean, max, standard deviation), measures of periodicity (e.g., Fast Fourier Transform coefficients, and lag coefficients), information (energy, entropy), and parameters of linear and nonlinear regressors.

An example of a time series feature is the quantile (q). This is calculated by ordering all values within a window and selecting the value exceeded by $q\%$ of all others.

3.1.3. Seismicity Label Vector Calculation

Binary labels were used to subdivide the values of each feature time series into two sets. By casting as a classification problem, we can look for features that might be linked with discrete changes in seismicity behavior, which is suspected to have occurred after the August 2013 operational change at Rotokawa.

We used a binary label to classify whether, at time, t , the seismicity rate in the near-future, $t + \Delta t_j$, would be high (1) or low (0). This allowed us to separate injection or pressure feature values according to whether they preceded high or low seismicity. Then, we checked for any significant difference between those two distributions and, if discovered, whether that could be justified by a physical argument.

The label vector is constructed as follows (Figure 8):

1. As with the calculation of seismicity rate in Section 2, we first compute a continuous estimate of the earthquake rate averaged over some time period, that is, $\lambda(t, \Delta t) = Eq_{s\Delta t}/\Delta t$. The term $Eq_{s\Delta t}$ is the number of events in the time period $[t - \Delta t, t]$. By increasing Δt , we construct an earthquake rate that is smoothed over longer time scales. Furthermore, by investigating feature patterns for different Δt , we can determine if

- different injection behaviors affect short or long-term seismicity rate. The seismicity rate calculation is backward looking, that is, calculated using knowledge of prior events only.
- Next, we compute an instantaneous seismicity label vector, Y_1 . This is a binary classification of the current seismicity rate as either high or low compared to a pre-determined threshold rate, λ_{th} (Figures 8b and 8c). If $\lambda > \lambda_{th}$, a value $Y_1 = 1$ is assigned, otherwise, $Y_1 = 0$. Here, we select λ_{th} as a quantile of all recorded λ values over the study interval, for example, λ_{80} is the seismicity rate below which 80% of the seismicity rate observations fall.
 - We are interested in whether particular injection patterns occur before high or low seismicity. Therefore, we defined a look back time gap Δt_j and used this to select the time windows that precede $Y_1 = 1$. Formally, the vector $Y_{\Delta t_j}$ (Figure 8d) for time gap Δt_j , is expressed $Y_{\Delta t_j} = \text{any}([Y_1([t, t + \Delta t_j]) = 1])$.

For example, the schematic label vector shown in Figure 8d picks out three periods of relatively elevated seismicity. The label $Y_{\Delta t_j}$ highlights time series features that occur before the elevated seismicity, so we can study them more closely and see if they are systematically different from feature values at low seismicity.

3.1.4. Statistical Tests

Statistical tests are used to determine which of the many possible injection features are relevant to the seismicity periods of interest. For each feature, we evaluate its significance by applying a Mann-Whitney-U test (Mann & Whitney, 1947). This assesses whether the distribution of feature values corresponding to label $Y_{\Delta t_j} = 0$ and the distribution corresponding to label $Y_{\Delta t_j} = 1$ have different medians. Each test outputs a p -value. By comparing the p -value to a minimum threshold, we can determine which features are likely to have different values prior to high seismicity periods. Furthermore, by ordering the p -values, we can assign a ranking to significant features.

Because of the large number of tests performed—one for each of approximately 700 features computed for 39 data streams—we took steps to minimize false discovery of non-significant features. Before reporting results for any individual feature, we checked it would not be rejected with respect to a downward-adjusted Benjamini-Yekutieli p -value. This controls false discovery of spurious associations to a rate of 5% by imposing a p -value threshold of 0.026 (Benjamini & Yekutieli, 2001; Figure S1 in Supporting Information S1).

In certain cases, for instance when evaluating the relative significance of a single data type, we found it useful to analyze a total p -value, $\mathcal{P}$, obtained by summing the logarithm of p -values over all features for a single data type:

$$\mathcal{P} = \sum_i^F \log(p_i) \quad (4)$$

where p_i is the p -value of the i th feature and F is the number of features. This score reports the bulk association of a particular data type with periods of high microseismicity. However, it does not give information about the nature of that association. Also, because it combines both significant and non-significant features, it can only be used as a comparative metric for understanding trends between wells, data-types, or across different time-scales, Δt . Larger associations between injection and seismicity are recorded as larger negative numbers.

To avoid spurious artifacts when calculating $\mathcal{P}$ for comparisons, we:

- Ensured a balanced label set, by performing random under-sampling of windows so they had equal numbers of 1's and 0's. This introduces some randomness that could generate different p -values depending on the random number seed. Therefore, we performed the under-sampling 100 times for different seeds and took the median p -value.
- Controlled for samples size. This is because, when comparing across different values of λ_{th} or for different wells, we cannot guarantee that the undersampled feature sets, although balanced by label, would have the same size. Therefore, the undersampled features were randomly downsampled so that they are all contained the same numbers of values. This was set at the minimum size among all undersampled feature sets for all values of λ_{th} and different wells.

Our analysis of time series features at Rotokawa and Húsmúli is structured as follows. In Section 3.2, we inspect a few individual feature distributions that are associated with high and low seismicity periods at Rotokawa. Table 3 includes detail about significant features used in this section. In Sections 3.3 and 3.4, we test for associations between long-term and short-term seismicity rate, and the injection rates at individual wells at both Rotokawa and

Table 3
Feature Description Used in the Figures 9–12 and 15

Feature name	Description
Mean (X)	Calculate the mean values of time series X .
Minimum (X)	Calculate the lowest value of time series X .
Maximum (X)	Calculate the largest value of time series X .
Quantile (X, q)	Calculate the q quantile of time series X , this is the values of X greater than $q\%$ of the ordered values from X . Features of quantile__q_0.3 and quantile__q_0.2 return the 30%, 20% values of time series X .
Sum_values	Calculate the sum over the time series values X .
Change_quantiles ($X, ql, qh, isabs, f_agg$)	First fixes a corridor given by the quantiles of ql and qh of the distribution of time series X , then calculate the average, absolute value of consecutive changes of the series X inside of this corridor. Change_quantiles__f_agg_“mean”isabs_True__qh_0.4__ql_0.0 applies $qh = 0.4$, $ql = 0.0$ quantile to time series to get the corridor of the time series first. Then, it calculates the average absolute value of changes in this corridor.
Ar_coefficient ($X, param$)	This feature is to fit the unconditional maximum likelihood of an autoregressive AR(k) process, the k parameter is the maximum lag of the process: $X_t = \varphi_0 + \sum_{i=1}^k \varphi_i X_{t-i} + \varepsilon_t$, where X is the time series to calculate the feature. $param$ contains two parameters of k as the maximum lag and i , the φ_i is the returned fitting value when getting the maximum likelihood. The “ Pressure_ar_coefficient_k_10_coeff_4 ” feature is a component of an autoregressive model constructed from the pressure data. The best fit model with maximum likelihood is when $\varphi_i = \varphi_4$, for lags tested up to 10 sample units.
Autocorrelation (X, lag)	Calculate the autocorrelation of the specified lag according to the below formula: $\frac{1}{(n-l)\sigma^2} \sum_{i=1}^{n-l} (X_i - \mu)(X_{i+l} - \mu)$, where n is the length of the time series X_i , σ^2 is the variance and μ is the mean of the time series X_i . l denotes the lag. Autocorrelation__lag_1 calculates the autocorrelation when lag = 1.
Number_peaks (X, n)	Calculate the number of peaks of at least support n in the time series. A peak of support n is defined as a sequence of x where a value occurs, which is bigger than its n neighbors to the left and to the right.
Approximate_entropy (X, m, r)	Implements a vectorized approximate entropy algorithm (Yentes et al., 2013). Where m is the length of compared with sun of the data, r is the filtering level. Approximate_entropy__m_2__r_0.9 calculates the approximate entropy when $m = 2$ and $r = 0.9$.
Agg_linear_trend ($X, param$)	Calculates a linear least squares regression for values of the time series that were aggregated over chunks versus the sequence from 0 up to the number of chunks minus one. Variable $param$ contains dictionaries {“attr”: x , “chunk_len”: l , “f_agg”: f }. The parameters $attr$ controls which of the characteristics are returned. Possible extracted attributes are “pvalue”, “rvalue”, “intercept”, “slope”, “stderr”; the chunksize is regulated by “chunk_len”. It specifies how many time series values are in each chunk; further, the aggregation function is controlled by “f_agg”, which can use “max”, “min” or, “mean”, “median.” Agg_linear_trend__f_agg_“max”__chunk_len_5__attr_“intercept” calculates the linear least squares regression for X based on the intercept characteristic when chunksize = 5 and max aggregation function.

Húsmúli. The goal is to rank the different wells in terms of their relative association with microseismicity. This is also an opportunity to compare with the coefficient ranking of the regression analysis in Section 2. Finally, in Section 3.5, we explore the relative importance of different data types: injection rate, wellhead pressure, hydraulic work rate, injection volume. As the relationship between injection rate reduction and induced seismicity remains largely enigmatic, we also explored for evidence that well throttling—rapid changes in injection rate, q —can trigger microseismicity.

3.2. Feature Analysis of Rotokawa Injection Data

Significant features of the data were identified on the basis of low p -values and then further understood by visualizing their distributions corresponding to label vectors 1 and 0 separately. The goal is to understand how aspects of the injection behavior can affect the microseismicity that follows. In this section, we plot a small selection of features to give an impression for how their distributions can vary across different data streams, features, and wells. A more exhaustive list of feature distributions is given in Supporting Information S1.

Figure 9 shows the distribution of the “median” feature for injection rate into Well #1. Values corresponding to “high seismicity” tend to plot to the right, that is, being preceded by larger values of injection rate. Increasing the amount of seismicity smoothing (Δt) tends to increase overlap of the two distributions (Figure 9, top row),

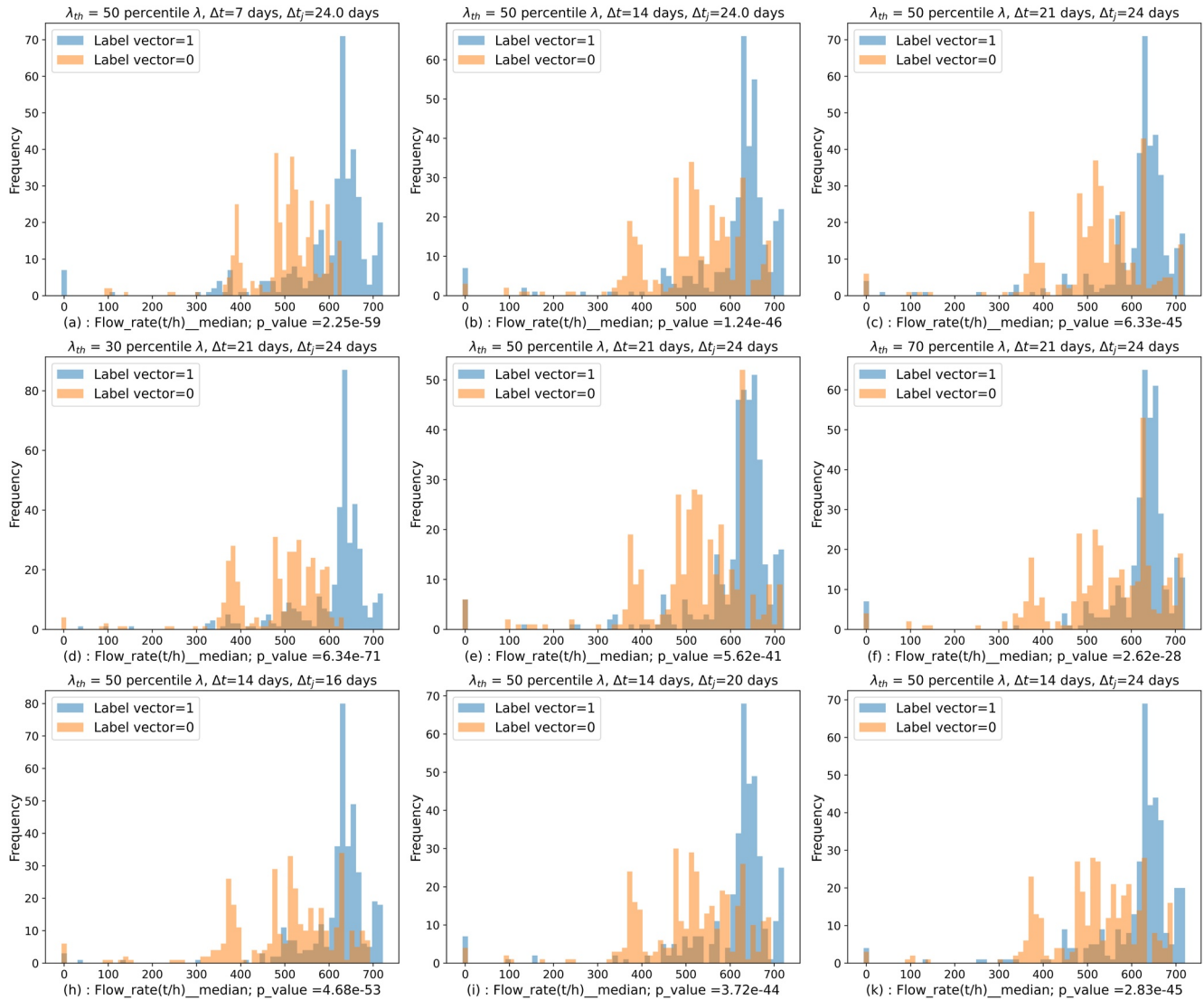


Figure 9. How binarized distributions of median injection rate into Well #1 change across ((a–c), top row) varying Δt , that is, increased smoothing of seismicity, ((d–f), middle row), varying threshold, λ_{th} , and, ((h–k), bottom row) look-forward period, Δt_f . t/h (tonnes per hour) shown in the x axes of each subplot is the unit of flow rate.

although this is not necessarily true for all features. Setting a lower threshold, λ_{th} , to subdivide high and low seismicity appears to lead to a clearer separation of the distributions (Figure 9, middle row). However, in this study, we are more interested in the fewer periods of higher microseismicity and so therefore have tended to select λ_{th} at the 50th or 80th percentile. Finally, the look forward period (Δt_f) has a minor impact on distributions for this feature (Figure 9, bottom row).

Figure 10 plots the distribution of the median injection rate features for individual wells and their sum (the median is the middle injection rate value in the 2-day window). High and low seismicity are here separated by the 50th percentile value (7.85 events per day) applied to a seismicity rate calculated using a 20-day smoothing window ($\Delta t = 20$ days).

First, we note a weak association between total injection rate and seismicity (Figure 10a), which is indicated by a lack of clear separation in the distributions of median injection corresponding to low and high seismicity periods. In contrast, Figures 10b and 10c show that higher injection rates precede higher rates of microseismicity for injection into Wells #1 and #2, whereas Figure 10d highlights a negative association with Well #3 injection rate.

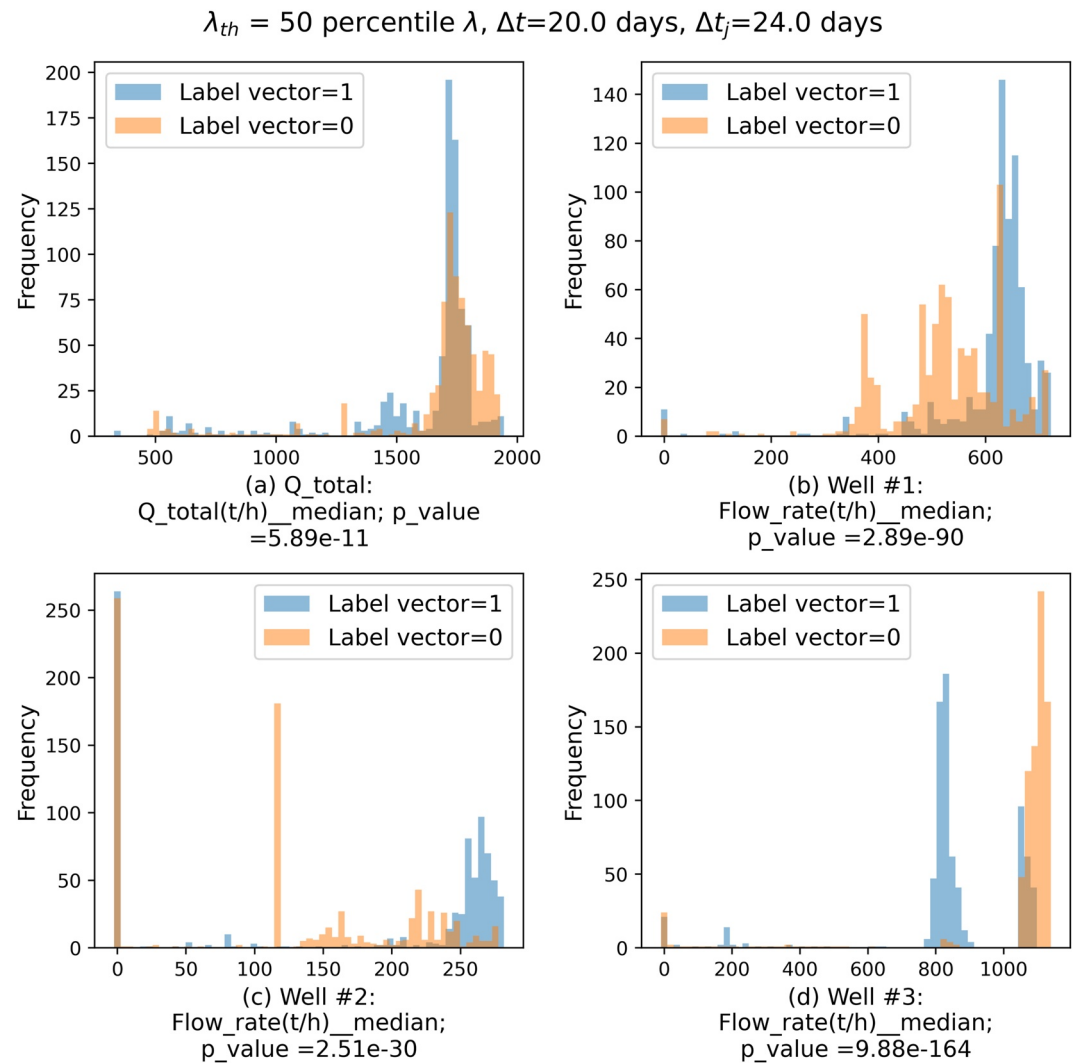


Figure 10. Median injection rate distributions for (a) total injection rate, (b) Well #1, (c) #2, and (d) #3. Here, we are showing the same time series feature (median) and data stream (q), for different wells and well combinations. t/h (tonnes per hour) shown in the x axes of each subplot is the unit of flow rate.

These trends are reflecting the shift in field operation that occurred in August 2013 when a portion of reinjection was switched from Well #3 to Well #2 (Hopp et al., 2020). At the same time, there was an increase to a new stable microseismicity rate, which accounts for the positive and negative associations with Well #2 and Well #3, respectively. Although these trends had been previously identified in Section 2, it is a useful verification that corroborates the earlier model.

Figure 11 shows wellhead pressure features for wells #1 and #3. Unfortunately, pressure data were not available for Well #2. As described in Table 3, the “Pressure_sum_values” feature is calculated as the sum of all values in each 2-day time window. Therefore, it is proportional to average pressure. The “Pressure_ar_coefficient_k_10_coeff_4” feature is a component of an autoregressive model constructed from the pressure data in a window. An autoregressive model fits a time series, X_t , in terms of lagged versions of itself, that is,

$$X_t = \varphi_0 + \sum_{i=1}^k \varphi_i X_{t-i} + \varepsilon_t \quad (5)$$

where $k = 10$ is the maximum lag of the process and the coefficient φ_4 (“coeff_4”) is the model term associated with the 4 hr lagged time-series.

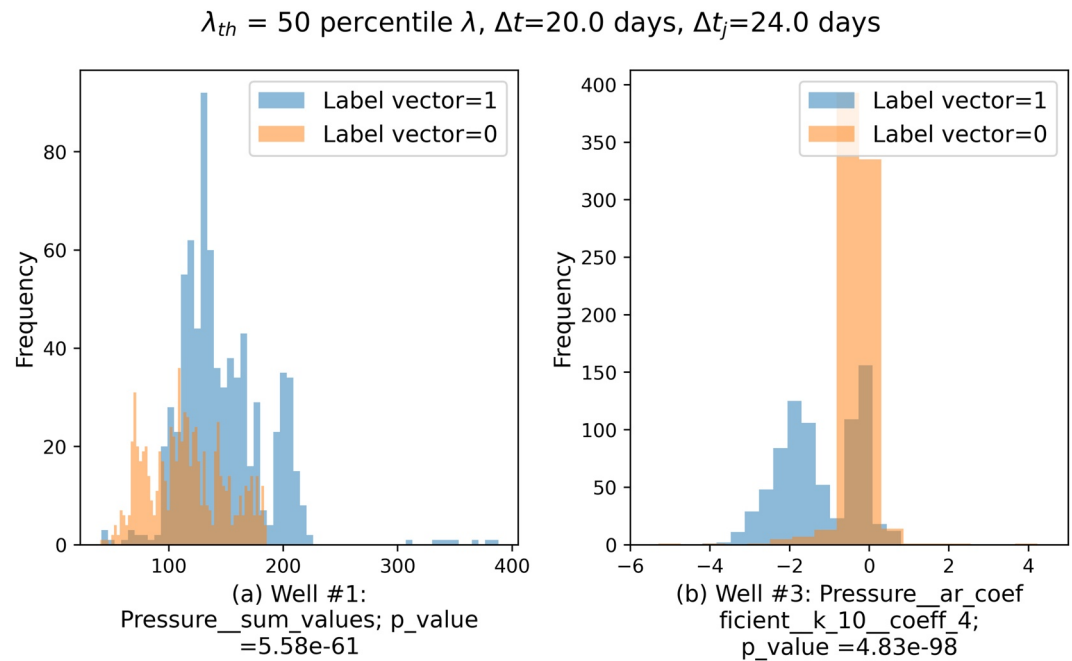


Figure 11. Significant features distributions of pressure (p) data stream for Well #1 (a) and Well #3 (b). Here, we are showing the most significant time series feature (smallest p -value) for the same data stream (P) at two different wells where it is available.

Both features in Figure 11 show some separation of the feature distributions associated with high and low microseismicity, although there is quite a large degree of overlap in their values. This means that, although we can develop a hypothesis to account for the difference, it is, at most, part of the story. Other technical processes that influence the seismicity rate are clearly in operation.

The association of higher wellhead pressure at Well #1 with elevated microseismicity (Figure 11a) is consistent with standard models that explain induced earthquake triggering in terms of effective pressure reduction on faults (Hubbert & Rubey, 1959). However, the overlap of distributions indicates that other factors are also important, which likely includes both causative (e.g., coincident operations at other wells) and random effects (e.g., heterogeneous fracture or stress distributions).

The association of negative autoregressive coefficients at Well#3 with high microseismicity (Figure 11b) is an example of a feature that is difficult to interpret (in contrast to Figures 10 and 11a). This is because the autoregressive model is complex and its coefficients need to be interpreted alongside other values φ_i and φ_0 . One approach could be to reconstruct the models and to plot their output across samples of label 1 and 0 windows. However, such reconstructions for all time series features is a non-trivial task. Our goal is to establish associations between particular wells or data-types, and the seismicity, for which the distribution analysis is sufficient. Developing and testing explanatory hypotheses for the autoregressive coefficients and other complex features is left to follow-up studies.

Figure 12 plots significant features for derivatives of injection rate at Rotokawa for individual wells and their sum. High and low seismicity are here separated by the 50th percentile value (7.85 events per day) for a rate calculated using a 7-day smoothing window ($\Delta t = 7$ days). A detailed description of these features is given in Table 3. Although the distributions are different, there are not really clear separations prior to low and high seismicity that would motivate development explanatory hypotheses.

3.3. Association of Rotokawa Microseismicity Rate With Injection Well Features

The previous section gives a basic indication of how one feature of one data type (e.g., injection rate) can associate differently with seismicity. However, it is not practical to visually inspect all features independently. Here, we extend the analysis to include more features by investigating the aggregate association, $\mathcal{P}$ (Equation 4). We

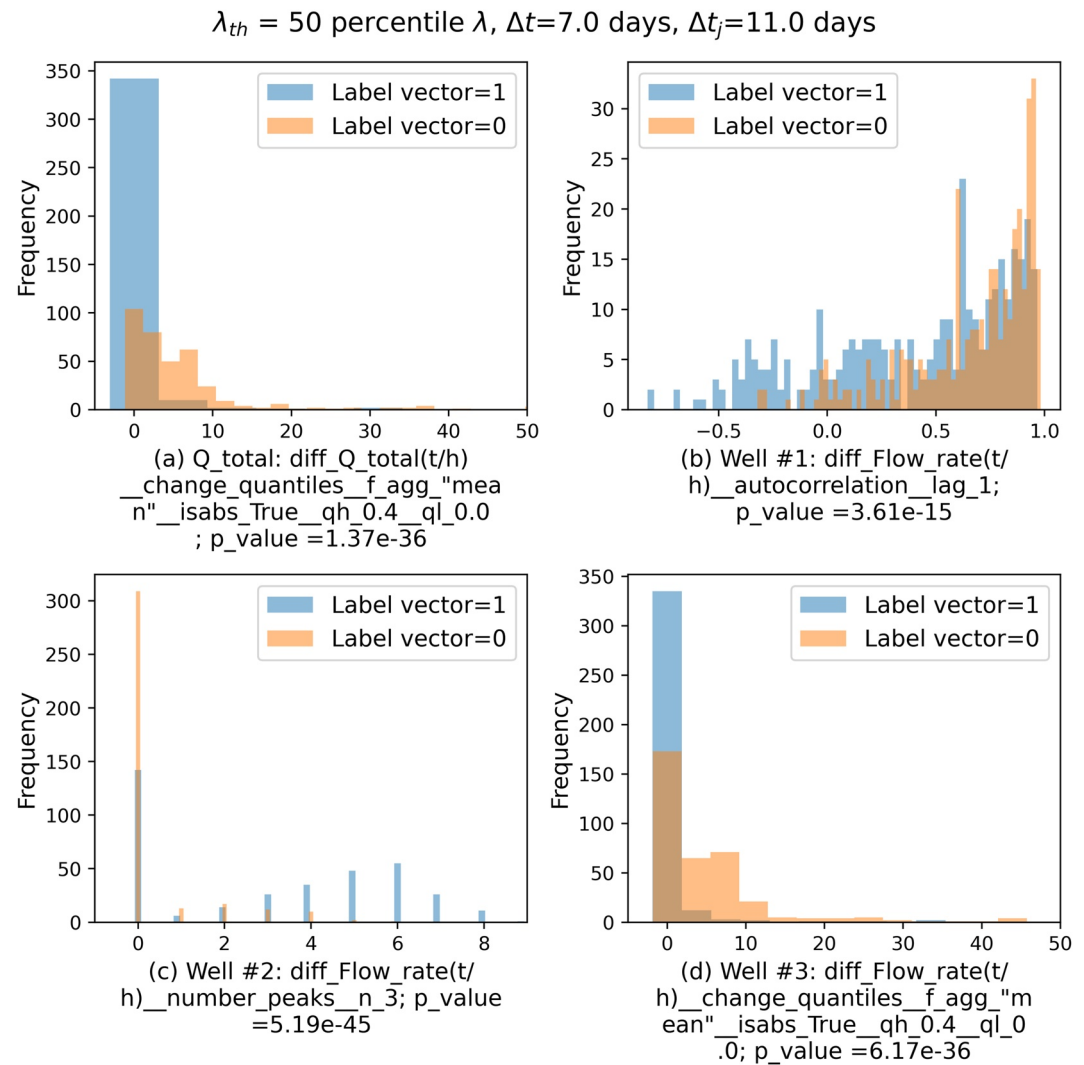


Figure 12. Significant feature distributions for the differenced injection rate ($\dot{q}$) data stream for (a) total injection rate, (b) Well #1, (c) #2, and (d) #3. Here, we are showing the most significant time series feature (lowest p -value) for the same data stream ($\dot{q}$) for different wells and well combinations. t/h (tonnes per hour) shown in the x axes of each subplot is the unit of flow rate.

explore how the injection-seismicity association shifts for different Rotokawa wells over time by comparing the pre-July and post-September 2013 responses (Figures 13 and 14). In this analysis, we only used features computed from normalized injection data, $\tilde{q}_i$, and considering two values of λ_{th} : 50th and 80th percentile. The latter value picks out associations with the highest seismicity periods.

Focusing on a specific well (e.g., Well #3) in Figures 13a–13b, we can assess the association with seismicity rate measured over different time-scales (Δt in the range 2.5–20 days). We see an increasing association (smaller aggregate p -values) between longer-term seismicity (larger Δt) and injection rate. The reduced association over periods of only a few days is best explained by the relative dominance of aleatoric variability in the microseismicity. Short-term earthquake rates, by construction, have fewer events from which to infer an underlying rate. This results in more day-to-day variability of the seismicity rate, and this in turn obscures any relationships with injection. In comparison, by averaging over several weeks, we smooth out short-term stochasticity, and any associations with injection rate, if present, become apparent.

The vertical separation between curves in Figures 13a and 13b indicates that injection into Well #2 in the pre-July 2013 phase had the largest association with microseismicity. This is true when both 50th and 80th percentile thresholds are used to separate periods of high and low rate. This result is consistent with regression modeling

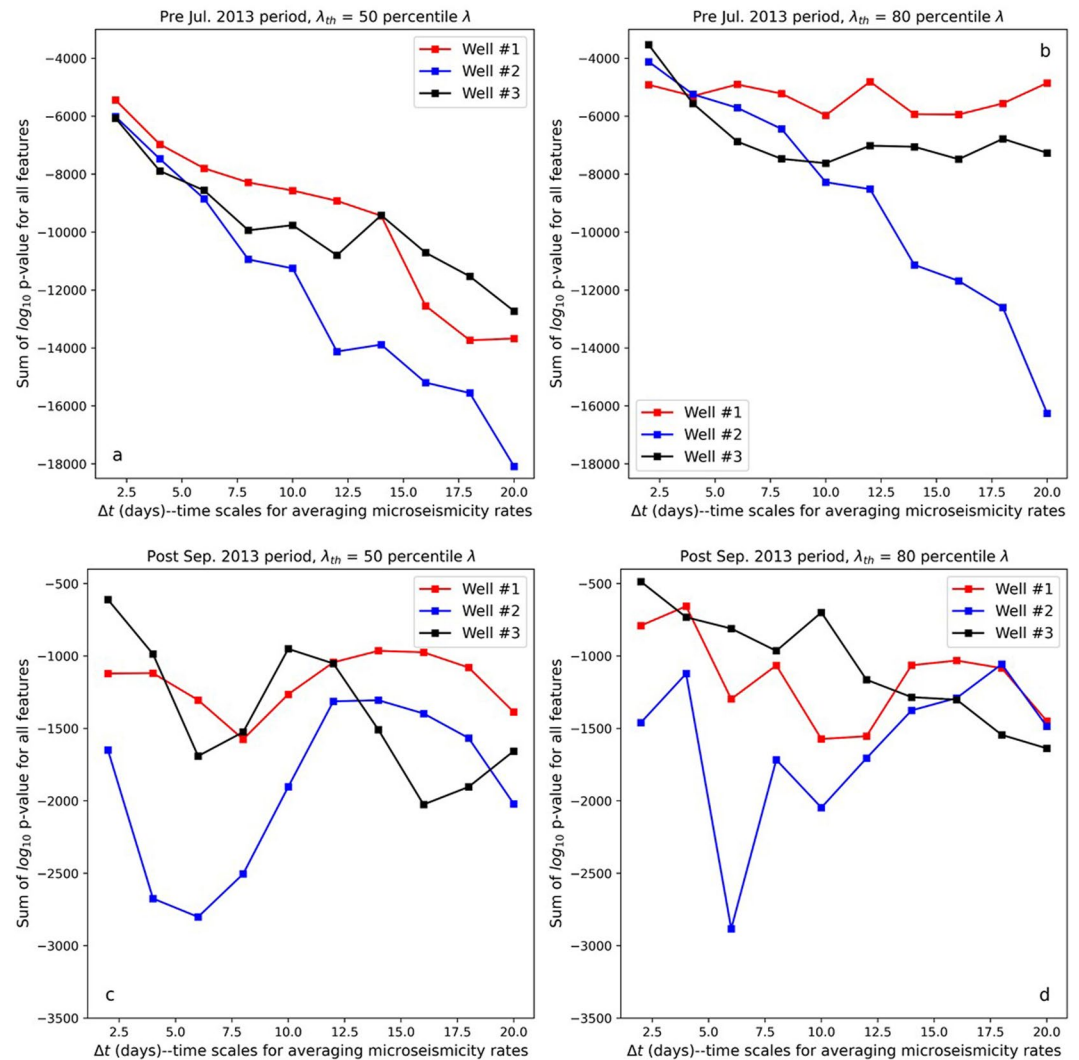


Figure 13. Aggregate association for seismicity smoothed over different Δt_i for the three different wells at Rotokawa pre-July 2013 and post September 2013 periods. (a) Using a 50th percentile threshold to separate high and low seismicity at pre July 2013 period. (b) Using an 80th percentile threshold at pre July 2013 period. (c) Using a 50th percentile threshold at post September 2013 period. (d) Using an 80th percentile threshold at post September 2013 period.

in Section 2, which assigned a larger coefficient to Well #2 when fitting these data. For $\lambda_{th} = 80$ th percentile, the association between Wells #1 and #3 and the microseismicity is reduced relative to Well #2.

In contrast, for the post-September 2013 data, associations between wells and the microseismicity rate are much less clear (higher values of $\mathcal{P}$ in general, Figures 13c and 13d). Compared with pre-July 2013 period (Figures 13a and 13b), they also do not exhibit the trend with seismicity averaging rate, which suggests there is no underlying association to be diluted by stochasticity.

To place these associations in context, we note that in the pre-July 2013 period, seismicity had been steadily increasing whereas, in the post-September 2013 period, it switched to fluctuating around a relatively constant rate. Although the regression model suggested a decreased emphasis on Well #2 injection in the post-September period (Table 1), the feature analysis here suggests that no significance should be attached to specific wells. Taken together, these findings provide further support that variations in the operational strategy achieved by rebalancing well injection volumes resulted in changes in the microseismicity behavior (rate and stability) at Rotokawa.

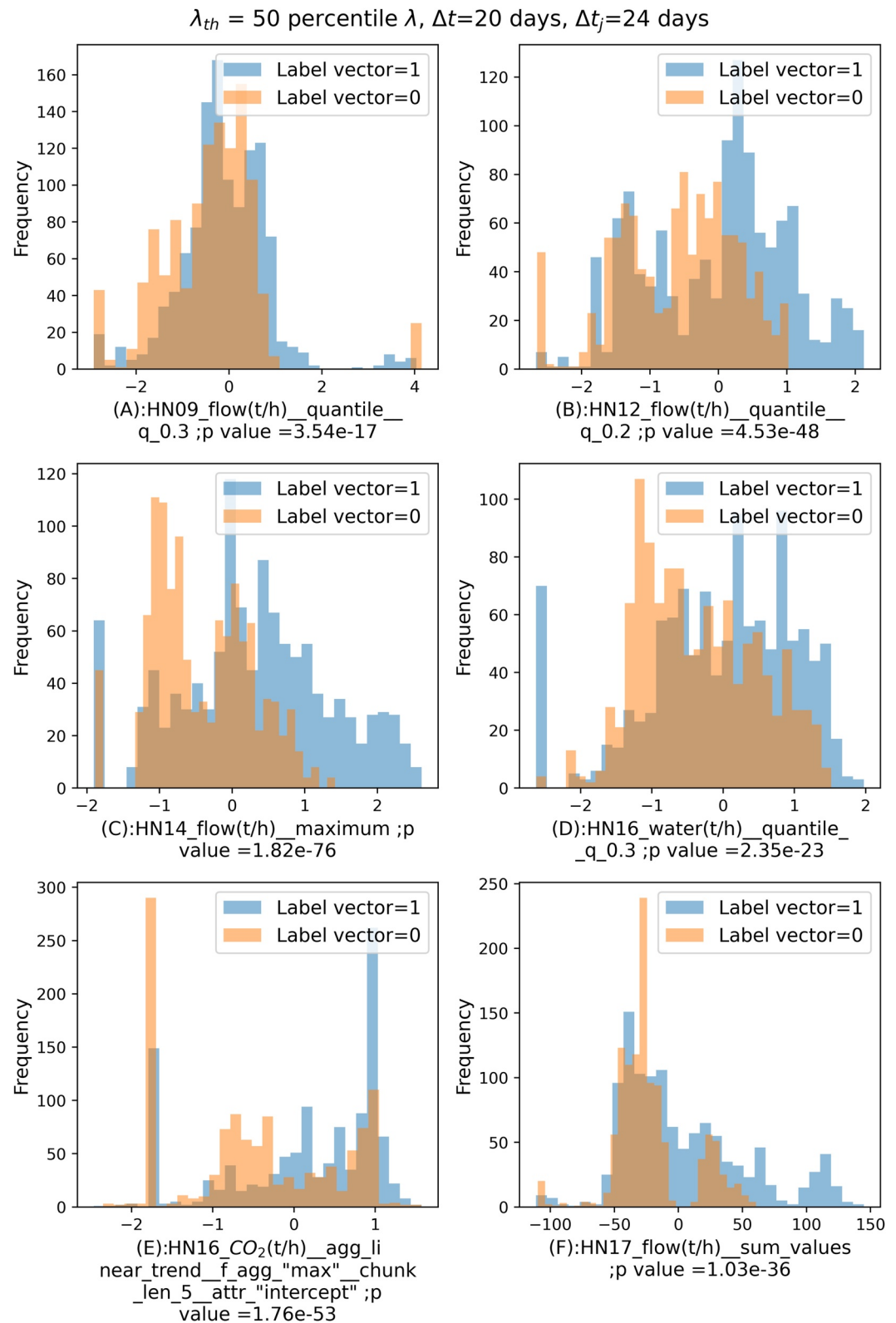


Figure 14. Significant features distributions of each individual wells of Húsmúli reinjection area. t/h (tonnes per hour) shown in the x axes of each subplot is the unit of flow rate.

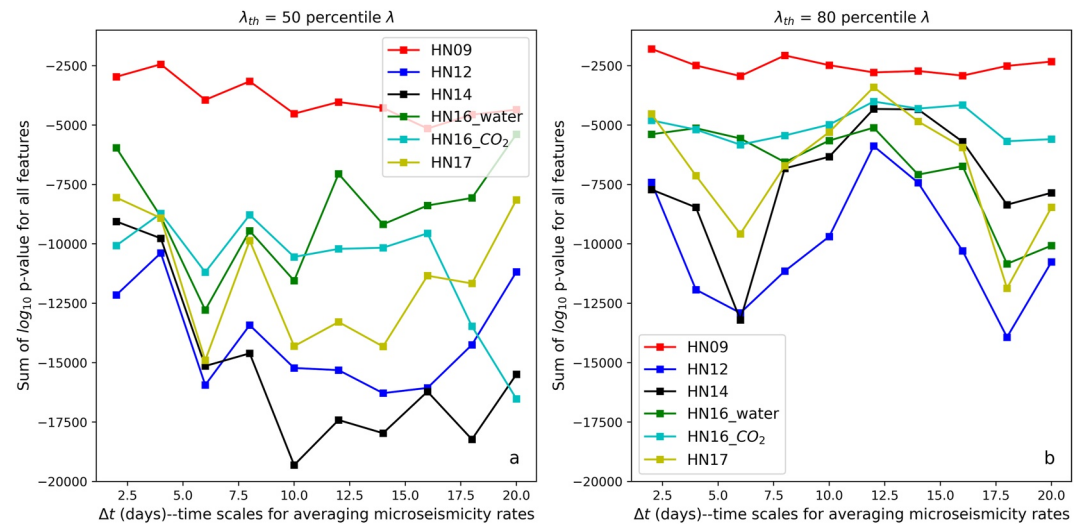


Figure 15. Aggregate association for all features under different Δt_i for Hús múli reinjection wells. (a) Using a 50th percentile threshold to separate high and low seismicity. (b) Using an 80th percentile threshold.

3.4. Microseismicity Rate Association With Hús múli Reinjection Wells

We performed a similar feature analysis for Hús múli, first plotting a selection of significant feature for individual wells to show how aspects of the injection behavior can affect the microseismicity that follows (Figure 14). An expanded selection of features is given in Supporting Information S1. As with the Hús múli regression model, we first declustered the catalog by removing all events in a 1-week window following events larger than M_L 2.5. High and low seismicity are here separated by the 50th percentile value (~ 2.1 events per week) for a rate calculated using a 20-day smoothing window ($\Delta t = 20$ days).

Analysis of the distribution of readily interpreted injection rate features (quantiles, maximum, intercept of linear regressors, sums) suggests weak associations between seismicity and individual Hús múli wells. There is a tendency for high seismicity rate (label vector 1) distributions to be slightly biased to higher injection rate measures. However, because the distributions overlap significantly, the associations are weak and, in this form, likely not suited for predictive modeling. For the explanation of the features shown in Figure 14, refer to Table 3.

We also checked aggregate associations of the individual Hús múli wells by computing the $\mathcal{P}$ (Equation 4) for each of them at different seismicity smoothing time-scales (Figure 15). First, separating low and high seismicity at the 50th percentile reveals there is perhaps a stochasticity effect that is being removed by averaging over periods of a week or more. However, at longer time-scales, this trend plateaus and is even reversed. This could suggest significant microseismicity variability tied to injection at one-to-two weeks time scales, which would tend to discount longer-term mechanisms such as accumulated thermal effects.

Of interest is the clear separation in terms of p -value for the six well curves in Figure 15a. From this we identify HN-12, HN-17, and HN-14 as the wells most strongly associated with one-to-two week seismicity. The HN-16 CO_2 injection zone has a stronger association with long term seismicity rates. As a fluid, CO_2 is more compressible than water and may therefore take longer to build pressures needed to trigger seismicity. When $\lambda_{th} = 80$ th percentile, the associations are generally weakened. This suggests that relationships between individual wells and microseismicity are somewhat dependent on the choice of λ_{th} and therefore could be spurious. Any further interpretation should be done with caution. Well HN-09 does not appear sensitive to microseismicity in the Hús múli area, independent of the threshold used to classify seismicity.

These general observations are roughly consistent with the findings of the regression analysis in Section 2.2, which highlighted HN-12 and HN-17 as particularly sensitive to microseismicity, and HN-14 and HN-16 (CO_2) less so. A caveat is that the time series feature analysis only shows an association between microseismicity and HN-16 (CO_2) when λ_{th} is the 50th percentile, but not the highest seismicity at $\lambda_{th} = 80$ th percentile. In contrast, the regression analysis fits all levels of seismicity simultaneously because, by construction, it does not arbitrarily

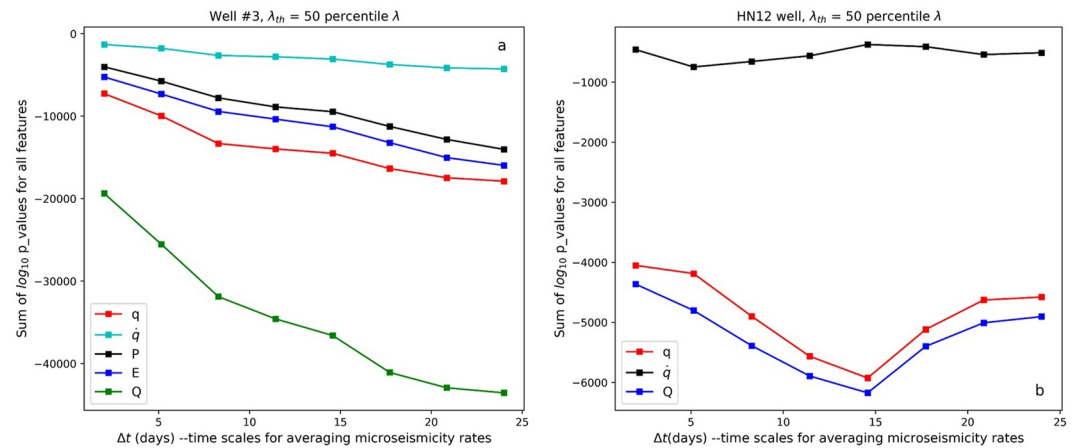


Figure 16. Aggregate association comparison of all features under different data type (flow rate q , flow rate derivative $\dot{q}$, cumulative injection volume Q , pore pressure P , and hydraulic work rate E) of Rotokawa Well #3 and Húsmúli HN-12 well under different short-term Δt .

categorize high and low seismicity. Therefore, although the analyses are not directly comparable, they at least do not contradict each other.

3.5. Feature Analysis of Different Data Types

In this section, we compare the relative importance of the different data types derived for Well #3 at Rotokawa geothermal field. We chose this well because it has a long record of uninterrupted data (Figure 2), some association with seismicity in the pre-July 2013 period (Figure 13a), and had available wellhead pressure data. We also evaluated relative significance of data streams (q , $\dot{q}$, Q) for HN-12 well at Húsmúli reinjection area, which has been identified as one of the wells that associated with seismic events (Figure 15). Following the approach of Section 3.4, we compute aggregate associations for the different normalized data types and compare their values across different smoothed earthquake rates.

Like injection rate, an increased wellhead pressure could plausibly be associated with microseismicity. For example, high injection pressure could increase the pressurized volume or pressurization rate, both of which link back to stability conditions on faults and earthquake triggering (Catalli et al., 2016; Ciardo & Rinaldi, 2022; Dempsey & Riffault, 2019; Hubbert & Rubey, 1959). However, Figure 16a shows that at Rotokawa, the pressure (P) actually has a weaker aggregate association than the flow rate q . Hydraulic work rate, E , which combines both pressure and injection rate, has an association that is approximately the average of its constituent parts. Further work is required to make full use of wellhead pressure data, especially because it does not necessarily correlate with downhole pressure conditions if the fluid is not in a liquid state. Cumulative injection volume, Q , shows the largest association of all data types. This likely reflects its monotonically increasing characteristic, which correlates well with building seismicity in the pre-July 2013 period (Figure 5a).

Figure 16b also shows the cumulative injection volume and injection rate are the two most significant data streams with seismicity rates at Húsmúli HN-12 well. Moreover, the P value decreases with the increase of Δt , and obtain the minimum value when $\Delta t = 2$ weeks. As noted earlier, this potentially reflects significant seismicity-injection variability at the 2-week timescales.

4. Discussion

4.1. Physical Interpretations of Well Association With Microseismicity

4.1.1. Rotokawa Injection Wells

Our analysis of the Rotokawa injection wells suggests that cumulative injection volume into Well #3 adjacent to the CFF is correlated with microseismicity. This is in agreement with previous studies (Sewell, Cumming, et al., 2015; Sherburn, Bannister, et al., 2015; Sherburn, Sewell, et al., 2015) of Rotokawa seismicity in the period 2008–2012. These studies showed that microseismicity clustered at the approximate depth of the permeable zone

of Well #3 and the surrounding area of the CFF, and were assumed to represent reactivation of this structure. The CFF has been identified as a major compartment controlling the microseismicity distribution in the Rotokawa reservoir. It acts as a conduit for the fluid migration along strike but also as a buffer for slowing the fluid migration across its strike (Addison et al., 2017; Sherburn, Bannister, et al., 2015; Sherburn, Sewell, et al., 2015), promoting pressure build-up.

Both statistical inference models indicate that injection into Rotokawa Well #2 is associated with microseismicity in the pre-July 2013 period, but that the association is reduced (Table 1) or absent (Figures 13c and 13d) in the post-September 2013 period. This is in spite of relatively higher injection after the operational change. One explanation is that reservoir rock in the region covering these three injection wells and CFF and IFF faults achieved some sort of equilibrium in the physical processes driving seismicity. Whether this is stabilization of fluid pressure, or attendant contributions by thermal and poroelastic stresses is unknown. However, the change in injectate to a higher temperature fluid is suggestive.

At Rotokawa, during the pre-July 2013 period, Well #2 was injected with 40°C condensate. This increased in the post-September 2013 period when a 130°C brine was instead injected, similar to Well #1 and Well #3 injection temperature. Therefore, one explanation for this difference is that induced thermal stresses were smaller after the change in injection, and this reduced their impact on microseismicity. Rutqvist et al. (2016a) have shown that thermal contraction of reservoir rock induces stress changes that contribute to microseismicity around injection wells at the Geysers EGS. Sewell, Cumming, et al. (2015) have argued that large temperature differences trigger microseismicity in the thermally perturbed near-borehole region at Rotokawa. If this is the case, then a smaller temperature difference could plausibly account for a reduced association with Well #2 injection.

An alternative explanation concerns pressurization (and perhaps poroelastic stress) re-equilibrium and then reduction in certain seismogenic reservoir compartments. This derives from Hopp et al. (2020) who adopted the interpretation of Wallis et al. (2013) that the strike of IFF separates wells #1 and Well #3 in one compartment from Well #2 in another. Wallis et al. (2013) describe the IFF as offsetting by 250–350m vertical stratigraphic separation between Well #2 and the rest of southern wells (e.g., Well #1 and #3). This indicates that Well #2 lies to the east of IFF whereas Well #1 and Well #3 locate to the west (Hopp et al., 2020). Thus, the stabilization of microseismicity in the post-September 2013 period might be explained by a reduced rate of pressurization in the reservoir compartment bounded by the CFF and IFF, while the increased pressurization rate in the Well #2 compartment may be somewhat offset by a shorter history of injection here (and, hence, lower overall pressure).

Both hypotheses described here are likely difficult to confirm without extensive pressure and temperature monitoring data in the relevant reservoir compartments, accompanied by future shifts in injection strategy that modulate pressure or thermo-poroelastic effects.

4.1.2. Húsmúli Injection Wells

For the Húsmúli analysis, we identified HN-12, HN-17 and HN-16 CO₂ injection feedzones as associated with microseismicity. One reason for the sensitivity of these particular wells is that they are drilled into active faults, or are near to faults that are becoming reactivated by stress perturbations associated with injection. This could include stresses generated both from thermal cooling and stress transfer from other microseismicity events (Ágústsson et al., 2015; Gunnarsson, 2011, 2013).

The three wells most associated with seismicity (HN12, HN16, and HN17) are located on the Eastern part of the reinjection zone, and target NE-SW faults identified for their ability to carry fluids (Ratouis et al., 2022; Tomasdóttir, 2018). The wells also interact with E-W striking faults associated with a slower fluid propagation toward the North and that act as a trap for the fluids. This build-up and longer residence time of the injected fluid could explain an increased association between activity in these wells and the observed microseismicity. HN-16 well CO₂ injection zone is associated with long-term seismicity rates (Figure 15a), which might be explained as the effect of CO₂ compressibility that requires a longer time to pressurize the faults. However, this also needs further investigation in the following studies to test this hypothesis.

On the other hand, wells HN09 and HN14 are located in the Western part of the reinjection area. HN09 is the deepest of the injection wells in Húsmúli, reaching into the altered crystalline basement (Ritz et al., 2021). HN14 lies on the Western-most part of the field and crosses paths with a NE-SW fault that seems to segment the reservoir and create a permeability boundary. This well has the lowest permeability of the five in Húsmúli

(Gunnarsson, 2011), which might result in a larger pressure build-up (and hence higher or more responsive seismicity rate) compared with a higher permeability formation under the same injection rate/volume. Its sensitivity to microseismicity could be linked to spatially clustered events happening between 2016 and 2018 (Figure 3c), which could correspond to the periods of higher rate seismicity (Figure 7). Surface movement and deformation have been observed near well HN14 since the start of reinjection (Bessason et al., 2012; Juncu et al., 2020) and this may be influenced by fracture slip nearby to this injector (Bessason et al., 2012; Gunnarsson, 2013).

The seismicity in the area could also be associated with large-scale poroelastic changes due to nearby steam production. Being located at the edge of the Hellisheiði production region, Húsmúli could be then affected by the large stress gradient caused by the subsidence induced by fluid production. This would create a long-term trend in seismicity rate that could not be directly associated with the reinjection wells.

4.2. No Evidence of Well Throttling Triggering Seismicity

Segall and Lu (2015) have proposed that rapid cycling of injection rates can induce transient undrained poroelastic stresses in reservoirs and that these can temporarily elevate the earthquake triggering rate. The seismicity transient is proposed to be short-lived, occurring shortly after the injection rate change.

Shirzaei et al. (2016) investigated this model and found evidence of elevated seismic activity in eastern Texas following a rapid decline in injection at wells there, which might partly be because of the poroelastic relaxation of compressive stress on a fault. Langenbruch and Zoback (2017) also observed aftershock sequences related to relatively large magnitude events after injection rate cuts, and which delayed the decay of the amount of induced seismicity in Oklahoma.

We have tested this hypothesis at the Rotokawa field by looking for associations between short-term seismicity rate (focusing on averages between 2 and 8 days) and the derivative of the injection rate, which will tend to have large values when rapidly shutting in or reopening wells. Figure 16 shows that the aggregate association ($\mathcal{P}$) of $\dot{q}$ has a smaller magnitude than q , indicating a reduced association with injection rate changes. Furthermore, there is no evidence of an increasing association with very short seismicity time-scales, where such effects might appear more prominent.

Thus, we find no evidence that rapidly throttling injection well rates, either by shutting-in or restarting, produces statistically elevated earthquake rates. However, we acknowledge that the absence of evidence is not evidence of absence. It may be that the data we have used are not sensitive enough to detect a poroelastic transient, particularly given that we had earlier indicated that short-term earthquake rates are contaminated by stochasticity. The corollary is those poroelastic transients, if present, are too small to be of concern. It may also be the case that Rotokawa is not as poroelastically sensitive as other reservoirs where this mechanism may prevail, possibly because pressure oscillations in permeable geothermal reservoirs are too small to induce a strong undrained poroelastic response. For fault systems that are far from critical state, injection-induced stress changes might be too small to trigger earthquakes, although this is unlikely to be the case in New Zealand and Iceland geothermal settings that are tectonically active with high natural seismicity. Finally, in this study, we have not tested whether throttling of production wells might be associated with elevated seismicity.

Note, this section does not consider or test for the presence of longer-term poroelastic stressing attributed, for example, to long-term pressurization of reservoir compartments. As noted earlier, the operational shift at Rotokawa in mid-2013 produced a seismicity response that might, in part, be explained by long-term poroelastic (and other) stress effects.

4.3. Limitations and Future Directions for Data Driven Modeling

Our goal in applying the time-series feature engineering method was to find unseen patterns or new and significant features linking injection and microseismicity. The overall concept is to use these time-series features to evaluate the relative sensitivity of injection wells to microseismicity at geothermal reservoirs, and in this way assist operational management, and possibly forecasting of unwanted effects.

In this instance, we are not convinced that time-series feature engineering has provided a better physical understanding of the system beyond simpler regression models. For example, both methods could associate the injection rate switch in August 2013 at Rotokawa with a change in microseismicity behavior. Furthermore, both methods identified the same wells as having an elevated association with microseismicity at Rotokawa and Húsmúli sites. Clearly, we should be cautious to not apply unnecessarily complex analysis methods when simple tools can achieve good results. Mordensky et al. (2023) applied several different ML methods to evaluate the geothermal energy assessments, which also found that complex ML methods (e.g., artificial neural networks) may not always outperform simple approaches (e.g., simple nonlinear algorithm).

At the same time, we note that this study has only tested one method (time series feature engineering) under a limited range of contexts (two geothermal fields, supervised learning). ML is a broad field with a diverse range of algorithms and modes of interrogation. For example, although both methods highlight the microseismicity change when directly informed of the injection rate switch, an unsupervised ML approach might detect this change, and other more subtle changes, without being directly led. This was the approach of Manley et al. (2021) who used anomaly detection to identify changes in patterns of volcanic state that may herald onset or the end of volcanic activity. Thus, a further test of our methods would be to try and identify the September 2013 change in the full Rotokawa data set while withholding from the algorithm knowledge of the change.

Neither modeling approach was able to satisfactorily eliminate the effect of aftershocks or other medium-term increases above the average rate, for example, swarms. Capturing these effects may be better achieved through physics-informed models that can build in their causative processes. It could also be useful to adopt different operational data streams, including downhole pressure data and or higher resolution catalogs subdivided into several field areas (e.g., as in Hopp et al., 2020). Using more advanced declustering methods on the catalogs to keep the microseismicity signal free of aftershocks might also prove helpful, providing it is done correctly and does not also remove underlying physical processes (e.g., Mizrahi et al., 2021). Integration of physical and data-driven models may further improve our understanding of induced microseismicity in geothermal reservoirs. Alternatively, other aspects of the event catalog might be more diagnostic than simply the seismicity rate. Future studies could investigate whether radiated seismic energy, moment rate or *b*-value changes can be linked to individual wells.

Finally, future applications of this work could include testing whether there are sufficient associations in the data to develop classification models and regression models with predictive capability.

5. Conclusions

We have investigated the relationship between injection and microseismicity data in the Rotokawa geothermal field (for the 2012–2015 period) and Húsmúli reinjection area (between 2015 and 2020) using data-driven modeling approaches. This includes a simple linear regression analysis of injection data and a more sophisticated ML method based on time-series feature engineering. Our goal was to infer associations and patterns between the complex injection schedule involving multiple wells and the time-varying seismicity. Our main conclusions are summarized below:

1. Based on the 2012–2015 data, there is an observed correlation between injection and long-term (i.e., $\Delta t = 20$ days) microseismicity rates in the Rotokawa geothermal field. The short-term earthquake rate is likely to be dominated by randomness when inferring rates from small event counts, and this obscures any association with changes in injection. Most seismicity at Rotokawa is smaller than magnitude 2 and these small events are unlikely to be felt outside the immediate vicinity.
2. For the Rotokawa geothermal field, both regression analysis and time series feature engineering has associated injection rates into Well #2 with elevated microseismicity prior to a field operation change (i.e., slight modification in the injection well configuration to adapt to changes in well injectivity) in August 2013. Negative and positive associations of the different wells in the field might reflect underlying connectivity to major seismogenic structures in the reservoir, in particular, the CFF and IFF. However, in the post-September 2013 period, the association between microseismicity rate and Well #2 and #3 becomes much weaker, and this is presumably linked to changes in reservoir operation.
3. Based on a similar analysis of Húsmúli wells, HN-17, HN-12, and HN-16 (CO₂ injection) are the three most significant injectors associated with microseismicity.

4. We find no evidence that rapid injection rate changes trigger elevated seismicity. Instead, cumulative injection volume and injection rate appear to be the most informative data for field microseismicity.
5. The ML method provided only marginal benefits over a simple regression analysis of the injection and microseismicity data. In this instance, it is not obvious that the approach has justified the additional time and complexity to implement it, beyond the evaluation of conclusion (4) above.

Data Availability Statement

The Húsmúli reinjection data and raw earthquake catalog are provided by Reykjavík Energy and the Icelandic Meteorological Office (IMO), Iceland, and are freely available at <https://doi.org/10.5281/zenodo.7041552>. Rotokawa injection data are provided by the Rotokawa Joint Venture, New Zealand, and are freely available at <https://doi.org/10.5281/zenodo.7592213>. The raw earthquake data for Rotokawa geothermal field were obtained from Hopp et al. (2020) shared at <https://doi.org/10.17605/OSF.IO/C2M6U>.

Acknowledgments

The authors thank Manuel Rivera at Mercury for discussions that improved our understanding of the Rotokawa geothermal field. P. Yu acknowledges Dr. Alberto Ardid for assistance producing Figures 1 and 3. D. Dempsey acknowledges funding from MBIE Endeavour “Empowering Geothermal” research grant. The authors wish to acknowledge the Centre for eResearch at the University of Auckland for their help in facilitating this research <http://www.eresearch.auckland.ac.nz>. A. P. Rinaldi and V. A. Ritz are funded through the COSEISMIQ and DEEP projects. The DEEP project (<http://deepgeothermal.org>) is funded through the ERANET Cofund GEOTHERMICA (Project 200320-4001) from the European Commission. The DEEP project benefits from an exploration subsidy of the Swiss federal office of energy for the EGS geothermal project in Haute-Sorne, canton of Jura (contract MF-021-GEO-ERK), which is gratefully acknowledged. The COSEISMIQ project (<http://www.coseismiq.ethz.ch>) is funded through the ERANET Cofund GEOTHERMICA (Project 731117) from the European Commission, and Geological Survey Ireland (GSI; Project 170167-44011). P. Yu acknowledges support from US Department of Energy Grand DE-EE008763. The authors thank editor Douglas Schmidt and reviewers Ben Holtzman and Chet Hopp for constructive reviews that improved this manuscript. Open access publishing facilitated by University of Canterbury, as part of the Wiley - University of Canterbury agreement via the Council of Australian University Librarians.

References

- Addison, S., Azwar, L., Clearwater, J., Hernandez, D., Mountain, B., Blair, A., & Siratovich, P. (2017). Improving the conceptual understanding through a recent injection of 200 GBq of Iodine - 125 at the Rotokawa geothermal field, New Zealand. *Forty-Second Workshop on Geothermal Reservoir Engineering Stanford University, February* (pp. 1–12).
- Ágústsson, K., Kristjánssdóttir, S., Flóvenz, Ó. G., & Guðmundsson, Ó. (2015). Induced seismic activity during drilling of injection wells at the Hellisheiði Power Plant, SW Iceland. *World Geothermal Congress 2015, April* (p. 10).
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/tac.1974.1100705>
- Amini, A., & Eberhardt, E. (2021). Machine-learning analysis of factors influencing induced seismicity susceptibility in the Montney Play Area, Northeastern British Columbia (NTS 093P, 094A, B, G, H). *Geoscience BC Summary of Activities 2020: Energy and Water, Geoscience BC, Report 2021-02* (pp. 45–56). Retrieved from <http://geosciencebc.com/updates/summary-of-activities/>
- Amirlatif, A., KC, B., Adibifard, M., Vahedifard, F., & Ghazanfari, E. (2021). Data analytics to investigate the cohort of injection wells with earthquakes in Oklahoma. *Earthquake Spectra*, 37(1_suppl), 1652–1672. <https://doi.org/10.1177/8755293021989725>
- Ardid, A., Dempsey, D., Caudron, C., & Cronin, S. (2022). Seismic precursors to the Whakaari 2019 phreatic eruption are transferable to other eruptions and volcanoes. *Nature Communications*, 13(1), 2002. <https://doi.org/10.1038/s41467-022-29681-y>
- Bachmann, C. E., Wiemer, S., Goertz-Allmann, B. P., & Woessner, J. (2012). Influence of pore-pressure on the event-size distribution of induced earthquakes. *Geophysical Research Letters*, 39(9), 9302. <https://doi.org/10.1029/2012GL051480>
- Bannister, S., Sherburn, S., & Bourguignon, S. (2016). Earthquake swarm activity highlights crustal faulting associated with the Waimangu-Rotomahana-Mt Tarawera geothermal field, Taupo Volcanic Zone. *Journal of Volcanology and Geothermal Research*, 314, 49–56. <https://doi.org/10.1016/J.JVOLGEORES.2015.07.024>
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4), 1165–1188. <https://doi.org/10.1214/AOS/1013699998>
- Bergen, K. J., Johnson, P. A., De Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363(6433). https://doi.org/10.1126/SCIENCE.AAU0323/ASSET/1DB6BFB4-5345-48FB-BA0E-AB037985D1E6/ASSETS/GRAPHIC/363_AAU0323_FA.JPEG
- Bessason, B., Ólafsson, E. H., Gunnarsson, G., Flóvenz, Ó. G., Jakobsdóttir, S. S., Björnsson, S., & Árnadóttir, P. (2012). *Verklag vegna örvaðrar skjálftavirkni í jarðhitakerfum*. Orkuveita Reykjavíkur.
- Bialas, J., Oommen, T., Rebbapragada, U., & Levin, E. (2016). Object-based classification of earthquake damage from high-resolution optical imagery using machine learning. *Journal of Applied Remote Sensing*, 10(3), 036025. <https://doi.org/10.1117/1.JRS.10.036025>
- Bibby, H. M., Caldwell, T. G., Davey, F. J., & Webb, T. H. (1995). Geophysical evidence on the structure of the Taupo Volcanic Zone and its hydrothermal circulation. *Journal of Volcanology and Geothermal Research*, 68(1–3), 29–58. [https://doi.org/10.1016/0377-0273\(95\)00007-H](https://doi.org/10.1016/0377-0273(95)00007-H)
- Broccardo, M., Mignan, A., Wiemer, S., Stojadinovic, B., & Giardini, D. (2017). Hierarchical Bayesian modeling of fluid-induced seismicity. *Geophysical Research Letters*, 44(22), 11–357. <https://doi.org/10.1002/2017gl075251>
- Brodsky, E. E., & Lajoie, L. J. (2013). Anthropogenic seismicity rates and operational parameters at the Salton Sea geothermal field. *Science*, 342(6165), 1446–1546. <https://doi.org/10.1126/science.342.6165.1446-a>
- Bryan, C. J., Sherburn, S., Bibby, H. M., Bannister, S. C., & Hurst, A. W. (1999). Shallow seismicity of the central Taupo Volcanic Zone, New Zealand: Its distribution and nature. *New Zealand Journal of Geology and Geophysics*, 42(4), 533–542. <https://doi.org/10.1080/00288306.1999.9514859>
- Candela, T., van der Veer, E. F., & Fokker, P. A. (2018). On the importance of thermo-elastic stressing in injection-induced earthquakes. *Rock Mechanics and Rock Engineering*, 51(12), 3925–3936. <https://doi.org/10.1007/S00603-018-1619-6/FIGURES/10>
- Cardiff, M., Lim, D. D., Patterson, J. R., Akerley, J., Spielman, P., Lopeman, J., et al. (2018). Geothermal production and reduced seismicity: Correlation and proposed mechanism. *Earth and Planetary Science Letters*, 482, 470–477. <https://doi.org/10.1016/J.EPSL.2017.11.037>
- Catali, F., Rinaldi, A. P., Gischig, V., Nespole, M., & Wiemer, S. (2016). The importance of earthquake interactions for injection-induced seismicity: Retrospective modeling of the Basel Enhanced Geothermal System. *Geophysical Research Letters*, 43(10), 4992–4999. <https://doi.org/10.1002/2016GL068932/FORMAT/PDF>
- Chen, H., Meng, X., Niu, F., Tang, Y., Yin, C., & Wu, F. (2018). Microseismic monitoring of stimulating shale gas reservoir in SW China: 2. Spatial clustering controlled by the pre-existing faults and fractures. *Journal of Geophysical Research: Solid Earth*, 123(2), 1659–1672. <https://doi.org/10.1002/2017JB014491/FORMAT/PDF>
- Christ, M., Kempa-Liehr, A. W., & Feindt, M. (2016). Distributed and parallel time series feature extraction for industrial big data applications. Retrieved from <https://arxiv.org/abs/1610.07717v3>
- Ciardo, F., & Rinaldi, A. P. (2022). Impact of injection rate ramp-up on nucleation and arrest of dynamic fault slip. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 8(1), 1–27. <https://doi.org/10.1007/s40948-021-00336-4>

- Cipolla, C., Weng, X., Mack, M., Ganguly, U., Gu, H., Kresse, O., & Cohen, C. (2011). Integrating microseismic mapping and complex fracture modeling to characterize fracture complexity. *Society of Petroleum Engineers—SPE/EAGE European Unconventional Resources Conference and Exhibition 2012* (pp. 27–48). <https://doi.org/10.2118/140185-MS>
- Clarke, D., Townend, J., Savage, M. K., & Bannister, S. (2009). Seismicity in the Rotorua and Kawerau geothermal systems, Taupo Volcanic Zone, New Zealand, based on improved velocity models and cross-correlation measurements. *Journal of Volcanology and Geothermal Research*, 180(1), 50–66. <https://doi.org/10.1016/j.jvolgeores.2008.11.004>
- Corbi, F., Bedford, J., Sandri, L., Funicello, F., Gualandi, A., & Rosenau, M. (2020). Predicting imminence of analog megathrust earthquakes with machine learning: Implications for monitoring subduction zones. *Geophysical Research Letters*, 47(7), e2019GL086615. <https://doi.org/10.1029/2019GL086615>
- Deichmann, N., & Giardini, D. (2009). Earthquakes induced by the stimulation of an enhanced geothermal system below Basel (Switzerland). *Seismological Research Letters*, 80(5), 784–798. <https://doi.org/10.1785/GSSRL.80.5.784>
- Dempsey, D. E., Cronin, S. J., Mei, S., & Kempa-Liehr, A. W. (2020). Automatic precursor recognition and real-time forecasting of sudden explosive volcanic eruptions at Whakaari, New Zealand. *Nature Communications*, 11(1), 1–8. <https://doi.org/10.1038/s41467-020-17375-2>
- Dempsey, D. E., & Riffault, J. (2019). Response of induced seismicity to injection rate reduction: Models of delay, decay, quiescence, recovery, and Oklahoma. *Water Resources Research*, 55(1), 656–681. <https://doi.org/10.1029/2018WR023587>
- Dempsey, D. E., & Suckale, J. (2016). Collective properties of injection-induced earthquake sequences: 1. Model description and directivity bias. *Journal of Geophysical Research: Solid Earth*, 121(5), 3609–3637. <https://doi.org/10.1002/2015JB012550>
- Dempsey, D. E., & Suckale, J. (2017). Physics-based forecasting of induced seismicity at Groningen gas field, the Netherlands. *Geophysical Research Letters*, 44(15), 7773–7782. <https://doi.org/10.1002/2017GL073878>
- Deng, K., Liu, Y., & Chen, X. (2020). Correlation between poroelastic stress perturbation and multidisposal wells induced earthquake sequence in Cushing, Oklahoma. *Geophysical Research Letters*, 47(20), e2020GL089366. <https://doi.org/10.1029/2020GL089366>
- DeVries, P. M. R., Viégas, F., Wattenberg, M., & Meade, B. J. (2018). Deep learning of aftershock patterns following large earthquakes. *Nature*, 560(7720), 632–634. <https://doi.org/10.1038/s41586-018-0438-y>
- Di Stefano, R., Aldersons, F., Kissling, E., Baccheschi, P., Chiarabba, C., & Giardini, D. (2006). Automatic seismic phase picking and consistent observation error assessment: Application to the Italian seismicity. *Geophysical Journal International*, 165(1), 121–134. https://doi.org/10.1111/j.1365-246X.2005.02799.x/3/M_165-1-121-EQ013.JPEG
- Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, 341(6142), 1225942. <https://doi.org/10.1126/science.1225942>
- Ellsworth, W. L., Giardini, D., Townend, J., Ge, S., & Shimamoto, T. (2019). Triggering of the Pohang, Korea, earthquake (Mw 5.5) by enhanced geothermal system stimulation. *Seismological Research Letters*, 90(5), 1844–1858. <https://doi.org/10.1785/0220190102>
- Frank, J., Rebbapragada, U., Bialas, J., Oommen, T., & Havens, T. C. (2017). Effect of label noise on the machine-learned classification of earthquake damage. *Remote Sensing*, 9(8), 803. <https://doi.org/10.3390/RS9080803>
- Franzson, H., Kristjánsson, B. R., Gunnarsson, G., Björnsson, G., Hjartarson, A., Steingrímsson, B., et al. (2005). The Hengill-Hellisheiði geothermal field. Development of a conceptual geothermal model. *Proceedings World Geothermal Congress 2005, April 2005* (pp. 24–29).
- Garwood, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika*, 28(3/4), 437–442. <https://doi.org/10.2307/2333958>
- Gunnarsson, G. (2011). Mastering reinjection in the Hellisheiði field, SW-Iceland: A story of successes and failures. In *Thirty-Sixth Workshop on Geothermal Reservoir Engineering*. Stanford University.
- Gunnarsson, G. (2013). Temperature dependent injectivity and induced seismicity—Managing reinjection in the Hellisheiði field, SW-Iceland. *GRC Transactions*, 37, 1020–1025.
- Gunnarsson, G., Kristjánsson, B. R., Gunnarsson, I., & Júlíusson, B. M. (2015). Reinjection into a fractured reservoir-induced seismicity and other challenges in operating reinjection wells in the Hellisheiði field, SW-Iceland. *Proceedings World Geothermal Congress* (pp. 19–25).
- He, M., Li, Q., & Li, X. (2020). Injection-induced seismic risk management using machine learning methodology—A perspective study. *Frontiers of Earth Science*, 227. <https://doi.org/10.3389/FEART.2020.00227>
- Healy, J. H., Rubey, W. W., Griggs, D. T., & Raleigh, C. B. (1968). The Denver earthquakes. *Science*, 161(3848), 1301–1310. <https://doi.org/10.1126/science.161.3848.1301>
- Hofmann, H., Zimmermann, G., Zang, A., & Min, K. B. (2018). Cyclic soft stimulation (CSS): A new fluid injection protocol and traffic light system to mitigate seismic risks of hydraulic stimulation treatments. *Geothermal Energy*, 6(1), 27. <https://doi.org/10.1186/s40517-018-0114-3>
- Holtzman, B. K., Paté, A., Paisley, J., Waldhauser, F., & Repetto, D. (2018). Machine learning reveals cyclic changes in seismic source spectra in Geysers geothermal field. *Science Advances*, 4(5), 1–8. <https://doi.org/10.1126/sciadv.aao2929>
- Hopp, C., Sewell, S., Mroczek, S., Savage, M., & Townend, J. (2019). Seismic response to injection well stimulation in a high-temperature, high-permeability reservoir. *Geochemistry, Geophysics, Geosystems*, 20(6), 2848–2871. <https://doi.org/10.1029/2019GC008243>
- Hopp, C., Sewell, S., Mroczek, S., Savage, M., & Townend, J. (2020). Seismic response to evolving injection at the Rotokawa geothermal field, New Zealand. *Geothermics*, 85, 101750. <https://doi.org/10.1016/j.geothermics.2019.101750>
- Hubbert King, M., & Rubey, W. W. (1959). Role of fluid pressure in mechanics of overthrust faulting: I. Mechanics of fluid-filled porous solids and its application to overthrust faulting. *Geological Society of America Bulletin*, 70(2), 115–166.
- Hulbert, C., Rouet-Leduc, B., Jolivet, R., & Johnson, P. A. (2020). An exponential build-up in seismic energy suggests a months-long nucleation of slow slip in Cascadia. *Nature Communications*, 11(1), 1–8. <https://doi.org/10.1038/s41467-020-17754-9>
- Illsley-Kemp, F., Barker, S. J., Wilson, C. J. N., Chamberlain, C. J., Hreinsdóttir, S., Ellis, S., et al. (2021). Volcanic Unrest at Taupō Volcano in 2019: Causes, mechanisms and implications. *Geochemistry, Geophysics, Geosystems*, 22(6), e2021GC009803. <https://doi.org/10.1029/2021GC009803>
- Juncu, D., Árnadóttir, T., Geirsson, H., Guðmundsson, G. B., Lund, B., Gunnarsson, G., et al. (2020). Injection-induced surface deformation and seismicity at the Hellisheiði geothermal field, Iceland. *Journal of Volcanology and Geothermal Research*, 391, 106337. <https://doi.org/10.1016/j.jvolgeores.2018.03.019>
- Jung, R. (2013). EGS—Goodbye or back to the future. <https://doi.org/10.5772/56458>
- Khodayar, M., Axelsson, G., & Steingrímsson, B. (2015). Potential structural flow paths for tracers and source faults of earthquakes at Hús múli, Hengill, South Iceland. Report Iceland GeoSurvey, iSOR-2015/035.
- Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, 90(1), 3–14. <https://doi.org/10.1785/0220180259>
- Kwiatk, G., Martínez-Garzón, P., Dresen, G., Bohnhoff, M., Sone, H., & Hartline, C. (2015). Effects of long-term fluid injection on induced seismicity parameters and maximum magnitude in northwestern part of the Geysers geothermal field. *Journal of Geophysical Research: Solid Earth*, 120(10), 7085–7101. <https://doi.org/10.1002/2015JB012362>
- Langenbruch, C., Weingarten, M., & Zoback, M. D. (2018). Physics-based forecasting of man-made earthquake hazards in Oklahoma and Kansas. *Nature Communications*, 9(1), 1–10. <https://doi.org/10.1038/s41467-018-06167-4>

- Langenbruch, C., & Zoback, M. D. (2017). Response to comment on "How will induced seismicity in Oklahoma respond to decreased saltwater injection rates?". *Science Advances*, 3(8). <https://doi.org/10.1126/sciadv.aao2277>
- Larson, J., Kramar, D., & Leonard, K. (2020). A geostatistical analysis of seismicity in Oklahoma using regression trees and neural networks. *Physical Geography*, 42(4), 334–350. <https://doi.org/10.1080/02723646.2020.1762982>
- Lubbers, N., Bolton, D. C., Mohd-Yusof, J., Marone, C., Barros, K., & Johnson, P. A. (2018). Earthquake catalog-based machine learning identification of laboratory fault states and the effects of magnitude of completeness. *Geophysical Research Letters*, 45(24), 13269–13276. <https://doi.org/10.1029/2018GL079712>
- Mahzari, P., Stanton-Yonge, A., Sanchez-Roa, C., Saldi, G., Mitchell, T., Oelkers, E. H., et al. (2021). Characterizing fluid flow paths in the Hellisheidi geothermal field using detailed fault mapping and stress-dependent permeability. *Geothermics*, 94, 102127. <https://doi.org/10.1016/j.geothermics.2021.102127>
- Majer, E. L., & Peterson, J. E. (2007). The impact of injection on seismicity at the Geysers, California Geothermal Field. *International Journal of Rock Mechanics and Mining Sciences*, 44(8), 1079–1090. <https://doi.org/10.1016/j.ijrmms.2007.07.023>
- Manley, G. F., Mather, T. A., Pyle, D. M., Clifton, D. A., Rodgers, M., Thompson, G., & Roman, D. C. (2021). Machine learning approaches to identifying changes in eruptive state using multi-parameter data sets from the 2006 eruption of Augustine Volcano, Alaska. *Journal of Geophysical Research: Solid Earth*, 126(12), e2021JB022323. <https://doi.org/10.1029/2021jb022323>
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50–60. <https://doi.org/10.1214/aoms/117730491>
- Martínez-Garzón, P., Kwiatak, G., Sone, H., Bohnhoff, M., Dresen, G., & Hartline, C. (2014). Spatiotemporal changes, faulting regimes, and source parameters of induced seismicity: A case study from the Geysers geothermal field. *Journal of Geophysical Research: Solid Earth*, 119(11), 8378–8396. <https://doi.org/10.1002/2014jb011385>
- Matter, J. M., Stute, M., Snæbjörnsdóttir, S. Ó., Oelkers, E. H., Gislason, S. R., Aradóttir, E. S., et al. (2016). Rapid carbon mineralization for permanent disposal of anthropogenic carbon dioxide emissions. *Science*, 352(6291), 1312–1314. <https://doi.org/10.1126/science.aad8132>
- McNamara, D. D., Sewell, S., Buscarlet, E., & Wallis, I. C. (2016). A review of the Rotokawa geothermal field, New Zealand. *Geothermics*, 59, 281–293. <https://doi.org/10.1016/j.geothermics.2015.07.007>
- Mehrabifard, A., & Eberhardt, E. (n.d.). Investigation of the dependence of induced seismicity magnitudes on differential stress and pore pressure using supervised machine learning, Northeastern British Columbia (NTS 093, 094A, B, G, H) and globally.
- Mignan, A., & Broccardo, M. (2019). One neuron versus deep learning in aftershock prediction. *Nature*, 574(7776), E1–E3. <https://doi.org/10.1038/s41586-019-1582-8>
- Mizrahi, L., Nandan, S., & Wiemer, S. (2021). The effect of declustering on the size distribution of mainshocks. *Seismological Society of America*, 92(4), 2333–2342. <https://doi.org/10.1785/0220200231>
- Mordensky, S. P., Lipor, J. J., DeAngelo, J., Burns, E. R., & Lindsey, C. R. (2023). When less is more: How increasing the complexity of machine learning strategies for geothermal energy assessments may not lead toward better estimates. *Geothermics*, 110, 102662. <https://doi.org/10.1016/j.geothermics.2023.102662>
- Mossop, A., & Segall, P. (1999). Volume strain within the Geysers geothermal field. *Journal of Geophysical Research: Solid Earth*, 104(B12), 29113–29131. <https://doi.org/10.1029/1999jb900284>
- 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>
- Norbeck, J. H., McClure, M. W., & Horne, R. N. (2018). Field observations at the Fenton Hill enhanced geothermal system test site support mixed-mechanism stimulation. *Geothermics*, 74, 135–149. <https://doi.org/10.1016/j.geothermics.2018.03.003>
- O'Sullivan, M. J., Pruess, K., & Lippmann, M. J. (2001). State of the art of geothermal reservoir simulation. *Geothermics*, 30(4), 395–429. [https://doi.org/10.1016/S0375-6505\(01\)00005-0](https://doi.org/10.1016/S0375-6505(01)00005-0)
- 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. *The Journal of Machine Learning Research*, 12, 2825–2830.
- Raleigh, C. B., Healy, J. H., & Bredehoeft, J. D. (1976). An experiment in earthquake control at Rangely, Colorado. *Science*, 191(4233), 1230–1237. <https://doi.org/10.1126/SCIENCE.191.4233.1230>
- Ratouis, T. M. P., Snæbjörnsdóttir, S. Ó., Sigfússon, B., Gunnarsson, I., Gunnarsson, G., & Aradóttir, E. S. (2019). Modeling the flow paths at the CarbFix2 reinjection site, Hellisheidi geothermal power plant, SW-Iceland. *Proceedings 44th Workshop on Geothermal Reservoir Engineering* (pp. 1–12).
- Ratouis, T. M. P., Snæbjörnsdóttir, S. Ó., Voigt, M. J., Sigfússon, B., Gunnarsson, G., Aradóttir, E. S., & Hjörleifsdóttir, V. (2022). CarbFix 2: A transport model of long-term CO₂ and H₂S injection into basaltic rocks at Hellisheidi, SW-Iceland. *International Journal of Greenhouse Gas Control*, 114, 103586. <https://doi.org/10.1016/j.ijggc.2022.103586>
- Riffault, J., Dempsey, D., & Archer, R. (2019). Microearthquake enhanced permeability imaging of a stimulated well. *Fifty-Third U.S. Rock Mechanics/Geomechanics Symposium*. Retrieved from <http://onepetro.org/ARMAUSRMS/proceedings-pdf/ARMA19/AII-ARMA19/ARMA-2019-0106/1125832/arma-2019-0106.pdf>
- Riffault, J., Dempsey, D., Karra, S., & Archer, R. (2018). Microseismicity cloud can be substantially larger than the associated stimulated fracture volume: The case of the Paralana enhanced geothermal system. *Journal of Geophysical Research: Solid Earth*, 123(8), 6845–6870. <https://doi.org/10.1029/2017JB015299>
- Rinaldi, A. P., & Nespole, M. (2017). TOUGH2-seed: A coupled fluid flow and mechanical-stochastic approach to model injection-induced seismicity. *Computers and Geosciences*, 108, 86–97. <https://doi.org/10.1016/j.cageo.2016.12.003>
- Rintamäki, A. E., Hillers, G., Vuorinen, T. A. T., Luhta, T., Pownall, J. M., Tsarsitalidou, C., et al. (2021). A seismic network to monitor the 2020 EGS stimulation in the Espoo/Helsinki Area, Southern Finland. *Seismological Research Letters*, 93(2A), 1046–1062. <https://doi.org/10.1785/0220210195>
- Ritz, V., Rinaldi, A. P., Karvounis, D., Castilla, R., Colas, E., Meier, P. M., et al. (2021). Preliminary modeling activities for an adaptive traffic light system for the Hengill geothermal field. In *World Geothermal Congress (WGC 2020+ 1)* (Vol. 2, pp. 3861–3869).
- Ritz, V. A., Rinaldi, A. P., & Wiemer, S. (2022). Transient evolution of the relative size distribution of earthquakes as a risk indicator for induced seismicity. *Communications Earth & Environment*, 3(1), 249. <https://doi.org/10.1038/s43247-022-00581-9>
- Ritz, V. A., Rinaldi, A. P., Zbinden, D., Nespole, M., Karvounis, D., & Wiemer, S. (2020). Modeling induced seismicity with a hydraulic-mechanical-stochastic simulator: Review of case studies. *OnePetro*. Retrieved from <http://onepetro.org/ARMAUSRMS/proceedings-pdf/ARMA20/AII-ARMA20/ARMA-2020-1654/2255758/arma-2020-1654.pdf>

- Ross, Z. E., Yue, Y., Meier, M. A., Hauksson, E., & Heaton, T. H. (2019). PhaseLink: A deep learning approach to seismic phase association. *Journal of Geophysical Research: Solid Earth*, 124(1), 856–869. <https://doi.org/10.1029/2018JB016674>
- Rouet-Leduc, B., Hulbert, C., & Johnson, P. A. (2018). Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nature Geoscience*, 12(1), 75–79. <https://doi.org/10.1038/s41561-018-0274-6>
- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, 44(18), 9276–9282. <https://doi.org/10.1002/2017GL074677>
- Rouet-Leduc, B., Hulbert, C., McBrearty, I. W., & Johnson, P. A. (2020). Probing slow earthquakes with deep learning. *Geophysical Research Letters*, 47(4). <https://doi.org/10.1029/2019GL085870/FORMAT/PDF>
- Rowland, J. V., Wilson, C. J. N., & Gravley, D. M. (2010). Spatial and temporal variations in magma-assisted rifting, Taupo Volcanic Zone, New Zealand. *Journal of Volcanology and Geothermal Research*, 190(1–2), 89–108. <https://doi.org/10.1016/J.JVOLGEORES.2009.05.004>
- Rutqvist, J., Jeanne, P., Dobson, P. F., Garcia, J., Hartline, C., Hutchings, L., et al. (2016a). The Northwest Geysers EGS Demonstration Project, California—Part 2: Modeling and interpretation. *Geothermics*, 63, 120–138. <https://doi.org/10.1016/j.geothermics.2015.08.002>
- Rutqvist, J., Rinaldi, A. P., Cappa, F., Jeanne, P., Mazzoldi, A., Urpi, L., et al. (2016b). Fault activation and induced seismicity in geological carbon storage—Lessons learned from recent modeling studies. *Journal of Rock Mechanics and Geotechnical Engineering*, 8(6), 789–804. <https://doi.org/10.1016/j.jrmge.2016.09.001>
- Schoenball, M., Davatzes, N. C., & Glen, J. M. G. (2015). Differentiating induced and natural seismicity using space-time-magnitude statistics applied to the Coso geothermal field. *Geophysical Research Letters*, 42(15), 6221–6228. <https://doi.org/10.1002/2015GL064772>
- Segall, P., & Lu, S. (2015). Injection-induced seismicity: Poroelastic and earthquake nucleation effects. *Journal of Geophysical Research: Solid Earth*, 120(7), 5082–5103. <https://doi.org/10.1002/2015JB012060>
- Sepulveda, F. R., Siega, C., Bixley, P. F., Mannington, W. I., Milloy, S. F., Soengkono, S., & Andrews, J. (2014). Wairakei geothermal field boundary: Insights from recent geophysics and reservoir information. In *Proceedings Thirty-Ninth Workshop on Geothermal Reservoir Engineering*, Figure 1 (pp. 1–9).
- Sepulveda, F. R., Siega, C., Lim, Y. W., Urgel, A., & Boese, C. (2016). Observations of deep and shallow seismicity within the Wairakei-Tauhara geothermal system. In *Proceedings 38th New Zealand Geothermal Workshop*.
- Sewell, S. M., Addison, S. J., Hernandez, D., Azwar, L., & Barnes, M. L. (2015). Rotokawa conceptual model update 5 yr after commissioning of the 138 MWe NAP plant. In *Proceedings New Zealand Geothermal Workshop*.
- Sewell, S. M., Cumming, W., Bardsley, C. J., Winick, J., Quinao, J., Wallis, I. C., et al. (2015). Interpretation of microearthquakes at the Rotokawa geothermal field, 2008–2012. In *Proceedings World Geothermal Congress 2015 Melbourne, Australia, 19–25 April 2015* (pp. 19–25).
- Shapiro, S. A., Krüger, O. S., & Dinske, C. (2013). Probability of inducing given-magnitude earthquakes by perturbing finite volumes of rocks. *Journal of Geophysical Research: Solid Earth*, 118(7), 3557–3575. <https://doi.org/10.1002/jgrb.50264>
- Shapiro, S. A., Rentsch, S., & Rothert, E. (2005). Characterization of hydraulic properties of rocks using probability of fluid-induced microearthquakes. *Geophysics*, 70(2), F27–F33. <https://doi.org/10.1190/1.1897030>
- Sherburn, S., Bannister, S., Sewell, S., Bourguignon, S., & Bromley, C. (2015). New Zealand geothermal induced seismicity: An overview. In *Proceedings World Geothermal Congress*. Retrieved from <https://www.researchgate.net/publication/284578912>
- Sherburn, S., Bourguignon, S., Bannister, S., Sewell, S., Cumming, B., Bardsley, C., et al. (2013). Microseismicity at Rotokawa geothermal field, 2008–2012. *Thirty-Fifth New Zealand Geothermal Workshop: 2013 Proceedings*. Retrieved from <https://www.researchgate.net/publication/258764853>
- Sherburn, S., Sewell, S. M., Bourguignon, S., Cumming, W., Bannister, S., Bardsley, C., et al. (2015). Microseismicity at Rotokawa geothermal field, New Zealand, 2008–2012. *Geothermics*, 54, 23–34. <https://doi.org/10.1016/J.GEOTHERMICS.2014.11.001>
- Shirzaei, M., Ellsworth, W. L., Tiampo, K. F., González, P. J., & Manga, M. (2016). Surface uplift and time-dependent seismic hazard due to fluid injection in eastern Texas. *Science*, 353(6306), 1416–1419. <https://doi.org/10.1126/science.aag0262>
- Sigmundsson, F., Hooper, A., Hreinsdóttir, S., Vogfjörð, K. S., Ófeigsson, B. G., Heimisson, E. R., et al. (2014). Segmented lateral dyke growth in a rifting event at Bárðarbunga volcanic system, Iceland. *Nature*, 517(7533), 191–195. <https://doi.org/10.1038/nature14111>
- Sinha, S., Wen, Y., De Lima, R. P., & Marfurt, K. (2018). Statistical controls on induced seismicity. *SPE/AAPG/SEG Unconventional Resources Technology Conference 2018* (Vol. 2018, pp. 23–25). URTC. <https://doi.org/10.15530/URTEC-2018-2897507>
- Snæbjörnsdóttir, S., Tómasdóttir, S., Sigfússon, B., Aradóttir, E. S., Gunnarsson, G., Niemi, A., et al. (2018). The geology and hydrology of the CarbFix2 site, SW-Iceland. *Energy Procedia*, 146, 146–157. <https://doi.org/10.1016/j.egypro.2018.07.019>
- Tómasdóttir, S. (2018). Flow paths in the Húsúlí reinjection zone, Iceland. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-353580>
- Trugman, D. T., Shearer, P. M., Borsa, A. A., & Fialko, Y. (2016). A comparison of long-term changes in seismicity at the Geysers, Salton Sea, and Coso geothermal fields. *Journal of Geophysical Research: Solid Earth*, 121(1), 225–247. <https://doi.org/10.1002/2015JB012510>
- Wallis, I. C., Bardsley, C., Powell, T., Rowland, J. V., & O'Brien, J. M. (2013). A structural model for the Rotokawa geothermal field, New Zealand. *Thirty-Fifth New Zealand Geothermal Workshop: 2013 Proceedings* (pp. 17–20).
- Warpinski, N. (2009). Microseismic monitoring: Inside and out. *Journal of Petroleum Technology*, 61(11), 80–85. <https://doi.org/10.2118/118537-JPT>
- Webb, T. H., Ferris, B. G., & Harris, J. S. (2012). The Lake Taupo, New Zealand, earthquake swarms of 1983. *New Zealand Journal of Geology and Geophysics*, 29(4), 377–389. <https://doi.org/10.1080/00288306.1986.10422160>
- Wozniakowska, P., & Eaton, D. (2020a). Determination of factors controlling geological susceptibility to induced seismicity in the Montney Formation based on a machine-learning approach (pp. 19–26). Retrieved from <http://www.geosciencebc.com/updates/summary-of->
- Wozniakowska, P., & Eaton, D. W. (2020b). 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>
- Yentes, J. M., Hunt, N., Schmid, K. K., Kaipust, J. P., McGrath, D., & Stergiou, N. (2013). The appropriate use of approximate entropy and sample entropy with short data sets. *Annals of Biomedical Engineering*, 41(2), 349–365. <https://doi.org/10.1007/s10439-012-0668-3>
- Yu, P., Dempsey, D., & Archer, R. (2021). A three-dimensional coupled thermo-hydro-mechanical numerical model with partially bridging multi-stage contact fractures in horizontal-well enhanced geothermal system. *International Journal of Rock Mechanics and Mining Sciences*, 143, 104787. <https://doi.org/10.1016/J.IJRMMS.2021.104787>
- Zöller, G., & Holschneider, M. (2016). The maximum possible and the maximum expected earthquake magnitude for production-induced earthquakes at the gas field in Groningen, The Netherlands. *Bulletin of the Seismological Society of America*, 106(6), 2917–2921. <https://doi.org/10.1785/0120160220>