Statistical bounds on how induced seismicity stops

Ryan Schultz*, William L. Ellsworth & Gregory C. Beroza

Earthquakes caused by human activities receive scrutiny due to the risks and hazards they pose. Seismicity that occurs after the causative anthropogenic operation stops has been particularly problematic—both because of high-profile cases of damage caused by this trailing seismicity and due to the loss of control for risk management. With this motivation, we undertake a statistical examination of how induced seismicity stops. We borrow the concept of Båth’s law from tectonic aftershock sequences. Båth’s law anticipates the difference between magnitudes in two subsets of seismicity as dependent on their population count ratio. We test this concept for its applicability to induced seismicity, including ~ 80 cases of earthquakes caused by hydraulic fracturing, enhanced geothermal systems, and other fluid-injections with clear operational end points. We find that induced seismicity obeys Båth’s law: both in terms of the magnitude-count-ratio relationship and the power law distribution of residuals. Furthermore, the distribution of count ratios is skewed and heavy-tailed, with most earthquakes occurring during stimulation/injection. We discuss potential models to improve the characterization of these count ratios and propose a Seismogenic Fault Injection Test to measure their parameters in situ. We conclude that Båth’s law quantifies the occurrence of earthquake magnitudes trailing anthropogenic operations.

The injection of fluids in the subsurface has the potential to cause faults to reactivate with earthquake slip1,2. Numerous high-magnitude cases of induced seismicity causing significant impacts3–6 have led to increased attention. Study topics have covered a broad range of aspects including triggering mechanisms, maximum magnitude, potential risks/hazards, and management strategies7,8. Despite the proliferation of scientific studies, relatively little attention has been devoted to understanding how induced earthquake sequences stop. This is surprising because earthquakes trailing well shut-in have been identified as one of the most significant sources of risk for induced seismicity management9. In practice, there are many cases where the largest magnitude earthquake trails the shut-in of the responsible well10–12 or where abrupt ‘jumps’ in magnitude have occurred13.

In this study we build a conceptual framework for statistical bounds on how induced earthquake sequences stop. We start by following the development of Båth’s law for aftershocks of natural earthquakes and adapt it through a set of assumptions and justifications for induced earthquakes. We test the validity of this framework on sequences of induced seismicity and find that in most cases the majority of earthquakes occur during the stimulation/injection periods, rather than trailing well shut-in. As well, the largest events occurring during the stimulation/trailing periods follow Båth’s law. We also discuss the implications of our results and suggest recommendations for in situ tests that could be useful for constraining induced earthquake risks. Finally, we speculate on how improved physical models could be used to build upon our empirical analysis of how induced earthquakes stop.

Background information on tectonic aftershocks and Båth’s law

Båth’s law is an empirical observation that the average magnitude difference between a mainshock earthquake and its largest aftershock is approximately 1.2 magnitude units14. Statistical models of earthquakes15,16 have come to interpret Båth’s law as a probabilistic consequence of two more fundamental observations17–20. The first being an aftershock law (i.e., Omori’s law), which describes the temporal decay of earthquake rate following a main shock21–23.

\[ n^p(t) = K(t + c)^{-p} \]

(1)

In Omori’s law \( n(t) \) is the time-dependent rate of aftershocks, \( K \) describes the productivity of the aftershock sequence and strongly influences the total number of aftershocks, \( p \) is the decay exponent, and \( c \) is an ad hoc parameter added to ensure the equation doesn’t diverge at the time of the mainshock (i.e., \( t = 0 \)). In this formulation, the value of \( p \) is bounded such that \( p > 1 \) for a finite number of earthquakes to occur24. The second
fundamental observation is the Gutenberg Richter magnitude frequency distribution (GR-MFD), which describes the relative number of small and large earthquakes in a sample.\(^2\)\(^5\) In the GR-MFD, \(N\) is the sample size of the catalogue, \(a\) is the \(a\)-value which is related to the productivity of the sequence, and \(b\) is the \(b\)-value, which controls the proportionality of big-to-small magnitude earthquakes.

Omori’s law governs the total number of earthquakes that follow a mainshock \(N_A\) and the GR-MFD controls the distribution of earthquake magnitudes ‘randomly’ sampled in that population. The combination of these two observations gives the expression for the average magnitude difference (where \(N > N_A > 1\)).

\[
\langle \Delta M \rangle \approx \frac{1}{b} \log_{10}(N / N_A)
\]

where the total number of earthquakes in a catalogue \(N\) is an observable and the number of aftershocks \(N_A\) can be forecast/estimated from the Omori’s law parameters. Typical estimates of Omori and GR-MFD parameters yields estimates of the average magnitude difference \(\langle \Delta M \rangle\) that are similar to the empirical observation of \(\sim 1.2\) \(M\)\(^2\)\(^7\)\(^,\)\(^8\).

**Deriving a trailing earthquake model for induced seismicity**

Following a similar rationale, we derive a trailing earthquake model from the statistical viewpoint of induced earthquakes as a point-process. We first assume that induced earthquake magnitudes are independent and identically distributed random variables that follow the GR-MFD. From this assumption an expression for the distribution of the maximum magnitude earthquake follows (in the limit of large \(N\)).

\[
M_{\text{max}} \approx M_C + \frac{1}{b} \log_{10}(N) - \frac{1}{b} \log_{10}(-\ln(u))
\]

This is often referred to as the sample size hypothesis for the distribution of maximum magnitudes \(M_{\text{max}}\), where the newly introduced parameters are the magnitude of completeness \(M_C\)\(^1\)\(^1\) and a confidence level \(u\) that can be utilized to sample the \(M_{\text{max}}\) distribution. The distribution of expected maximum magnitudes for different population sizes is shown in Fig. 1a.

Based on this sample size hypothesis, we can consider an archetype of an induced earthquake sequence that is broken into two segments: the injection/stimulation phase (denoted by a subscript S) and the trailing shut-in phase (denoted by a subscript T) that follows (Fig. 1c). The stimulation phase would entail a population count of \(N_S\) stimulation earthquakes that occur until the anthropogenic operation is halted at time \(T_S\),
Table 1. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

| Model name | Reference(s) | Aftershock rate decay | Cumulative aftershock count | Total aftershock count |
|------------|--------------|-----------------------|----------------------------|-----------------------|
| Modified Omori⁴⁰⁻²³ | \( n^0(t) = K(t + c)^{-p} \) | \( N^0(t) = \left( \frac{1}{\tau} \right) (c^{1-p} - (t + c)^{1-p}) \) | \( N^0(\infty) = K \frac{1}{1-p} \) |
| Exponential⁴¹ | \( n^0(t) = n^0(0) e^{-\mu t} \) | \( N^e(t) = \tau n^0(0) (1 - e^{-\mu t}) \) | \( N^e(\infty) = \tau n^0(0) \) |
| Stretched exponential⁴² | \( n^0(t) = q K e^{-\mu t} \left( \frac{\gamma(t)}{\gamma(c)} \right)^{q} \left( \frac{(c^{1-p})}{(t^{1-p})} \right) \) | \( N^s(t) = N^0(1 - e^{-\left( \frac{\gamma(t)}{\gamma(c)} \right)^{q} \left( \frac{(c^{1-p})}{(t^{1-p})} \right)}) \) | \( N^s(\infty) = N^0 \) |
| Cut-off Power Law⁴³ | \( n^0(t) = K(t + c)^{-p} e^{-\mu t} \) | \( N^c(t) = K t^{1-p} e^{\mu t} (\Gamma(t - p, \frac{\mu}{\mu}) - \Gamma(t - p, \frac{\mu}{\mu} + 1)) \) | \( N^c(\infty) = K t^{1-p} e^{\mu t} \Gamma(t - p, \frac{\mu}{\mu} + 1) \) |
| Gamma⁴⁴ | \( n^0(t) = A t^{-\beta} (\gamma(q, A t) - \gamma(q, A t)) \) | \( N^\gamma(t) = \left( \frac{A^\beta}{\beta \Gamma(q, A t)} \right) \left( \Gamma(q, A t - 1) - \Gamma(q, A t) \right) + A t^{-\beta} e^{-\beta t} + (1 - e^{-\beta t}) \Gamma(q, A t) \) | \( N^\gamma(\infty) = \left( \frac{A^\beta}{\beta \Gamma(q, A t)} \right) \left( \Gamma(q, A t - 1) - \Gamma(q, A t) \right) + A t^{-\beta} + (1 - e^{-\beta t}) \) |
| Epidemic type⁴⁵ | \( n^0(t) = \sum K_i (t - c_i)^{-p} \) | \( N^K(t) = \sum K_i (c^{1-p} - (t - c_i)^{1-p}) \) | \( N^K(\infty) = \sum K_i \) |

Received: 19 November 2022 | Accepted: 8 March 2023

https://doi.org/10.1038/s41598-022-05216-9

In summary, we found that the time decay of the maximum magnitude prediction \( \Delta M \) for the induced MFD is consistent with the decay rate of the mainshock magnitude prediction \( M_{\text{max}, S} \).

### Acknowledgments

This work was supported by the National Science Foundation (NSF) under grant EAR-2130645.

### Competing Interests

The authors declare no competing interests.

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

### Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

### References

1. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

2. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

3. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

4. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

5. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

6. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

7. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

8. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

9. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

10. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

11. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.

12. Aftershock models proposed for tectonic earthquakes. For each model we provide an expression for the aftershock rate function \( n(t) \), the cumulative aftershock count \( N(t) \), and the total number of aftershocks \( N(\infty) \). Corresponding references are also provided for each model. Note that \( \gamma(a,x) \) and \( \Gamma(a,x) \) are the lower and upper incomplete gamma functions, respectively. Superscripts are used to denote the aftershock model used.
Testing the model’s validity
In adapting Båth’s law to induced seismicity, we’ve made several assumptions (Fig. 1): that the sequence is described by a nonhomogenous Poisson process, that the extended fault system can effectively be modelled as a point-process, that there’s a single lag time between operations and seismogenic response, that statistical parameters are stationary during the observation interval, and that there are no injection-volume or tectonic limits on the maximum magnitude. To validate (or refute) these assumptions, we begin by examining induced seismicity cases.

We focus only on cases where there has been a clearly delineated end of operations (Fig. S1). This criterion dominantly encompasses cases related to hydraulic fracturing or enhanced geothermal systems. We select cases of induced seismicity where both catalogue and injection time series information is available. Cases include the Duvernay11,13,46,47 and Montney48 in Canada, the Bowland Shale in the UK49,50, the Appalachian Basin51,52, SCOOP & STACK play53–55, Eagle Ford56, and Arkoma Basin57 in the United States, the Sichuan Basin in China58,59, Soultz-sous-Forêts in France60,61, Cooper Basin62 and Paralana63 in Australia, KTB64 and Groß Schönebeck65 in Germany, Basel66 and St Gallen67 in Switzerland, Helsinki in Finland68, Pohang in South Korea69, and Berlin in El Salvador70. Many of these cases have data available in online repositories71, within aggregate comparison studies12,29,72, or as supplements to the individual case studies.

Briefly, individual cases are truncated at their magnitude of completeness (MC) and then subdivided into two populations: those occurring during stimulation or those trailing stimulation. We report statistics such as earthquake counts, largest magnitudes, beta-value, and MC, which we summarize in Table S1. For additional details on the data acquisition and preprocessing, we direct the reader to the Supplementary Information (Section S1).

Examining the Statistics of $R_S$. From our compiled database (Table S1), we examine the parameter $R_S$ observed during numerous induced seismicity sequences (Fig. 2a). $R_S$ shows a highly skewed distribution—tending towards values of 77% on average and 86% as a median (Fig. S5). This is comparable to observations in prior studies on hydraulic fracturing that found ~ 90% of events occurred during stage stimulation in the Duvernay60 and ~ 87% of events in the Sichuan Basin were ‘forced’ under a modified Epidemic Type Aftershock Model58. While the distribution is skewed towards high $R_S$ values, we also note that it is heavy-tailed with a handful of cases with low $R_S$ values. For example, the enhanced geothermal stimulation at Pohang shows a large trailing aftershock sequence69.

To further examine the skewed nature of the $R_S$ distribution, we compare our empirical observations to the distribution expected from a random process (Fig. 2a). To model the random process, we consider a uniform random variable for both $N_t$ and $N_c$. We then perform the two-sample Kolmogorov–Smirnov test between our empirically observed $R_S$ and the random process model. The two-sample Kolmogorov–Smirnov test is a non-parametric hypothesis test that assesses the probability that two samples are drawn from the same (arbitrary) distribution73. In performing this test, we find extremely small $p$ values that are orders of magnitude less than 5% (Fig. S6)—which rejects the null hypothesis that the two samples are from the same distribution. The failure of $R_S$ to be modelled as a random process suggests that there is a physical process underlying this distribution (e.g., Eqs. 7 and 8).

A better representation of the statistics of the $R_S$ would be useful for estimation of differential magnitudes $\Delta M$ (e.g., Eq. 5). Towards this end, we fit our histogram results to a beta distribution—a two-parameter ($\alpha$ and $\beta$) family of distributions defined over the interval 0–174. We select the beta distribution due to its flexibility in fitting probability density functions over this interval. Further, to capture the variability in the fitted parameters...
In this section we discuss the interpretations and implications of our results. We divide the discussion into three parts: (1) the applicability of Båth’s law for induced seismicity, (2) the relative importance of parameters in Båth’s law, and (3) a diagnostic test for operators to use in situ when encountering seismogenic faults.
The applicability of Båth’s law for induced seismicity. To date, many models have been introduced to capture the anticipated maximum magnitude of induced earthquakes from either physical or statistical perspectives. Typically, these models have focused on the earthquakes occurring during the stimulation process. In this paper, we have described a statistical framework to extend these considerations to those earthquakes that trail after well shut-in based on Båth’s law for tectonic earthquakes. This conceptual extension is partly driven by the observation of large earthquake magnitude jumps, high-profile cases of large trailing events, and the contribution of trailing events to earthquake risk management.

Our results suggest that Båth’s law is a viable foundation for anticipating the magnitude of earthquakes trailing an operation that produces anthropogenic seismicity. This assertion is supported by the observation that: (1) induced seismicity cases are well fit by the linear relationship between magnitude differences $\Delta M$ and logarithmic population ratios $\frac{1}{2}\log_{10}(R_{TS})$ that is predicted by Båth’s law (Fig. 3a), and (2) the residuals from this relationship are approximately consistent with the power law distribution anticipated by Båth’s law (Fig. 3b). Based on this, we further suggest that the description of Båth’s law (Eq. 5) will have predictive power for quantitative forecasts of trailing seismicity. Of course, this is predicated on knowing/estimating the population ratios $R_{TS}$—either a priori or updating in real-time during an operation. In this sense, the empirical distribution of $R_S$ and its best fit to a beta distribution (Fig. 2) provides some utility—possibly as prior information in a Bayesian updating scheme. Much of the literature that has been described for tectonic aftershock forecasts through extreme value statistics and Bayesian approaches, will likely be adaptable for the purpose of constraining $N_T$. Additionally, models of stimulation seismicity that consider the full injection history should provide better predictability to the anticipated stimulation events $N_S$.

This interpretation is in line with recent work that has found that induced earthquake sequences have statistical properties that are similar to (or subtle modifications of) their tectonic counterparts. These results could be indicative of the importance of fluid-flow along faults for tectonic earthquakes, which has been suggested in some studies as important factor for the aftershock process. General links between statistical models and physical processes have been made for earthquakes and aftershocks. Thus, incorporating and testing physical models of aftershocks that consider realistic fault geometries, fault rheology, stress, material heterogeneity, and fluid flow will likely be important for better understanding post shut-in induced earthquakes.

Relative importance of Båth’s law parameters. A better understanding the physical processes that underlie the observed statistics would allow for better understanding of trailing seismicity. While the collected dataset (Table S1) is too small to meaningfully discern systematic trends in $R_S$ that vary with anthropogenic type, operational parameters, fault architecture, or geological setting, we may still apply our simple model to better understand the relative importance of parameters that contribute to a skewed distribution of $R_S$. We focus first on the exponential decay model (Eq. 8) both for its simplicity and the breadth of induced seismicity studies it has been applied to. If we assume representative values for a hydraulic fracturing operation ($\Delta T = 1 \text{ h}, r = 3 \text{ days}, f = 1.0, T_s = 2 \text{ weeks}$), we find that $R_S = 82\%$; a value between those reported for the mean/median of the empirical $R_S$. To build on this further, we assume that the parameters $\Delta T$ and $r$ follow a lognormal distribution while $f$ follows a Gaussian distribution so that we may compute a synthetic $R_S$ distribution. From this, we perform a Markov Chain Monte Carlo (MCMC) optimization to maximize the Kolmogorov–Smirnov $p$ value between the empirical and synthetic $R_S$. The MCMC approach randomly samples the parameter space to find optimal mean and standard deviations of these parameters ($\Delta T, r, f$) that are bounded within those reasonably observed in the literature.

The MCMC approach produces synthetic $R_S$ distributions that fit the empirical distribution well (likelihoods of up to 90%). The input parameters distributions needed to fit $R_S$ well (Fig. 4) suggests that a skewed distribution of input parameters ($\Delta T$ or $r$) is required to match the skewed $R_S$ distribution. Such a skewed distribution includes drivers for black swan occurrences of extreme events observed in cases of induced seismicity: relatively rare events of large consequence. For example, the earthquakes at Pohang were exceptional in that a dormant fault system was reactivated by the geothermal stimulation program. Similarly, most events in a 2018 episode in British Columbia occurred as part of a delayed trigger after hydraulic fracturing stage completion. The largest event during the second well stimulation at Preston New Road occurred 60 h after the end of the last stimulation. Earthquakes were apparently induced at distances of up to ~5 km away from the enhanced geothermal stimulation at Strasbourg-Vendenheim. These consequential trailing seismicity events suggest the importance of better modelling and constraining parameters relevant to Båth’s law. We note, however, that there are significant limitations to our MCMC approach, as the distributions of $\Delta T$ and $r$ trade-off with one another. Thus, empirical observations of aftershock parameters ($r$ or $c$ and $p$) and injection responses ($\Sigma, b, \Delta T$) would be needed to better constrain these values/distributions.

The Seismogenic fault injection test (SFIT). Previously we showed that skewed values of $\Delta T$ or $r$ are required for a skewed $R_S$—and represent an increased potential for large trailing magnitude differences $\Delta M$. Given the possibility of consequential impacts from large magnitude trailing events, it would be prudent to measure these parameters in situ to reduce the uncertainties associated with fault reactivation.

Similar to Diagnostic Fracture Injection Tests where fluids are intentionally injected into a reservoir to constrain its hydrological properties and stress state, we suggest an analogous Seismogenic Fault Injection Test (SFIT) to calibrate the seismic response to fluid injection (Fig. 5). Ideally, an inventory of faults would be available from a pre-assessment to injection. Injection would be performed with the intention of measuring both the time-delayed response to injection ($\Sigma, b, \Delta T$) and the aftershock-like decay of seismicity ($r$ or $c$ and $p$) for each of the known faults. The SFIT should be designed to constrain parameters of greatest consequence ($\Delta T$ and $r$). In the context of our study, unknown faults activated at distance from the stimulation interval (e.g., large
ΔT) or with prolonged trailing sequences (e.g., large \( \tau \)) pose the greatest risks. With this in mind, an individual injection phase would span a range of injection rates, potentially sweeping through them cyclically in a ‘chirped’ manner (Fig. 5); the intent would be to both discern any rate-dependent features and better highlight potential time lags ΔT. As well, individual injection phases should be repeated to bolster confidence in the measurements, in particular the decay rate \( \tau \) which only is well constrained after shut-in\(^{79} \). Ideally, these repeated injections phases would be spaced at non-periodic intervals to unambiguously measure delay times \( \Delta T_{\text{ss}, \text{se}} \). Finally, if we are making suggestions outside of what the simple models predict, varying the shut-in procedure for individual injection phases would be useful to discern safer responses to unwanted seismicity.

Figure 4. Synthetically fitting the \( R_S \) distribution to model parameters. (d) The archetypal model for trailing seismicity ratios (Eq. 8) is fit (grey bars) to the empirical distribution of \( R_S \) (blue bars). The distribution of model parameters (a) ΔT, (b) \( f \), and (c) \( \tau \) are also shown.

Figure 5. The Seismogenic Fault Injection Test (SFIT). Schematic diagram of a diagnostic test to measure a fault’s seismogenic response through parameters such as \( \Sigma, b, \Delta T, \) and \( \tau \) (or \( c \) and \( p \)). Repeated injection tests (at non-periodic intervals) are performed to verify the accuracy of measurements. Inset diagram shows how individual tests could be ‘chirped’ in their injection rate to better constrain parameters.
| Variable | Description |
|----------|-------------|
| $a$ | GR-MFD productivity parameter, $a$-value |
| $b$ | GR-MFD slope parameter, $b$-value |
| $M$ | Earthquake magnitude and magnitude difference, respectively |
| $M_c$ | Magnitude of completeness |
| $\Sigma$ | Seismogenic Index |
| $V(t), \dot{V}(t)$ | Cumulative injection volume and injection rate, respectively |
| $n(t)$ | Earthquake rate |
| $N, N_a, N_T, N_S$ | Cumulative earthquake counts: total, aftershocks, trailing, and stimulation |
| $R_S, R_T$ | Earthquake count ratios: related by $1/R_S = N/N_s = 1 + N_a/N_T = 1 + R_T$ |
| $u, u_T, u_S$ | Confidence level variables |
| $\Delta T$ | Simple archetype model, stimulation-earthquake response delay |
| $T_3$ | Stimulation/injection time interval |
| $f$ | Stimulation-trailing rate change factor |
| $K$ | Omori aftershock productivity parameter |
| $p$ | Omori aftershock decay exponent |
| $\tau$ | Omori aftershock singularity parameter |

Table 2. Summary of equation parameters and variables. Here we list the relevant parameters used in equations throughout our study, for convenience to the reader.

Conclusions
In summary, we have tested the applicability of Båth’s law to cases of induced seismicity with clear ends to fluid injection. Båth’s law fits the expected relationship, as confirmed by a linear regression of the collected data. Additionally, the residuals to the data follow the expected power law distribution, as confirmed by the Kolmogorov-Smirnov tests. Thus, our compilation of population ratio $R_S$ statistics, including a fit to the beta distribution, could be useful for forecasting trailing seismicity. Finally, simple point-like models describing parameters of interest ($\Delta T$ and $\tau$); we propose in situ tests to help quantify the risks associated with seismicity trailing an anthropogenic operation. Variables used throughout the paper are summarized in Table 2.

Data availability
Data used to produce the results of this study are available as an electronic supplement (Table S1). In addition, the codes used to derive these results are available online, through GitHub (https://github.com/RyanJamesSchultz/IS-Bath). Catalogue data and injection data are available in online repositories 51, within aggregate comparison studies12,29,72, as supplements to the individual case studies.

Received: 21 October 2021; Accepted: 7 January 2022
Published online: 24 January 2022

References
1. Ellsworth, W. L. Injection-induced earthquakes. *Science* [https://doi.org/10.1126/science.1225942](https://doi.org/10.1126/science.1225942) (2013).
2. Schultz, R. et al. Hydraulic fracturing-induced seismicity. *Rev. Geophys.* 58(3), e2019RG000695. [https://doi.org/10.1029/2019RG000695](https://doi.org/10.1029/2019RG000695) (2020).
3. Keranen, K. M., Savage, H. M., Abers, G. A. & Cochran, E. S. Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence. *Geology* 41(6), 699–702. [https://doi.org/10.1130/G34045.1](https://doi.org/10.1130/G34045.1) (2013).
4. Van der Voort, N. & Vanclay, F. Social impacts of earthquakes caused by gas extraction in the Province of Groningen, The Netherlands. *Environ. Impact Assess. Rev.* 50, 1–15. [https://doi.org/10.1016/j.eiar.2014.08.008](https://doi.org/10.1016/j.eiar.2014.08.008) (2015).
5. Ellsworth, W. L., Giardini, D., Townend, J., Ge, S. & Shimamoto, T. Triggering of the Pohang, Korea, earthquake (Mw 5.5) by enhanced geothermal system stimulation. *Seismol. Res. Lett.* 90(5), 1844–1858. [https://doi.org/10.1785/0220190102](https://doi.org/10.1785/0220190102) (2019).
6. Lei, X., Wang, Z. & Su, J. The December 2018 ML 5.7 and January 2019 ML 5.3 earthquakes in South Sichuan Basin induced by enhanced geothermal system stimulation. *Seismol. Res. Lett.* 90(3), 1099–1110. [https://doi.org/10.1785/0220190029](https://doi.org/10.1785/0220190029) (2019).
7. McGarr, A. Maximum magnitude earthquakes induced by fluid injection. *J. Geophys. Res. Solid Earth* 119(2), 1008–1019. [https://doi.org/10.1002/2013JB010597](https://doi.org/10.1002/2013JB010597) (2014).
8. Lee, K. K. et al. Managing injection-induced seismic risks. *Science* 364(6442), 730–732. [https://doi.org/10.1126/science.aax1878](https://doi.org/10.1126/science.aax1878) (2019).
9. Schultz, R., Beroza, G. C. & Ellsworth, W. L. A risk-based approach for managing hydraulic fracturing–induced seismicity. *Science* 372(6541), 504–507. [https://doi.org/10.1126/science.abg4541](https://doi.org/10.1126/science.abg4541) (2021).
10. Segall, P. & Lu, S. Injection-induced seismicity: Poroplastic and earthquake nucleation effects. *J. Geophys. Res. Solid Earth* 120(7), 5082–5103. [https://doi.org/10.1002/2015JB012860](https://doi.org/10.1002/2015JB012860) (2015).
11. Schultz, R., Wang, R., Gu, Y. J., Haug, K. & Atkinson, G. A. Seismological overview of the induced earthquakes in the Duvernay play near Fox Creek, Alberta. *J. Geophys. Res. Solid Earth* 122(1), 492–505. [https://doi.org/10.1002/2016JB013570](https://doi.org/10.1002/2016JB013570) (2017).
12. Verdon, J. P. & Bommer, J. J. Green, yellow, red, or out of the blue? An assessment of traffic light schemes to mitigate the impact of hydraulic fracturing-induced seismicity. *J. Seismol.* [https://doi.org/10.1007/s10950-020-09966-9](https://doi.org/10.1007/s10950-020-09966-9) (2020).
13. Schultz, R. & Wang, R. Newly emerging cases of hydraulic fracturing induced seismicity in the Duvernay East Shale Basin. *Tectonophysics* 779, 228393. [https://doi.org/10.1016/j.tecto.2020.228393](https://doi.org/10.1016/j.tecto.2020.228393) (2020).
57. Yoon, C. E., Huang, Y., Ellsworth, W. L. & Beroza, G. C. Seismicity during the initial stages of the Guy-Greenbrier, Arkansas, earthquake sequence. J. Geophys. Res. Solid Earth 122(11), 9253–9274. https://doi.org/10.1002/2017JB014946 (2017).

58. Lei, X. et al. Fault reactivation and earthquakes with magnitudes of up to Mw4.7 induced by shale-gas hydraulic fracturing in Sichuan Basin, China. Sci. Rep. 7(1), 1–12. https://doi.org/10.1038/s41598-017-08557-y (2017).

59. Tan, Y. et al. Hydraulic fracturing induced seismicity in the southern Sichuan Basin due to fluid diffusion inferred from seismic and injection data analysis. Geophys. Res. Lett. 47(4), e2019GL084885. https://doi.org/10.1029/2019GL084885 (2020).

60. Charlety, J. et al. Large earthquakes during hydraulic stimulations at the geothermal site of Soultz-sous-Forêts. Int. J. Rock Mech. Min. Sci. 44(8), 1091–1105. https://doi.org/10.1016/j.ijrmms.2007.06.003 (2007).

61. Dorbath, L., Cuenot, N., Genter, A. & Frogneux, M. Seismic response of the fractured and faulted granite of Soultz-sous-Forêts. Geothermics 103, 1–10. https://doi.org/10.1016/j.geothermics.2020.03.009 (2020).

Supplementary Information
The authors declare no competing interests.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-05216-9.
