Variability of Seismicity Rates and Maximum Magnitude for Adjacent Hydraulic Stimulations

G. Kwiatek[1], I. Grigoratos[2], and S. Wiemer[2]

1. Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Section 4.2 Geomechanics and Scientific Drilling, Telegrafenberg, 14473 Potsdam, Germany (kwiatek@gfz-potsdam.de)
2. ETH Zurich, Swiss Seismological Service, Sonneggstrasse 5, 8092 Zurich, Switzerland (iason.grigoratos@sed.ethz.ch; stefan.wiemer@sed.ethz.ch)

Corresponding Author:

G. Kwiatek, Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Section 4.2 Geomechanics and Scientific Drilling, Telegrafenberg, 14473 Potsdam, Germany, email: kwiatek@gfz-potsdam.de

Highlights

1. We forecast seismicity rates and next largest earthquake magnitudes using hydraulic data from two hydraulic stimulations at St1 Helsinki
2. Critical components of the seismicity rate model differ significantly between the two stimulations despite their close proximity
3. Forecasting the next largest magnitude using different models led to a wide range of outcomes that are inconsistent across stimulations

Abstract

We hindcasted the seismicity rates and the next largest earthquake magnitude using seismic and hydraulic data from two hydraulic stimulation campaigns carried out in adjacent (500 m apart) ultra-deep wells in Finland. The two campaigns performed in 2018 and 2020 took place in the frame of St1 Helsinki project produced stable, pressure-controlled induced seismic activity with maximum magnitudes of M_W 1.3 and 1.7, respectively. The seismicity rates were modeled using simplified physics-based approaches tailored to varying injection rates. This is the first time that this framework was applied to a cyclical injection protocol. The next largest earthquake magnitude was estimated using several existing models from the literature. Despite the close proximity of the two hydraulic stimulations and associated seismicity, we obtained strongly different parameterization of the critical model components, questioning the use of a-priori seismic hazard analysis tools in the planning of a neighboring stimulation. The differences in parameterization were attributed to the contrasting hydraulic energy rates observed in each stimulation, small differences in the structural inventory of the reservoir and resulting seismic injection efficiency, and potentially to variations in the injection protocol itself. As far as the seismicity rate model is concerned, despite a good performance during the 2018 campaign, the fit during the 2020 stimulation was suboptimal. Forecasting the next largest magnitude using different models led to a very wide range of outcomes. Moreover, their relative ranking across stimulations was inconsistent, including the situation whether the best performing model in 2018 stimulation was the worst performing one in the 2020 stimulation.

Plain language summary

We modeled the seismicity rate and the next largest earthquake magnitude using seismic and hydraulic data from two stimulation campaigns with high-pressure injection, carried out in adjacent deep wells in Finland. The two campaigns took place in 2018 and 2020 in the frame of St1 Deep Heat (Enhanced Geothermal) project and led to prominent seismic activity with the largest earthquakes reaching magnitudes of M_W 1.3 and 1.7, respectively. The seismicity rates were modeled using simplified physics-based approaches tailored to the cyclical injection rates, whereas the next largest earthquake was sequentially hindcasted using several existing models from the literature. Despite the close proximity of the two stimulation campaigns (500 m apart), they led to fundamentally different parametrization of most of the model-parameters. As a result, the models derived from the first stimulation could not be reliably applied to the second stimulation campaign, negating the use of a-priori seismic hazard analysis tools in the planning of a neighboring stimulation. In terms of real-time forecasting of the next largest magnitude, the applied models produced a wide range of magnitude estimates. Moreover, the latter estimates were sometimes inconsistent between the two stimulations, with the best performing model in 2018 being the worst performing one in 2020. The observed modeling discrepancies were attributed primarily to differences in hydraulic energy, to geological and tectonic variations within the reservoir, and potentially to the variations in the injection protocol.

Keywords

enhanced geothermal systems, maximum magnitude, induced seismicity, seismogenic index, hydraulic stimulation, seismic injection efficiency

Introduction

Enhanced Geothermal Systems (EGS) inject cold water into hot underground reservoirs, with the resulting hot water being then pumped up to the surface and used to power a turbine or a binary power plant system to generate electricity and/or heating. They are called “enhanced” because low-permeability reservoirs require man-made fracture networks to be established beforehand to enable the stable circulation of fluids between the injecting and producing well. The popularity of EGS have been increasing worldwide in recent decades, since they are considered a source of clean renewable base-load power and in principle they are deployable in a wide range of geological settings. The formation of EGS reservoirs, via permeability enhancement, can be accomplished through either chemical, thermal or hydraulic stimulation. The latter are the most common and considered the most seismogenic. During hydraulic stimulation, high-pressure fluids are pumped into the rock mass creating new fractures and thus fluid pathways, leading to the permeability enhancement and occurrence of associated microseismicity. Evolving pore fluid pressure interacts as well with existing nearby faults or fracture networks leading to earthquakes of potentially significant size. Large seismic events associated with development of EGSs display a negative socio-economic impact, posing a risk to the local infrastructure as well as safety of people, undermining the public acceptance of pending and future EGS projects (Giardini, 2009; Trutnevyte and Wiemer, 2017). In particular, seismicity from reservoir stimulations has contributed to the early termination of at least four EGS projects, e.g. in Basel (Häring et al., 2008; Bachmann et al., 2011), Pohang (Ellsworth et al., 2019), and in rare cases to building damages, injuries and even one death (Pohang M_W 5.6). Therefore, the estimation of both the rate and amplitude of the generated seismicity is crucial for risk mitigation.

Assessment of the maximum possible magnitude or the next sequentially larger magnitude during fluid injection operations can be done in a probabilistic or deterministic way, typically using physics-based concepts. For example, the next record-breaking technique is deterministic, but based solely on seismic catalogs (Cooke, 1979), without any physics-based input. Probabilistic technique of (van der Elst et al., 2016) uses the seismogenic index concept (Shapiro et al., 2010). The method relies on the Gutenberg-Richter (G-R) Magnitude-Frequency Distribution (MFD) and the Poisson assumption. Deterministic methods uses physics-based concepts and combine information from the seismic catalog (e.g. spatio-temporal evolution of seismicity), industrial parameters (such as fluid volume, injection rate or injection pressure), and reservoir geomechanical properties, to name a few, into the assessment of the upper limit to the maximum magnitude (e.g. Shapiro et al., 2013; Hallo et al., 2014; McGarr, 2014; Galis et al., 2017; Li et al., 2022, see Supplementary Text S1). However, existing physical limits to specific model applicability, arbitrary choices regarding selection of the model parameters (Kwiatek et al., 2015), as well as uncertainties related to estimation of model parameters may substantially bias maximum magnitude estimates. Structural inhomogeneities frequently observed within one reservoir (distinct faults, fracture networks, varying lithologies, limits to formation thickness) may lead to spatio-temporal variations in the partitioning of input hydraulic energy into the seismic and out-of-the seismic band processes (Goodfellow et al., 2015; Bentz et al., 2020). This can invalidate modeling assumptions, such as time-invariance of the MFD parameterization (cf. discussion in Igonin et al., 2018; Kozłowska et al., 2018), or space-invariance of the energy partitioning, that is typically represented with time-dependent quantities such as the seismogenic index (Shapiro et al., 2010) or seismic efficiency factor (Hallo et al., 2014). In an applied real-time forecasting scenario, understanding the relations between the structural complexity of the reservoir and its effects on the modeling parameters is then of crucial importance.

Exploitation of an EGS often involves drilling multiple wells beforehand that will serve as heat exchangers. As the earthquakes are proxy for the damage in the crust (e.g. Main, 1991), detected clusters of microseismicity associated with fluid injection are frequently guiding the drilling of the second well, which then will be also stimulated enabling hydraulic communication between the (doublet) wells. In some geothermal sites, hydraulic stimulations have been performed in adjacent wells, e.g. Soultz-sous-Forets, France (Charléty et al., 2007), Cooper Basin, Australia (Baisch et al., 2006; Hogarth and Heinz-Gerd, 2017), and Helsinki, Finland (Rintamäki et al., 2021; Kwiatek et al., 2022b). Accordingly, there is an intrinsic tendency to extrapolate the developed hazard assessment procedures and traffic light systems from one stimulation to the other assuming that the key seismo-mechanical modeling parameters are very similar. However, it is important to understand under which conditions this assumption and extrapolations are valid.

In this study, we analyze the seismic data (Leonhardt et al., 2021a; Kwiatek et al., 2022a) collected from two stimulation campaigns performed in 2018 and 2020 in Helsinki, Finland, in the frame of St1 Deep Heat project (Ader et al., 2019; Kwiatek et al., 2019, 2022b; Hillers et al., 2020; Leonhardt et al., 2021b; Rintamäki et al., 2021; Holmgren et al., 2023). For both campaigns, we use simple physics-based semi-empirical models to hindcast the evolution of the seismicity rate, the magnitude frequency distribution, and the next largest earthquake. We find how fine-grain changes in the reservoir structural inventory, common (yet crude) modeling assumptions and changes in the hydraulic input may lead to misleading interpretations even in the apparently simple case of pressure-control induced seismicity associated with two adjacent hydraulic stimulations.

Data and methods

Overview of site and stimulations campaigns

The St1 Helsinki site consists of two wells OTN-3 and OTN-2 (Figure 1). The deeper OTN-3 well reached 6,100 m b.s.l. Last 1,000 m of the well was an open-hole dipping 45° towards the NE. Well OTN-2 located ca. 500 m NW from OTN-3 was drilled parallel to the OTN-3 reaching the depth of 5,765 m b.s.l. The bottom hole section of OTN-2 well started at 4.9 km depth.

The seismic monitoring network in both stimulations constituted of 12 3-component 4.5 Hz geophones located in boreholes of 0.3 - 1.1 km depth, located 0.4 – 11 km away, surrounding the project site to ensure azimuthal coverage. This network was complemented with a *borehole array* of up to 12 3-component 15 Hz geophones installed in vertical portions of OTN-2 (2018 stimulation) or OTN-3 (2020 stimulation) well at approximately 2.7 km depth (see Kwiatek et al., 2019, 2022b for details).

Between June 2018 and August 2018, a massive stimulation campaign was performed over 60 days in the inclined portion of the OTN-3 well in 5 stages, separated with inflatable packers (Kwiatek et al., 2019). A total volume of 18,160 m^3 of water was injected over 50 days into the crystalline basement rocks to create the reservoir around the bottom part of OTN-3 well. The stimulation was flow-rate controlled with varying injection rates 400-1200 l/min and wellhead pressures reaching 95 MPa. The injection was performed in a quasi-cyclic manner, where the fluid injection performed at constant rates were alternated with resting periods. Towards the end of stimulation, the resting periods were progressively elongated responding to enhanced seismic hazard. Second stimulation campaign was performed in May 2020 over 16 days in the open-hole section of OTN-2 well (e.g. Rintamäki et al., 2021; Kwiatek et al., 2022b), but the active fluid injection was maintained only for half of the time. A total of 2,875 m^3 of water (16% of that injected in 2018) was injected to establish communication between the two wells. The maximum wellhead pressure did not exceed 70 MPa with injection rates kept at a low level of 400 l/min. Later phases of the 2020 injection were characterized by a repetitive pattern of ~1.5 hr intervals of constant rate injection, followed by 1.5 hr of resting period.

The full seismic catalog of the 2018 stimulation contains 55,707 detected and 5,456 located events (with observed maximum moment magnitude M_W=1.7) originating from the direct vicinity of the stimulated volume of rocks (Leonhardt et al., 2021a). The catalog of the 2020 OTN-2 stimulation consists of 6318 detected and 72 located events (maximum observed M_W=1.3 (Kwiatek et al., 2022a).

The seismicity of 2018 and 2020 shared common features (see Kwiatek et al., 2022b and references therein). The radiated seismic energy and seismicity rate evolved following the hydraulic energy rate with a short time lag. The seismic activity tended to cease within 1 week following the shut-in phases. The temporal evolution of the maximum observable magnitude was found to qualitatively follow pressure-controlled models (e.g. van der Elst et al., 2016; Galis et al., 2017; see also discussion in Bentz et al., 2020). No signatures of runaway behavior were observed. However, the two seismicity datasets display visible differences as well. The seismic injection efficiency, the ratio of seismic-to-hydraulic energy (Maxwell, 2011; Goodfellow et al., 2015) of 2018 stimulation is approximately 3 times larger than that observed in 2020 stimulation (see Fig. 3c in Kwiatek et al., 2022b). The staged 2018 OTN-3 stimulation formed four major clusters along the last 1-km long portion of the injection well that expanded during the stimulation following the diffusion law (e.g. Shapiro et al., 2002, see also Leonhardt et al., 2021b), whereas the open-hole section 2020 stimulation of OTN-2 well led to a single dominant cluster that did not give clear signatures of spatial expansion.

Data analysis

For the purpose of this study, the local "Helsinki" magnitude from input catalogs (Leonhardt et al., 2021a; Kwiatek et al., 2022a) has been first converted to seismic moment and then to moment magnitude (see e.g. Kwiatek et al., 2022b for details). The magnitude of completeness for both seismic catalogs has been selected to limit already identified effects related to day-night background noise variations and transient noises related to injection operations (Kwiatek et al., 2019, 2022b). As shown in Kwiatek et al. (2022b), ignoring temporal variations in completeness may lead to a seismicity that significantly deviates from the non-stationary Poissonian process. To suppress undesirable effects related to varying completeness, we selected very conservative bounds of $M_{W,C}$=-0.5 and $M_{W,C}$=-0.8 for the entire seismicity catalog of 2018 and 2020 stimulation campaigns, respectively, leading to selection of N=24,296 and N=2772 earthquakes for the analysis. The lower magnitude of completeness reached in the 2020 stimulation was due to the lower overall noise level of the pumping system visibly reducing the detection threshold, as well as improved AI-aided processing techniques (see Kwiatek et al., 2022b for details).

The *b*-value was estimated based on maximum likelihood statistics (assuming a Poisson process), and by bootstrapping 1000 catalog-samples to account for measurement/conversion uncertainties behind the cataloged magnitudes (assumed to have a normally distributed standard deviation of 0.2). The magnitude binning interval was 0.1. The *b*-value was computed following two different assumptions for the maximum size of an arrested rupture. First, we assumed an unbounded maximum magnitude, in line with the classic formulation of the Gutenberg–Richter (G-R) relation, and we employed the regression by Weichert (1980). Then, following the finite-volume (FV) formulation of Shapiro et al. (2013), we assumed that a rupture can nucleate only within the stimulated rock-volume and cannot propagate outside of it. This applies a geometrical constraint on the size of any rupture. As a proxy for the stimulated rock-volume we use a fitted ellipsoid around the evolving seismicity cloud (e.g. Kwiatek et al., 2015) to represent the expanding triggering front during pore-pressure diffusion (Supplementary Text S1).

The seismicity rate was modeled using semi-empirical models from the literature. During increasing or stable injection rates, we used the Seismogenic Index (SI) model (Shapiro et al., 2010, p.20; Mignan et al., 2021), while during decreasing injection rates we used the modified Omori decay function (Langenbruch and Shapiro, 2010). The SI model was originally developed for hydraulic fracturing stimulations (Shapiro, 2015), but has been generalized to any type of fluid injection operation (Grigoratos et al., 2020, 2022). The original SI model is a modified version of the G-R relation. The simplified physics-based formula predicts that the total number of induced events (above the magnitude of completeness, M_C) is proportional to the injected fluid volume, which is considered a proxy for pore-pressure perturbation. The ratio of proportionality is then governed by the parameter Σ, the Seismogenic Index. The SI model is a 1D point-source model applicable to non-decreasing pore-pressure conditions, and it does not consider poroelastic stress transmission (Segall and Lu, 2015) or earthquake interactions. When the injection rate drops significantly, during cyclical injection or after shut-in, the decay rate of seismicity can be approximated by the Omori law (Langenbruch and Shapiro, 2010), which originally describes the decay rate of aftershock activity after tectonically driven earthquakes (Omori, 1894). The law states that the number of aftershocks (N) in a given time period (t) after the main shock is proportional to t^{-p}, with common values for p ranging around 1. For induced seismicity, this decay function depends on the fracture strength, with p normally being larger than 2 (Langenbruch and Shapiro, 2010).

The calibration for the parameters Σ and p was performed using maximum likelihood statistics while assuming a Poisson process using the earthquake catalog and the injection rate as input data. For Σ in particular, the regression can be done either in the time-domain (assuming a pre-fitted b-value) or in the magnitude-domain (jointly maximizing the likelihood for b and Σ). In the former case, the rate of earthquakes above M_C was binned based on 2-hour intervals and the b-value was obtained from the Weichert (1980) regression (b_W). In the latter case, the regression assumed a bounded G-R curve, following the FV formulation (Shapiro et al., 2013), and simultaneously solved for the b-value (b_{FV}), the Seismogenic Index (Σ_{FV}), and the stress drop. The Omori p-value was always fitted in the time-domain.

Finally, we tested how various models from the literature perform at directly estimating or simply constraining the size of the next largest earthquake. Some of these models use as key input the total injected volume (Hallo et al., 2014; McGarr, 2014; Galis et al., 2017; Li et al., 2022), while the others also require calibration of the SI (Shapiro et al., 2013; van der Elst et al., 2016). Notably, the Next Record-Breaking Event (NRBE) only uses the cataloged magnitudes as input (Mendecki, 2016; Cao et al., 2020; Verdon and Bommer, 2021). Further descriptions of all these methods, as well as their governing formulations, can be found in the Supplementary Text S1. We note that all the methods except for van der Elst et al. (2016) and NRBE assume self-arrested ruptures that are more or less contained within the volume of rocks affected by pore-pressure changes and do not exhibit characteristics of overextended runaway ruptures that release predominantly energy accumulated via tectonic strain. Furthermore, all the methods except for Li et al. (2022) and NRBE assume that the MFD follows either a bounded or an unbounded G-R distribution. All the methods that employ the fitted *b*-value used the corresponding value of b_W for that time-step, with the exception of Shapiro et al. (2013) which employs b_{FV}.

Results

Quasi-stationary character of seismicity

The regression for b_W, Σ and p assumes that the seismicity data follow a nonhomogeneous Poisson process, which is often the case for hydraulic stimulations (Langenbruch et al., 2011). The seismicity associated with 2020 stimulation was found to follow a nonhomogeneous (quasi-stationary) Poissonian process modulated by injection rates, with limited temporal clustering or anti-clustering, lack of magnitude correlations, and presence of exponential distribution of interevent times (cf. Kwiatek et al., 2022b). Kwiatek et al. (2019) reported that the 2018 seismicity associated with OTN-3 stimulation displayed limited clustering with ~88% of background seismicity following the Poissonian processes. To complete this analysis we followed Kwiatek et al., (2022b) and selected a subset of the 2018 seismic catalog and analyzed whether it follows a quasi-stationary *Poissonian* process using magnitude correlation and interevent time ratio statistics (Supplementary Text S2). We conclude that both catalogs display properties of nonhomogeneous Poissonian processes (see Figures S1-S2 and Supplementary Text S2).

Calibration of model-parameters

Fig. 1cd shows the G-R magnitude-frequency distributions and *b*-value estimations using the two formulations (unbounded and bounded). For the entire 2018 dataset, the classical G-R curve results in b_{W} = 1.43 and Shapiro's FV formulation results in b_{FV}=1.33 (Fig. 1c). The bounded FV approximation seems to provide a much better fit at larger magnitudes, which matter the most for seismic hazard. The entire catalog from 2020 stimulation is characterized by b_{W}=1.51 and b_{FV}=1.41. This time, the FV approach did not visibly improve the fit mainly because the G-R distribution displays a self-similar behavior. Fracturing of the entire extent of the stimulated rock volume would result in a maximum magnitude $M_Y = 2.1$, a value almost identical to the one obtained for 2018 ($M_Y = 2.0$). This is despite the fact that the 2020 stimulation had a slightly higher b_{FV}, much lower Σ values and much lower observed maximum magnitude. The stability of M_{Y}, i.e. of the projected maximum magnitude from the FV approach, is noteworthy.

Figure 2 shows temporal evolution of the *b*-value and SI (the *p*-value temporal evolutions are shown in the Supplementary Figure S3) for the two stimulation campaigns. The 2018 stimulation resulted in steadily increasing *b*-value during the first two weeks from b_W=1.2 to b_W=1.4 (Fig. 2a). This is followed by a stationary period until the end of the stimulation and regardless of the employed regression method. For the 2020 stimulation, the *b*-value started from high values (b_W=1.8, Fig. 2b) and rapidly decreased in the second phase of injection starting around May 13th. Like in the first stimulation, the *b*-value stabilized in the second half of the 2020 injection campaign. When looking at the finite-volume regression of the *b*-value, b_{FV} yields 0.1 to 0.4 lower values in both years, with the largest deviations coming at the start of the stimulation campaigns. Notably, in 2020, b_{FV} needed a lot less data to converge to the eventually stable *b*-value of around 1.4.

The Σ values obtained from the FV constraints of Shapiro et al. (2013) and the classical time-series regression were not identical, neither for 2018 (Σ_{FV}=-0.7; Σ_t=-0.4), nor for 2020 (Σ_{FV}=-1.4; Σ_t=-1.1) stimulations (Fig. 2cd). This highlights that the obtained estimates of the SI are sensitive to the paired *b*-value, which is in turn sensitive to the upper bound constraints of the MFD. Overall, the fitted values of Σ_{FV} were a little more stable in time, compared to those for Σ_t, as expected. Remarkably, the 2020 Σ values were overall much smaller than those observed in 2018. This is in agreement with the observable discrepancy in seismic injection efficiency between the two stimulations (Kwiatek et al., 2022b), suggesting that the seismic process was somewhat less efficient (Maxwell, 2011; Goodfellow et al., 2015; Kwiatek et al., 2018) in the case of 2020 stimulation.

The Omori's law p-value ranged between p=1 and p=2.5 during the first three weeks of the 2018 stimulation (Fig. S3). Fluid-injection driven sequences usually have $2<p<10$ (Langenbruch and Shapiro, 2010). The higher this p-value is, the faster the decay of the seismicity rate. In the end, both stimulations eventually converged to a p=1, a typical value for tectonic sequences (Utsu et al., 1995; Schmid and Grasso, 2012), which is notable and somewhat unexpected. It implies that (given enough time) the entire perturbed volume reached a steady-state of increased yet somewhat consistent stress levels.

Hindcasting seismicity rates and maximum magnitude

Having the key seismic catalog-derived parameters established for both stimulation campaigns, we start with the calibration of our seismicity rate model at 2-hour increments using the available data from the beginning of the stimulation on June 4th, 2018 and until about half of the stimulation period (Figure 3a, gray area). In the following, we hindcasted the remaining period with the injection rate acting as a known input variable. To the best of our knowledge, this is the first time that the Seismogenic Index model and the modified Omori-decay law have been applied to cyclical injection rates. Qualitatively, we find the model performance for 2018 stimulation very good, as it replicates well the amplitudes of the seismicity rate, both during increasing and decreasing injection rates.

For the 2020 stimulation, we followed the exact same procedure and first calibrated our seismicity rate model through May 13th, 2020 (Fig. 3b) and then hindcasted the remaining days in a forward looking way. This time, the qualitative performance of the simulation visibly degraded with respect to that from the 2018 stimulation. First, the simulated time-series tends to underestimate the observed seismicity rates. Second, the simulated rates do not seem to align well with the peaks of the observed seismicity rates, because the latter do not coincide with the local peaks in the injection rate. This results in a highly time-dependent Σ_t value for 2020, and thus in a poor forecasting performance. For this reason, we would favor the use of Σ_{FV} and b_{FV} (instead of Σt and bW) parameterization for this hindcasting exercise (cf. Figure S4).

Figure 4 shows the evolution of the maximum observed magnitude and sequentially updated estimates of the next largest earthquake from different models from the literature. For the 2018 stimulation (Fig. 4a), the best performing models (minimal yet positive deviation) were the ones by Hallo et al. (2014), the NRBE and by van der Elst et al. (2016), the latter using default formulation. Both McGarr (2014) and Galis et al. (2017), the latter with γ derived from Σ (see Supplementary Text S1) largely overestimated the size of the largest event. Shapiro et al. (2013) performed well overall, but underestimated the most crucial record magnitude by about 0.4 units. Finally, the approach of Li et al., (2022) did not yield realistic results, indicating that the assumptions made by the authors were not fully applicable to this stimulation. For the 2020 stimulation, the best performing models were the NRBE, the version of van der Elst et al. (2016) with the lowered probability of exceedance and McGarr (2014). Shapiro et al. (2013) and the default formulation of van der Elst et al. (2016) performed generally well, but they underestimated the most crucial record magnitude jump around May 12th by about 0.6 units. In contrast to 2018, Hallo et al. (2014) delivered persistent and large underestimation of the maximum magnitude. Finally, Li et al. (2022), after an initial reasonable assessment of the maximum magnitudes for the first part of the stimulation campaign broke down in the second part of the stimulation.

Discussion

As part of the St1 Deep Heat project, two hydraulic stimulation campaigns were carried out in two neighboring wells in the Helsinki suburban area in 2018 and 2020. They resulted in transient permeability enhancement and no established communication between the wells due to heterogeneous conductivity of the formations (Kukkonen et al., 2023). Each stimulation led to a stable (pressure-controlled) seismic activity that could be managed by changing the injection procedures (see discussion in Kwiatek et al., 2019, 2022b). The seismicity responded to injection operations with the hydraulic energy being proportional to the seismicity rates. It displayed the prevailing properties of the background seismicity that are well described by a non-stationary *Poisson* process. In absence of significant *b*-value trends (cf. Fig. 2ab), limited earthquake-to-earthquake interactions, and with injection into a distributed network of pre-existing fractures (Leonhardt et al., 2021b), the seismic hazard was controlled by the induced seismicity rates without the signatures of runaway behavior. This stable behavior of the St1 reservoir is different from other cases that display unstable (runaway) behavior associated with a combination of different factors, e.g. existing or emerging structural features (e.g. pre-existing major faults, see McGarr, 2014; Martínez-Garzón et al., 2018; Ellsworth et al., 2019), or earthquake-to-earthquake interactions (e.g. Schoenball et al., 2012; Brown and Ge, 2018; Shen et al., 2021; Verdecchia et al., 2021). It would be therefore expected that the seismicity associated with the two stimulations at the St1 Deep Heat makes a presumably ideal case to train, calibrate, test and optimize models of (adaptive) traffic light systems, as well as probabilistic and deterministic seismic hazard assessment tools. However, this study highlights distinct differences and difficulties in constraining the model parameters, and in forecasting the seismicity rates and the next sequentially largest magnitude.

Despite the fact that the two stimulation campaigns have been performed only approximately 500 m apart, key modeling parameters such as the Seismogenic Index (Σ) and at times the *b*-value (b_W) were very different. Consequently, the relative performance of the models trying to capture the next largest magnitude was also very unstable (Fig. 4). For example, the best performing model in 2018, was the worst performing one in 2020. Only the NRBE and Shapiro et al. (2013) performed consistently across both stimulations. This means that using the first stimulation to fix relative model-weights (on a logic tree basis) to aid forecasting of the subsequent stimulations is not sustainable.

The discrepancy in fitted Σ values between the two stimulation campaigns (cf. Fig. 2cd), could be attributed to the different stimulation protocol applied. The staged stimulation performed in 2018 should result in higher stress concentration near isolated portions of the well in comparison to the 2020 injection, where the stress would be distributed more uniformly along the open-hole section of the well. However, in the 2018 stimulation the isolation of stages was in fact not fully successful, as shown by a near-simultaneous activation of multiple clusters trending SSW-NNE after the injection began. In addition, the 2020 seismicity was in fact confined to a limited portion of the open-hole section of the well. Moreover, lack of significant interevent triggering in both seismicity catalogs, as well as ambiguous directivity of larger events (cf. Holmgren et al., 2023) suggest overall limited stress perturbation in the reservoir linked to its volumetric character and injection into the distributed network of parallel fractures (Leonhardt et al., 2021b; Kwiatek et al., 2022b). Therefore, it is more likely that the observable differences in Σ between the two stimulations could be related to the fine-grain differences in the structural settings, e.g. locally increased permeability and/or existence of preferably oriented/localized structures rather than the different injection protocol used. Indeed, apart from apparent spatial clusters of seismicity, larger events during 2018 stimulation tended to occur in a narrow SE-NW trending zone (see Fig. S4 in Leonhardt et al., 2021b). Kukkonen et al. (2023) concluded heterogeneous hydraulic conductivity was strongly affected by fracturing and local lithological variations of the rock mass with only a small fraction of the natural fractures open and conductive. This as well supports that within broader damage zones activated in the 2018 stimulation there might be more localized structures enhancing Σ and hosting larger events.

The arbitrary selection of the G-R relation may have severe consequences in the probabilistic seismic hazard assessment. While comparing the two different formulations for the G-R relation, it becomes immediately clear that seismicity from both stimulations display predominantly self-similar behavior at lower magnitudes (M_W<0.8) and eventually quite similar *b*-values (Fig. 1cd). However, the 2018 catalog presents a roll-off of the G-R relation at higher magnitudes that is not visible in the 2020 data, likely because the 2020 magnitudes did not grow large enough for this to be observable due to the lower seismicity rates. Notably, the finite-volume formulation of Shapiro et al. (2013) performs very well even when the 2020 data indicate a linear trend (Fig. 1d), and is remarkably able to identify the same potential roll-off magnitude (M_Y) as in 2018. Bootstrapping for magnitude uncertainty provides benefits to this analysis, distinguishing linear from non-linear trends via noise-reduction (Fig. 1d, Figure S5). We encourage wider adoption of this bootstrapping process. The use of bounded G-R model for the analyzed seismicity is also explainable in the context of limited extends to the fluid-perturbed zone originating from spatially varying fracturing and lithology (cf. Kukkonen et al., 2023), i.e. earthquakes cannot exceed the size of discrete damage zone, or with respect to the limited faults strength and lack of significant tectonic stresses (faults cannot slip beyond certain magnitude as they run out of the energy (cf. Kwiatek et al., 2019).

The high *b*-value observed at the beginning of the 2020 stimulation may be of physical origin, suggesting low level of stress at the reservoir (b-value as stress-meter, e.g. Scholz, 1990; Schorlemmer et al., 2005). This could originate from a significant M_L 1.2 earthquake that occurred two weeks before the 2020 stimulation during engineering operations in the well (see observations in Kwiatek et al., 2022b). Following this argument, low *b*-values observable during the initial days of the 2018 OTN-3 stimulation may be a signature of pre-existing (tectonic) stresses in the previously un-stimulated reservoir. Regardless, both stimulations resulted in fairly comparable *b*-values towards the end of stimulation campaigns. However, fundamental differences in Σ values between the two stimulations did not allow using training data from 2018 stimulation to forecast the seismicity rate 500m away for the 2020 stimulation. This is despite a very good performance of the seismicity rate model during the 2018 campaign (Fig. 3a). The 2020 stimulation was characterized by unexpected lags between the local peaks in the injection and seismicity rates and very unstable Σ values regardless of the fitting domain (Σ_t or Σ_{FV}).

Modeling of the next largest sequential earthquake magnitude led to an extreme range of forecasted values (Fig. 4). That said, accounting for the data-driven and time varying *b*-value led to improved performance for the McGarr (2014) model, as expected. Using a fixed *b*-value of 1 should be avoided, since it leads to overly conservative estimates. The modifications that Hallo et al. (2014) applied to McGarr (2014) worked very well in 2018, but failed completely in 2020.

Simplified, yet practical, formulation of Hallo et al. (2014) aimed to address the partitioning between seismic and aseismic processes by using Seismic Efficiency Factor (SEF) (Supplementary Text S1) calibrated from seismic data. Despite identical procedures applied in both stimulations, the 2020 yielded SEF values over 1 order lower than that from 2018. This led to a severe underestimation of the maximum magnitude, pointing out to the intrinsic difficulties in SEF assessment from the seismic data. The modifications that Li et al. (2022) applied to Hallo et al. (2014) were somewhat incompatible with both stimulation campaigns and should be treated with great caution. Finally, the default probability of exceedance behind the model of van der Elst et al. (2016) worked well in 2018, but was not conservative enough in 2020.

The finite volume model of Shapiro et al. (2013) performed better as an alternative G-R formulation than as a real-time upper limit on the next largest earthquake. Interestingly, regardless of the stimulation campaign and the total amount of fluid injected, the maximum magnitude was eventually capped at around M_W 2.0 (Fig. 1cd), which coincided with the red alert setup of the traffic light system during the 2018 (Ader et al., 2019) and the following 2020 stimulation. Although the model failed in constraining the upper limit to the next maximum magnitude in real-time, it overall provided reasonable constraints on maximum size of event in the long run. The latter supports the use of relation between the spatial extension of the activated fracture network to calculate the general constraints on the largest earthquake. It also suggests that the total volume of fluid injected alone is a suboptimal field parameter for constraining the upper limit to the maximum magnitude in the St1 Helsinki project, favoring application of spatial proxy parameters (as in Shapiro et al., 2013) or perturbed volume/pore pressure increase (see eq. 1 in McGarr, 2014, p.2), as discussed e.g. for The Geysers field (Kwiatek et al., 2015). However, an additional downside of Shapiro's finite-volume approach is that it considers a single ellipsoid cluster of seismicity, which is clearly not true for the 2018 stimulation (Fig. 1a). Furthermore, the refined hypocenters we used were the result of heavy post-processing; during real-time applications the extent of the seismicity cloud is poorly constrained and highly uncertain.

The application of Galis et al. (2017) formulation with the key parameter γ derived from Σ (Supplementary Equation S11) led to a significant overestimation of the maximum magnitude in both stimulations, clearly because it has as prerequisite setting the *b*-value equal to 1. However, the temporal evolution of maximum observable magnitude closely followed Galis' model when γ is calculated from four overall poorly constrained geomechanical and geometrical parameters (Kwiatek et al., 2019). We conclude that deriving γ from Σ can lead to overly conservative results if the real *b*-value is much larger than 1.

Surprisingly, the best performing model across both stimulation campaigns was the one that did not utilize any hydraulic data, and only used the jumps between successive record-breaking magnitudes as input. This implies that none of the other approaches has modeled the injection input in a consistently optimal way. Another advantage of the NRBE is that it does not assume any underlying statistical distribution (e.g. Poisson process or G-R curve), nor does it exclude the formation of runaway ruptures. We would encourage further utilization of this method by the scientific literature.

Conclusions

This study combined the Seismogenic Index model and the modified Omori-decay law to forecast the induced seismicity rates of two hydraulic stimulation campaigns performed in the frame of St1 Deep Heat project in Helsinki, Finland. This is the first time that this framework was applied to a cyclical injection protocol. Furthermore, we also tested the performance of 8 existing models in capturing the magnitude of the next largest sequential earthquake.

1. Despite the fact that the two hydraulic stimulations were performed ca. 500 meters apart, they resulted in different seismic responses attributed primarily to the contrasting hydraulic energy rates, but also to the fine differences in the structural inventory of the reservoir, and potentially to variations in the injection protocol. These led to difficulties in the assessment of key seismic hazard parameters such as the Seismogenic Index, the *b*-value and the boundary conditions of the magnitude-frequency distribution. As a result, simple extrapolation of model parameters and assumptions from one stimulation to the other was impossible, negating the use of a-priori seismic hazard and risk analysis tools in the planning of the second stimulation campaign. We conclude that real-time monitoring and modeling of induced seismicity remains a necessity.
2. As far as the seismicity rate model is concerned, despite a very good performance during the 2018 campaign, the fit during the 2020 stimulation was suboptimal. The 2020 stimulation was characterized by unexpected lags between the local peaks in the injection and seismicity rates, rendering the Σ values very unstable, regardless of the fitting domain. Notably, the finite-volume formulation of Shapiro et al. (2013) for the G-R

distribution performed consistently well in both stimulations, projecting the same potential roll-off magnitude (MY) regardless of the magnitude range present in each dataset. Finally, bootstrapping for magnitude uncertainty provided great benefits to our b-value analysis, despite the large number of events present in our catalogs.

3. As far as forecasting the next largest magnitude is concerned, the models produced a very wide range of outcomes. Furthermore, their relative performance across stimulation-campaigns was inconsistent. For example, the best performing model in 2018, was the worst performing one in 2020. Surprisingly, the best performing model across both stimulation campaigns was the one that did not utilize any hydraulic data, and only used the jumps between successive record-breaking magnitudes as input. This implies that the other approaches are not generalized enough to be able to handle common variations in the injection protocol. Perhaps, using an ensemble approach would yield more stable results. That said, the calibration of their relative weights cannot be performed once a-priori, but rather needs to be dynamically updated in near real-time in a data-driven way.

Data and resources

This study used publicly available data (seismicity catalogs) from Leonhardt et al. (2021a) and Kwiatek et al., (2022a).

Declaration of competing interests

The authors declare no competing interests

Acknowledgements

This work was performed in the frame of the EU HORIZON DT-GEO project (HORIZON-INFRA-2021-TECH-01, project number 101058129).

Bibliography

Ader, T. et al. (2019). Design and implementation of a traffic light system for deep geothermal well stimulation in Finland, *J. Seismol.* doi 10.1007/s10950-019-09853-y.

Bachmann, C. E., S. Wiemer, J. Woessner, and S. Hainzl (2011). Statistical analysis of the induced Basel 2006 earthquake sequence: introducing a probability-based monitoring approach for Enhanced Geothermal Systems, *Geophys. J. Int.* **186**, no. 2, 793–807, doi 10.1111/j.1365-246X.2011.05068.x.

Baiesi, M., and M. Paczuski (2004). Scale-free networks of earthquakes and aftershocks, *Phys. Rev. E* **69**, no. 6, 066106, doi 10.1103/PhysRevE.69.066106.

Baisch, S., R. Weidler, R. Vörös, D. Wyborn, and L. de Graaf (2006). Induced Seismicity during the Stimulation of a Geothermal hfr Reservoir in the Cooper Basin, Australia, *Bull. Seismol. Soc. Am.* **96**, no. 6, 2242–2256.

Bentz, S., G. Kwiatek, P. Martínez-Garzón, M. Bohnhoff, and G. Dresen (2020). Seismic moment evolution during hydraulic stimulations, *Geophys. Res. Lett.* **47**, e2019GL086185, doi 10.1029/2019GL086185.

Brown, M. R. M., and S. Ge (2018). Small Earthquakes Matter in Injection-Induced Seismicity, *Geophys. Res. Lett.* **45**, no. 11, 5445–5453.

Cao, N.-T., L. Eisner, and Z. Jechumtálová (2020). Next record breaking magnitude for injection induced seismicity, *First Break* **38**, no. 2, 53–57, doi 10.3997/1365-2397.fb2020010.

Charléty, J., N. Cuenot, L. Dorbath, C. Dorbath, H. Haessler, and M. Frogneux (2007). Large earthquakes during hydraulic stimulations at the geothermal site of Soultz-sous-Forêts, *Int. J. Rock Mech. Min. Sci.* **44**, no. 8, 1091–1105, doi 10.1016/j.ijrmms.2007.06.003.

Cooke, P. (1979). Statistical inference for bounds of random variables, *Biometrika* **66**, no. 2, 367–374, doi 10.1093/biomet/66.2.367.

Davidsen, J., and G. Kwiatek (2013). Earthquake interevent time distribution for induced micro-, nano- and picoseismicity, *Phys. Rev. Lett.* **110**, 068501, doi 10.1103/PhysRevLett.110.068501.

Davidsen, J., G. Kwiatek, and G. Dresen (2012). No Evidence of Magnitude Clustering in an Aftershock Sequence of Nano- and Picoseismicity, *Phys. Rev. Lett.* **108**, 038501, doi 10.1103/PhysRevLett.108.038501.

Davidsen, J., G. Kwiatek, E.-M. Charalampidou, T. H. W. Goebel, S. Stanchits, M. Rueck, and G. Dresen (2017). Triggering processes in rock fracture, *Phys. Rev. Lett.* **119**, 068501, doi 10.1103/PhysRevLett.119.068501.

Davidsen, J., T. Goebel, G. Kwiatek, S. Stanchits, J. Baró, and G. Dresen (2021). What Controls the Presence and Characteristics of Aftershocks in Rock Fracture in the Lab?, *J. Geophys. Res. Solid Earth* **126**, no. 10, e2021JB022539, doi 10.1029/2021JB022539.

Ellsworth, W. L., D. Giardini, J. Townend, S. Ge, and T. Shimamoto (2019). Triggering of the Pohang, Korea, Earthquake (Mw 5.5) by Enhanced Geothermal System Stimulation, *Seismol. Res. Lett.* **90**, no. 5, 1844–1858, doi 10.1785/0220190102.

Elst, N. J. van der, and E. E. Brodsky (2010). Connecting near-field and far-field earthquake triggering to dynamic strain, *J. Geophys. Res. Solid Earth* **115**, no. B7, doi 10.1029/2009JB006681.

van der Elst, N. J., M. T. Page, D. A. Weiser, T. H. W. Goebel, and S. M. Hosseini (2016). Induced earthquake magnitudes are as large as (statistically) expected, *J. Geophys. Res. Solid Earth* **121**, no. 6, 4575–4590, doi 10.1002/2016JB012818.

Galis, M., J. P. Ampuero, P. M. Mai, and F. Cappa (2017). Induced seismicity provides insight into why earthquake ruptures stop, *Sci. Adv.* **3**, no. 12, doi 10.1126/sciadv.aap7528.

Giardini, D. (2009). Geothermal quake risks must be faced, *Nature* **462**, no. 7275, 848–849, doi 10.1038/462848a.

Goodfellow, S. D., M. H. B. Nasseri, S. C. Maxwell, and R. P. Young (2015). Hydraulic fracture energy budget: Insights from the laboratory, *Geophys. Res. Lett.* **42**, no. 9, 3179–3187, doi 10.1002/2015GL063093.

Grigoratos, I., E. Rathje, P. Bazzurro, and A. Savvaidis (2020). Earthquakes Induced by Wastewater Injection, Part I: Model Development and Hindcasting, *Bull. Seismol. Soc. Am.* **110**, no. 5, 2466–2482, doi 10.1785/0120200078.

Grigoratos, I., A. Savvaidis, and E. Rathje (2022). Distinguishing the Causal Factors of Induced

Seismicity in the Delaware Basin: Hydraulic Fracturing or Wastewater Disposal?, *Seismol. Res. Lett.* **93**, no. 5, 2640–2658, doi 10.1785/0220210320.

Hallo, M., I. Oprsal, L. Eisner, and M. Y. Ali (2014). Prediction of magnitude of the largest potentially induced seismic event, *J. Seismol.* **18**, no. 3, 421–431, doi 10.1007/s10950-014-9417-4.

Häring, M. O., U. Schanz, F. Ladner, and B. C. Dyer (2008). Characterisation of the Basel 1 enhanced geothermal system, *Geothermics* **37**, no. 5, 469–495, doi 10.1016/j.geothermics.2008.06.002.

Hillers, G., T. A. T. Vuorinen, M. R. Uski, J. T. Kortström, P. B. Mäntyniemi, T. Tiira, P. E. Malin, and T. Saarno (2020). The 2018 Geothermal Reservoir Stimulation in Espoo/Helsinki, Southern Finland: Seismic Network Anatomy and Data Features, *Seismol. Res. Lett.* **91**, no. 2A, 770–786, doi 10.1785/0220190253.

Hogarth, R., and H. Heinz-Gerd (2017). Lessons Learned from the Habanero EGS Project, *Geotherm. Resour. Counc. Trans.* **41**, 865–877.

Holmgren, J. M., G. Kwiatek, and M. J. Werner (2023). Nonsystematic Rupture Directivity of Geothermal Energy Induced Microseismicity in Helsinki, Finland, *J. Geophys. Res. Solid Earth* **128**, no. 3, e2022JB025226, doi 10.1029/2022JB025226.

Igonin, N., M. Zecevic, and D. W. Eaton (2018). Bilinear Magnitude-Frequency Distributions and Characteristic Earthquakes During Hydraulic Fracturing, *Geophys. Res. Lett.* **45**, no. 23, 12,866-12,874, doi 10.1029/2018GL079746.

Kozłowska, M., M. R. Brudzinski, P. Friberg, R. J. Skoumal, N. D. Baxter, and B. S. Currie (2018). Maturity of nearby faults influences seismic hazard from hydraulic fracturing, *Proc. Natl.*

Acad. Sci. **115**, no. 8, E1720–E1729, doi 10.1073/pnas.1715284115.

Kukkonen, I. T., P. J. Heikkinen, P. E. Malin, J. Renner, G. Dresen, A. Karjalainen, J. Rytkönen, and J. Solantie (2023). Hydraulic conductivity of the crystalline crust: Insights from hydraulic stimulation and induced seismicity of an enhanced geothermal system pilot reservoir at 6 km depth, Espoo, southern Finland, *Geothermics* **112**, 102743, doi 10.1016/j.geothermics.2023.102743.

Kwiatek, G., P. Martínez-Garzón, G. Dresen, M. Bohnhoff, H. Sone, and C. Hartline (2015). Effects of long-term fluid injection on induced seismicity parameters and maximum magnitude in northwestern part of The Geysers geothermal field, *J. Geophys. Res. Solid Earth* **120**, doi 10.1002/2015JB012362.

Kwiatek, G., P. Martínez-Garzón, K. Plenkers, M. Leonhardt, A. Zang, S. Specht, G. Dresen, and M. Bohnhoff (2018). Insights Into Complex Subdecimeter Fracturing Processes Occurring During a Water Injection Experiment at Depth in Äspö Hard Rock Laboratory, Sweden, *J. Geophys. Res. Solid Earth* **123**, no. 8, 6616–6635, doi 10.1029/2017JB014715.

Kwiatek, G. et al. (2019). Controlling fluid-induced seismicity during a 6.1-km-deep geothermal stimulation in Finland, *Sci. Adv.* **5**, no. 5, eaav7224, doi 10.1126/sciadv.aav7224.

Kwiatek, G., P. Martinez-Garzon, and A. Karjalainen (2022a). Earthquake catalog of induced seismicity associated with 2020 hydraulic stimulation campaign at OTN-2 well in Helsinki, Finland, *GFZ Data Serv.* doi 10.5880/GFZ.4.2.2022.001.

Kwiatek, G., P. Martínez-Garzón, J. Davidsen, P. Malin, A. Karjalainen, M. Bohnhoff, and G. Dresen (2022b). Limited Earthquake Interaction During a Geothermal Hydraulic Stimulation in Helsinki, Finland, *J. Geophys. Res. Solid Earth* **127**, no. 9, e2022JB024354, doi 10.1029/2022JB024354.

Langenbruch, C., and S. A. Shapiro (2010). Decay rate of fluid-induced seismicity after termination of reservoir stimulations, *GEOPHYSICS* **75**, no. 6, MA53–MA62, doi 10.1190/1.3506005.

Langenbruch, C., C. Dinske, and S. A. Shapiro (2011). Inter event times of fluid induced earthquakes suggest their Poisson nature, *Geophys. Res. Lett.* **38**, no. 21.

Leonhardt, M., G. Kwiatek, P. Martinez-Garzon, and P. Heikkinen (2021a). Earthquake catalog of induced seismicity recorded during and after stimulation of Enhanced Geothermal System in Helsinki, Finland, *GFZ Data Serv.* doi 10.5880/GFZ.4.2.2021.001.

Leonhardt, M., G. Kwiatek, P. Martínez-Garzón, M. Bohnhoff, T. Saarno, P. Heikkinen, and G. Dresen (2021b). Seismicity during and after stimulation of a 6.1 km deep enhanced geothermal system in Helsinki, Finland, *Solid Earth* **12**, no. 3, 581–594, doi 10.5194/se-12-581-2021.

Li, Z., D. Eaton, and J. Davidsen (2022). Short-term forecasting of Mmax during hydraulic fracturing, *Sci. Rep.* **12**, no. 1, 12509, doi 10.1038/s41598-022-15365-6.

Main, I. G. (1991). A modified Griffith criterion for the evolution of damage with a fractal distribution of crack lengths: application to seismic event rates and b-values, *Geophys. J. Int.* **107**, no. 2, 353–362, doi 10.1111/j.1365-246X.1991.tb00830.x.

Martínez-Garzón, P., I. Zaliapin, Y. Ben-Zion, G. Kwiatek, and M. Bohnhoff (2018). Comparative Study of Earthquake Clustering in Relation to Hydraulic Activities at Geothermal Fields in California, *J. Geophys. Res. Solid Earth* **123**, no. 5, 4041–4062, doi 10.1029/2017JB014972.

Maxwell, S. C. (2011). What does microseismic tell us about hydraulic fracture deformation, *CSEG Rec. - Focus Artic.* **36**, no. 8.

McGarr, A. (1976). Seismic moments and volume changes, *J. Geophys. Res.* **81**, no. 8, 1487–1494, doi 10.1029/JB081i008p01487.

McGarr, A. (2014). Maximum magnitude earthquakes induced by fluid injection, *J. Geophys. Res. Solid Earth* **119**, no. 2, 1008–1019, doi 10.1002/2013JB010597.

Mendecki, A. (2016). *Mine Seismology Reference Book: Seismic Hazard*, Institute of Mine Seismology.

Mignan, A., M. Broccardo, and Z. Wang (2021). Comprehensive Survey of Seismic Hazard at Geothermal Sites by a Meta-Analysis of the Underground Feedback Activation Parameter afb, *Energies* **14**, no. 23, doi 10.3390/en14237998.

Omori, F. (1894). On the Aftershocks of Earthquakes, *J. Coll. Sci. Imp. Univ. Tokyo* **7**, 111–120.

Rintamäki, A. E. et al. (2021). A Seismic Network to Monitor the 2020 EGS Stimulation in the Espoo/Helsinki Area, Southern Finland, *Seismol. Res. Lett.* doi 10.1785/0220210195.

Ritz, V. A., A. P. Rinaldi, and S. Wiemer (2022). Transient evolution of the relative size distribution of earthquakes as a risk indicator for induced seismicity, *Commun. Earth Environ.* **3**, no. 1, 249, doi 10.1038/s43247-022-00581-9.

Schmid, A., and J.-R. Grasso (2012). Omori law for eruption foreshocks and aftershocks, *J. Geophys. Res. Solid Earth* **117**, no. B7, doi 10.1029/2011JB008975.

Schoenball, M., C. Baujard, T. Kohl, and L. Dorbath (2012). The role of triggering by static stress transfer during geothermal reservoir stimulation, *J. Geophys. Res. Solid Earth* **117**, no. B9, doi 10.1029/2012JB009304.

Scholz, C. H. (1990). *The mechanics of earthquakes and faulting*, Cambridge University Press, Cambridge (U.K.).

Schorlemmer, D., S. Wiemer, and M. Wyss (2005). Variations in earthquake-size distribution across different stress regimes, *Nature* **437**, no. 7058, 539–542, doi 10.1038/nature04094.

Segall, P., and S. Lu (2015). Injection-induced seismicity: Poroelastic and earthquake nucleation effects, *J. Geophys. Res. Solid Earth* **120**, no. 7, 5082–5103, doi 10.1002/2015JB012060.

Shapiro, S. A. (2015). *Fluid-Induced Seismicity*, Cambridge University Press, Cambridge.

Shapiro, S. A., E. Rothert, V. Rath, and J. Rindschwentner (2002). Characterization of fluid transport properties of reservoirs using induced microseismicity, *Geophysics* **67**, no. 1, 212, doi 10.1190/1.1451597.

Shapiro, S. A., C. Dinske, C. Langenbruch, and F. Wenzel (2010). Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations, *Lead. Edge* **29**, no. 3, 304–309.

Shapiro, S. A., O. S. Krüger, and C. Dinske (2013). Probability of inducing given-magnitude earthquakes by perturbing finite volumes of rocks, *J. Geophys. Res. Solid Earth* **118**, no. 7, 3557–3575, doi 10.1002/jgrb.50264.

Shen, L. W., D. R. Schmitt, R. Wang, and T. E. Hauck (2021). States of In Situ Stress in the Duvernay East Shale Basin and Willesden Green of Alberta, Canada: Variable In Situ Stress States Effect Fault Stability, *J. Geophys. Res. Solid Earth* **126**, no. 6, e2020JB021221, doi 10.1029/2020JB021221.

Trutnevyte, E., and S. Wiemer (2017). Tailor-made risk governance for induced seismicity of geothermal energy projects: An application to Switzerland, *Geothermics* **65**, 295–312, doi 10.1016/j.geothermics.2016.10.006.

Utsu, T., Y. Ogata, R. S, and Matsu'ura (1995). The Centenary of the Omori Formula for a Decay Law of Aftershock Activity, *J. Phys. Earth* **43**, no. 1, 1–33, doi 10.4294/jpe1952.43.1.

Verdecchia, A., E. S. Cochran, and R. M. Harrington (2021). Fluid-Earthquake and Earthquake-Earthquake Interactions in Southern Kansas, USA, *J. Geophys. Res. Solid Earth* **126**, no. 3, e2020JB020384, doi 10.1029/2020JB020384.

Verdon, J. P., and J. J. Bommer (2021). 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.* **25**, no. 1, 301–326, doi 10.1007/s10950-020-09966-9.

Weichert, D. H. (1980). Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes, *Bull. Seismol. Soc. Am.* **70**, no. 4, 1337–1346, doi 10.1785/BSSA0700041337.

Zaliapin, I., and Y. Ben-Zion (2013a). Earthquake clusters in southern California I: Identification and stability, *J. Geophys. Res. Solid Earth* **118**, no. 6, 2847–2864, doi 10.1002/jgrb.50179.

Zaliapin, I., and Y. Ben-Zion (2013b). Earthquake clusters in southern California II: Classification and relation to physical properties of the crust, *J. Geophys. Res. Solid Earth* **118**, no. 6, 2865–2877, doi 10.1002/jgrb.50178.

Zaliapin, I., A. Gabrielov, V. Keilis-Borok, and H. Wong (2008). Clustering Analysis of Seismicity and Aftershock Identification, *Phys. Rev. Lett.* **101**, no. 1, 018501, doi 10.1103/PhysRevLett.101.018501.

Authors

Grzegorz Kwiatek, Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Section 4.2 Geomechanics and Scientific Drilling, Telegrafenberg, 14473 Potsdam, Germany (kwiatek@gfz-potsdam.de)

Iason Grigoratos, ETH Zurich, Swiss Seismological Service, Sonneggstrasse 5, 8092 Zurich, Switzerland (iason.grigoratos@sed.ethz.ch)

Stefan Wiemer, ETH Zurich, Swiss Seismological Service, Sonneggstrasse 5, 8092 Zurich, Switzerland (stefan.wiemer@sed.ethz.ch)

List of Figure Captions

Figure 1. Overview of injection wells and seismicity associated with hydraulic simulations performed in 2018 (OTN-3) and 2020 (OTN-2). (a): Map view; (b): SW-NE-trending depth section along 45° (SW-NE) azimuth. The double-difference relocated seismicity from 2018 stimulation (Leonhardt et al., 2021b) is shown with circles color-coded with injection phases 1-5 into different sections of the OTN-3 well isolated with inflatable packers (sections are marked with corresponding color). The 2020 injection was performed in the open-hole section of the OTN-2 well (magenta highlight), and the associated double-difference relocated seismicity (Kwiatek et al., 2022b) is shown with magenta circles and squares. The size of symbols reflects earthquake magnitudes. (c-d): Corresponding cumulative (solid black dots) and non-cumulative (open squares) MFDs and G-R fits following Weichert (1980) (solid black lines) and Shapiro et al. (2013) (dashed blue line) for 2018 and 2020 stimulation, respectively.

Figure 2. Temporal evolution of the b-value (a,b) and seismogenic index Σ (c,d) for the two stimulation campaigns. (a,c): 2018 stimulation; (b,d): 2020 stimulation. (a,b): Grey circles and dotted lines represent seismic events and magnitude of completeness used, respectively. (c,d): Black solid line shows the seismicity rate above MC averaged over 2-hour bins. Corresponding temporal evolution of the p-value is shown in Fig. S3

Figure 3. Hindcasted seismicity rates using the developed time-domain models, i.e. the Seismogenic Index (Σt, bW) and Omori's law. Observed and simulated seismicity rates are shown with solid black and dashed dark magenta and red lines, respectively. Flow rate and well head

pressure is shown with dotted blue and brown lines, respectively. Time period used for calibration of the model is shown with a gray background. (a): 2018 stimulation (staged injection), (b): 2020 stimulation (open hole injection), cf. Fig. 1.

Figure 4. Evolution of the maximum observed magnitude and sequentially updated estimates of the next largest earthquake from different models from the literature (see Supplementary Text S1 for details). (a): 2018 stimulation (staged injection), (b): 2020 stimulation (open hole injection).

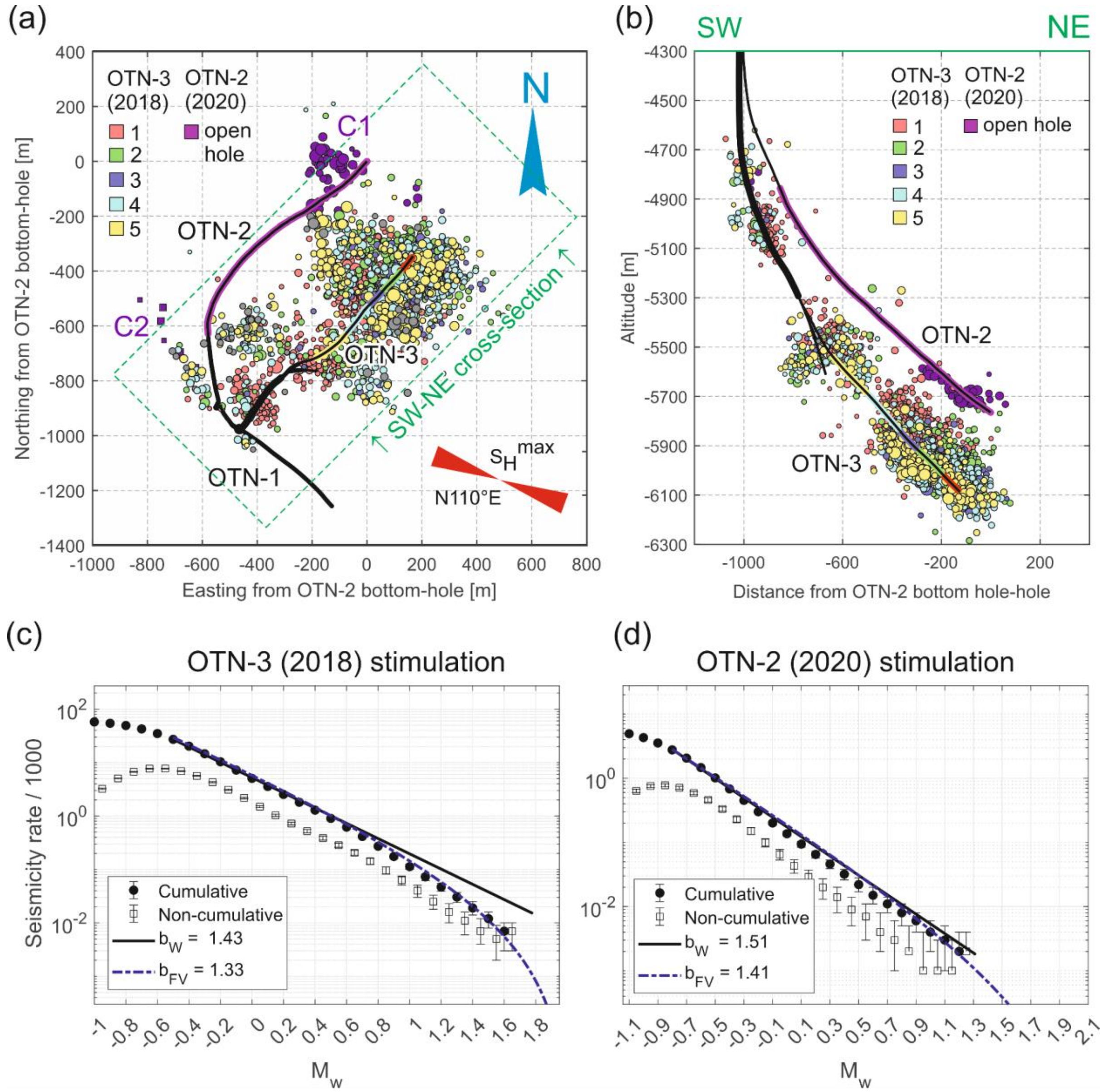

Figure 1. Overview of injection wells and seismicity associated with hydraulic simulations performed in 2018 (OTN-3) and 2020 (OTN-2). (a): Map view; (b): SW-NE-trending depth section along 45° (SW-NE) azimuth. The double-difference relocated seismicity from 2018 stimulation (Leonhardt et al., 2021b) is shown with circles color-coded with injection phases 1-5 into different sections of the OTN-3 well isolated with inflatable packers (sections are marked with corresponding color). The 2020 injection was performed in the open-hole section of the OTN-2

well (magenta highlight), and the associated double-difference relocated seismicity (Kwiatek et al., 2022b) is shown with magenta circles and squares. The size of symbols reflects earthquake magnitudes. (c-d): Corresponding cumulative (solid black dots) and non-cumulative (open squares) MFDs and G-R fits following Weichert (1980) (solid black lines) and Shapiro et al. (2013) (dashed blue line) for 2018 and 2020 stimulation, respectively.

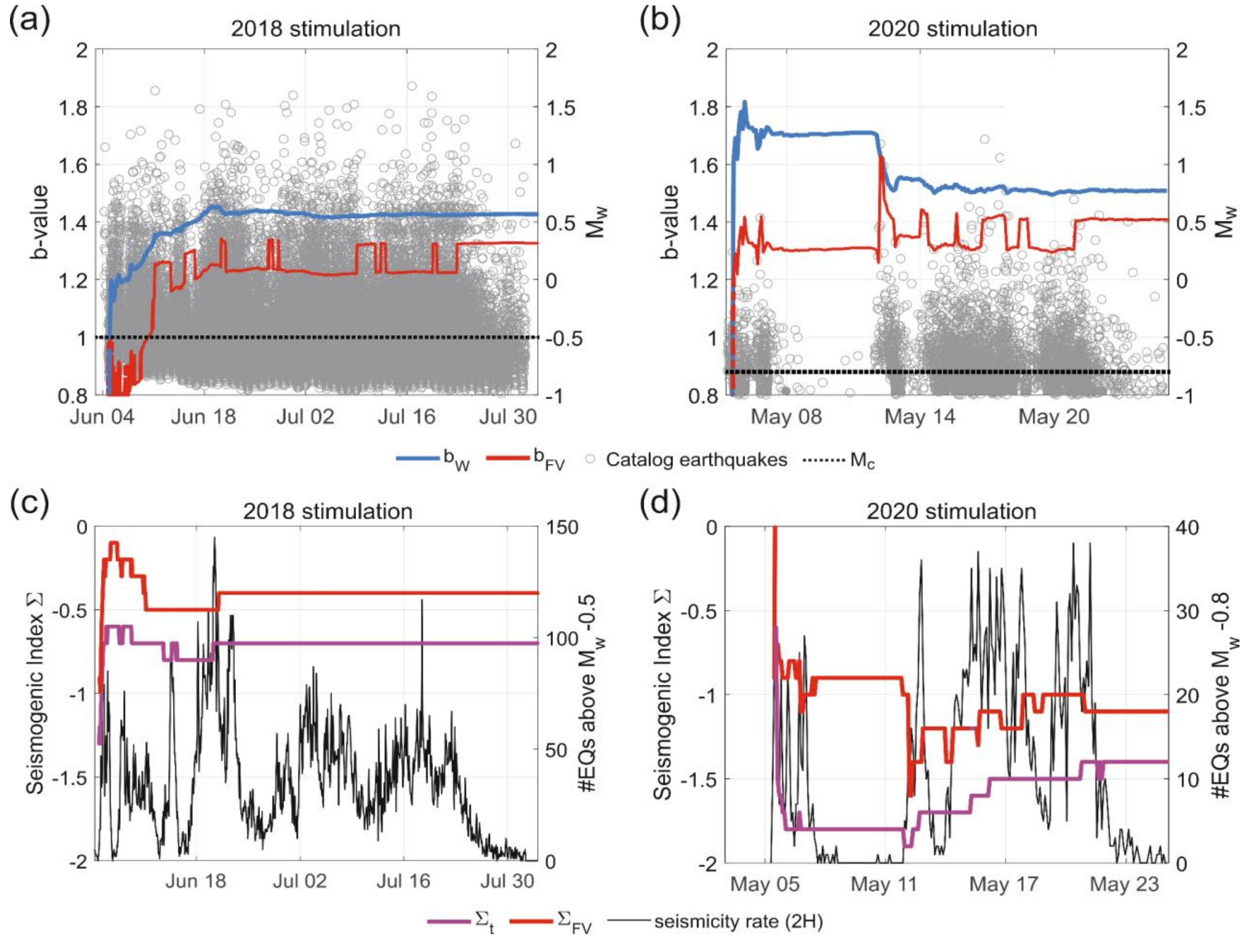

Figure 2. Temporal evolution of the b-value (a,b) and seismogenic index Σ (c,d) for the two stimulation campaigns. (a,c): 2018 stimulation; (b,d): 2020 stimulation. (a,b): Grey circles and dotted lines represent seismic events and magnitude of completeness used, respectively. (c,d): Black solid line shows the seismicity rate above MC averaged over 2-hour bins. Corresponding temporal evolution of the p-value is shown in Fig. S3

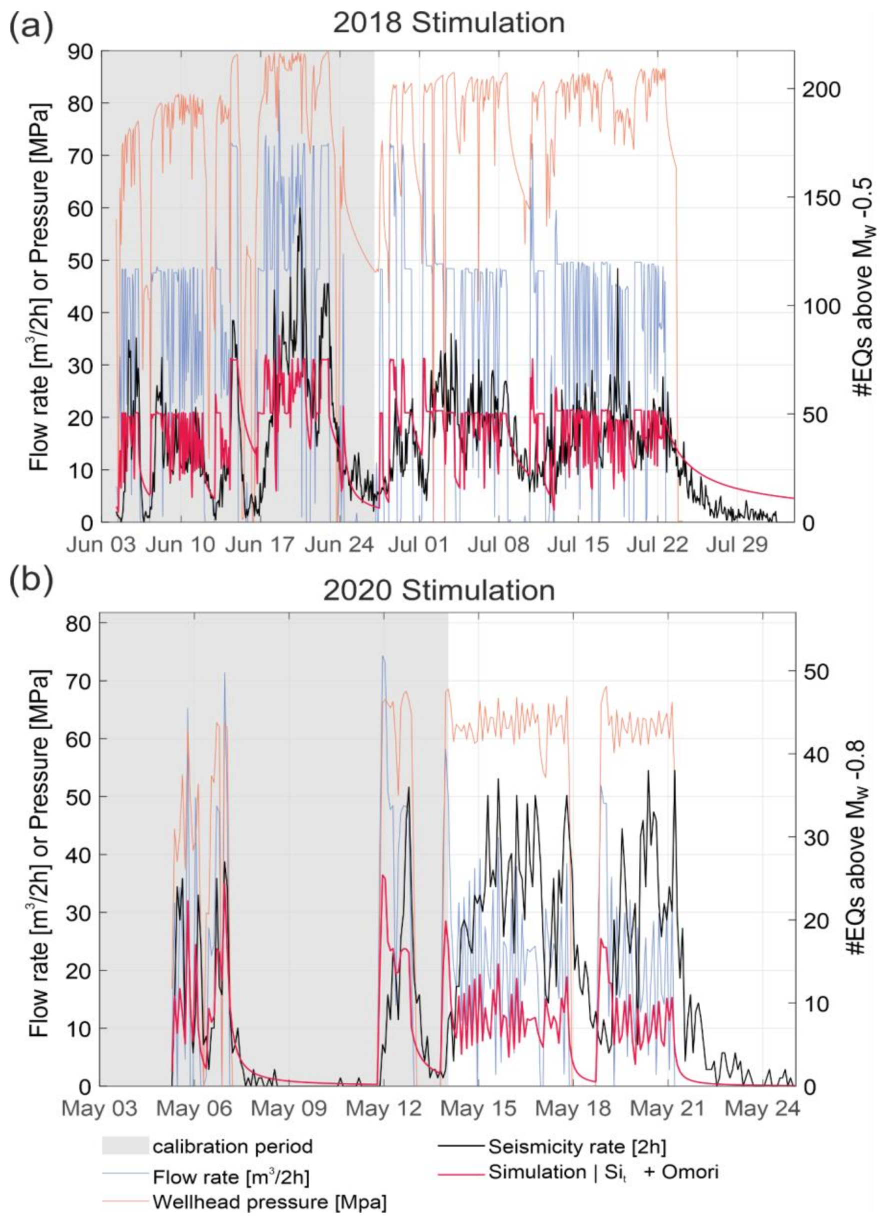

Figure 3. Hindcasted seismicity rates using the developed time-domain models, i.e. the Seismogenic Index (Σ_t, b_W) and Omori's law. Observed and simulated seismicity rates are shown with solid black and dashed dark magenta and red lines, respectively. Flow rate and well head pressure is shown with dotted blue and brown lines, respectively. Time period used for calibration of the model is shown with a gray background. (a): 2018 stimulation (staged injection), (b): 2020 stimulation (open hole injection), cf. Fig. 1.

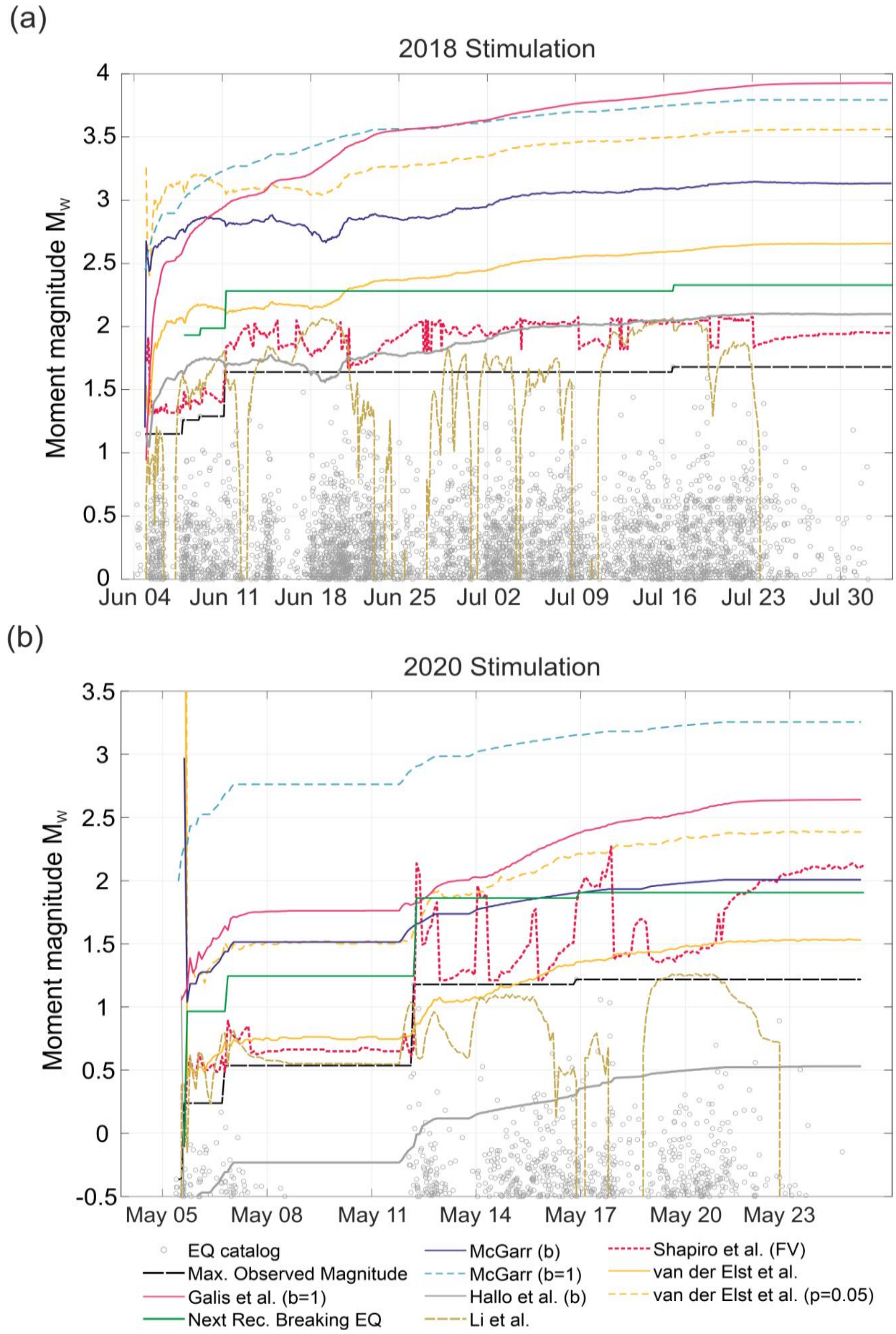

Figure 4. Evolution of the maximum observed magnitude and sequentially updated estimates of the next largest earthquake from different models from the literature (see Supplementary Text S1 for details). (a): 2018 stimulation (staged injection), (b): 2020 stimulation (open hole injection).

Supplementary Information for:

Variability of Seismicity Rates and Maximum Magnitude for Adjacent Hydraulic Stimulations

G. Kwiatek[1], I. Grigoratos[2], and S. Wiemer[2]

1. Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Section 4.2 Geomechanics and Scientific Drilling, Telegrafenberg, 14473 Potsdam, Germany
2. ETH Zurich, Swiss Seismological Service, Sonneggstrasse 5, 8092 Zurich, Switzerland

Content:

Supplementary Text S1: Models forecasting the next largest earthquake

McGarr (1976, 2014) proposed a simple formula to compute an upper bound for the cumulative seismic moment that can be released during a fluid-induced earthquake sequence.

$$\sum M_0 = 2 * G * \Delta V, \tag{S1}$$

where G is the shear modulus, often assumed equal to 30 GPa, and ΔV is the net total injected fluid volume. The approach is based on volumetric changes inducing seismic slip in a linear fashion. It assumes that, on average, each fault patch is about half a seismic stress drop below the yield stress and so it only takes half as much stress change imposed by the volumetric change to induce seismic slip.

Assuming a G-R relation, we can employ the b-value to convert this cumulative seismic moment into a single maximum magnitude, which is given by:

$$M_0{}^{max} = \frac{1-B}{B}\frac{2\mu(3\lambda+2G)}{3}\Delta V, \tag{S2}$$

where $B = 2b/3$. Assuming $\lambda = G$ and μ=0.6 leads to:

$$M_0{}^{max} = G * \Delta V. \tag{S3}$$

Then, the corresponding moment magnitude can be derived as:

$$M_W{}^{max} = \frac{2}{3} log_{10}(M_0{}^{max}) - 6.033. \tag{S4}$$

Modifying McGarr (1976), Hallo et al. (2014) introduced the Seismic Efficiency Factor (SEF) that accounts for partitioning of the elastic strain energy release associated with fluid injection into seismic and aseismic processes. SEF is a free parameter to be calibrated; in the case of a self-arrested rupture, it falls into the range 0 < SEFF ≤ 1. During a given sequence, we always used the maximum computed value for SEF till that time as the current one. Here, we apply Hallo et al. (2014) modification to the formulation of McGarr (2014), and not to the original 1976 equation, hence:

$$\sum M_0 = SEF * 2G\Delta V. \tag{S5}$$

Li et al. (2022) introduced a further modification to the approach of Hallo et al. (2014). They first define as dM0 the difference between the upper limit of the ΣMo (derived assuming Hallo et al. (2014) and the observed ΣMo from the earthquake catalog (at any given time t):

$$dM_0 = 2G * SEF * \Delta V(t) - \sum M_0\,(t). \tag{S6}$$

Then, they assume dMo is stored energy that could potentially be released as residual seismic moment in a single rupture. This results in:

$$M_W{}^{mmax} = (log_{10}(dM_0) - 9.1)/\,1.5. \tag{S7}$$

The main benefit of using dMo is that one does not have to assume a G-R relation or fit a *b*-value to derive the maximum magnitude estimate from ΣMo.

Based on the premise that induced seismicity is Poissonian and follows the G-R distribution, van der Elst et al. (2016) showed that the peak of the posterior probability density function for the expected maximum magnitude can be expressed as:

$$M^{max} = \frac{1}{b}(\Sigma + log_{10}\Delta V), \tag{S8}$$

where *b* is the slope of the frequency-magnitude distribution and Σ is the Seismogenic Index (Shapiro et al., 2010). According to the Poisson statistics, this implies a 63% probability of observing this magnitude value (single occurrence). Of course, we can calculate magnitude values for any Probability of Occurrence (POE) using eq. S7. For comparison, we also show results for a 5% probability.

$$M^{max} = \frac{1}{b}(\Sigma - log_{10}(-ln(1 - poe)\quad /\Delta V)), \tag{S9}$$

Galis et al. (2017) proposed an estimate for the maximum moment M_0^{max} that can be released during an arrested rupture based on the notion that such rupture is controlled by a competition between two sources of elastic energy: injection-induced fluid pressure and tectonic prestress. The contributions of these two sources are both positive. However, the energy contributed by injection-induced fluid pressure decays with increasing rupture size, whereas the energy contributed by tectonic prestress increases, thereby creating a trade-off between these two strain-energy sources. At the maximum arrest size, both contributions are comparable. The value of $M_0{}^{max}$ is dependent on the total net injected volume ΔV with an exponent of 3/2, instead of slope equal to 1 that is characteristic of McGarr (2014) model.

The rest of the formulation includes various geomechanical parameters related to the target-reservoir that can be combined in a single parameter γ:

$$M_0{}^{max} = \gamma * \Delta V^{3/2}. \quad \text{(S10)}$$

Galis et al. (2017), following the rationale of van der Elst et al. (2016), demonstrated that while assuming b=1, $\Sigma = 2/3 log_{10}\gamma - 6.07$. Therefore, in order to estimate the maximum magnitude following Galis et al. (2017), at every iteration *i*, we computed an extra Σ value equal to:

$$\Sigma_\gamma(i) = log_{10}\left(\frac{N(i)}{10^{(-M_c)}\Delta V(i)}\right), \quad \text{(S11)}$$

that assumes a *b*-value of 1. N is the total number of events above Mc.

The so-called “lower-bound” formulation of Shapiro et al. (2013) results in a maximum magnitude equal to:

$$M^{max,L} = log_{10}L^2 + 2/3(log_{10}\Delta\sigma - log_{10}C) - 6.07, \quad \text{(S11)}$$

where *L* denotes a characteristic scale of the stimulated volume, $\Delta\sigma$ is a static stress drop, and *C* is a geometrical constant close to 1. For volumes perturbed by fluid injection that can be surrounded with ellipsoid with principal axes characterized by $L_{min} < L_{int} < L_{max}$, Shapiro et al. (2013) found that:

$$L = \left[\frac{1}{3}\left(1/L_{min}{}^3 + 1/L_{int}{}^3 + 1/L_{max}{}^3\right)\right]^{-1/3} \quad \text{(S12)}$$

often provides a good estimate of the characteristic scale. The values for $L_{min}, L_{int}, L_{max}$ are derived from a fitted ellipsoid based on the seismicity cloud, while Δσ, Σ and b are derived jointly using grid-search maximum likelihood regression (*Poisson* assumption).

The Next Record Breaking Earthquake (NRBE) is a method that estimates the upper bound of the next largest event expected to occur based on a given catalog of earthquakes. It does not rely on any injection data, nor does it assume any magnitude frequency distribution. It only uses the earthquake magnitudes and the magnitude of completeness Mc. It does not even need the hypocenters. It is estimated using order statistics on random variables (Cooke, 1979). First, we compute the jumps in record-magnitude between time-ordered events above Mc. For example, if Mc is 1.5 and the observed magnitudes were 1.2, 1.5, 2, 1.7, 1.9, 2.2, 2.5, 2.1, 3.1, then the jumps were 0.5 [2 - 1.5], 0.2 [2.2 - 2.0], 0.3 [2.5 - 2.2], 0.6 [3.1 - 2.5]. Next, we order the jumps from the smallest difference to the largest. The maximum expected jump ΔM^{max} is estimated to be:

$$\Delta M^{max} = 2 * \Delta M_n - \sum_{i=0}^{n-1} \left[(1 - \frac{i}{n})^n - (1 - \frac{i+1}{n})^n \right] * \Delta M_{n-i} \quad \text{(S13)}$$

where $\Delta M_{i=1:n}$ are the ordered magnitude-jumps. Then, the NRBE value is simply the observed maximum magnitude $M^{(max,obs)}$ plus ΔM^{max}:

$$M^{NRBE} = M^{(max,obs)} + \Delta M^{max}. \quad \text{(S14)}$$

Supplementary Text S2: Clustering properties of seismicity associated with 2018 OTN-3 stimulation.

To investigate induced earthquakes interactions, we followed Kwiatek et al. (2022b) and calculated two additional statistical parameters derived from the representative portion of the seismic catalog associated with the 2018 hydraulic stimulation performed in the OTN-3 well. The statistical measures included interevent time ratio and magnitude correlations and has been discussed in detail (Davidsen et al., 2012, 2017, 2021; Davidsen and Kwiatek, 2013; Kwiatek et al., 2022b).

We selected the catalog above the magnitude of completeness (M_C=-0.5) as used throughout this study, and then removed 12 events with small magnitudes (M_W<0.3) that indicate duplicated events in the original catalog of Leonhardt et al. (2012). In the following, the catalog was constrained between the first event related to the stimulation (04 Jun 2018 05:27 UTC and shut-in of the well (23 Jul 2018 07:30 UTC), i.e. it did not contain the post-stimulation seismicity.

We first tested for potential correlation between the magnitudes of the consecutive earthquakes. Statistically significant correlations between magnitudes in the analyzed catalog suggest that the population is not behaving as randomly drawn for the G-R distribution as is expected from a Poissonian process. The statistics is calculated as:

$$\Delta M = [\Delta M_i] = M_{i+1} - M_i, \tag{S15}$$

where $[\Delta M_i]$ is the origin time ordered vector of magnitude differences. Following Davidsen et al. (2012), the probability density function (PDF), $p(\Delta M)$, built upon empirical ΔM vector is correlated once it significantly deviates from the distribution of magnitude differences which contains uncorrelated magnitudes $\Delta M^{\star}$, $p(\Delta M^{\star})$. The latter PDF can be realized many times by reshuffling the order of magnitudes in the input empirical catalog which destroys any potential correlations. In the following, one can calculate the difference between the cumulative distribution function (CDF) of empirical $p(\Delta M < \Delta m)$ and reshuffled data $p(\Delta M^{\star} < \Delta m)$ (cf. Kwiatek et al., 2022b):

$$\delta p(\Delta M) = p(\Delta M < \Delta m) - p(\Delta M^{\star} < \Delta m). \tag{S16}$$

Magnitudes from empirical catalog will be correlated, if $\delta p(\Delta M)$ deviate from zero baseline for any considered Δm. Statistically significant deviations from random distribution of magnitudes in time may suggest existence of local-in-time accelerations or decelerations of seismic processes that are not expected from the stationary Poissonian process (e.g. accelerated seismic release or aftershock sequences).

In addition, we calculated interevent time ratio statistics (Elst and Brodsky, 2010). This statistic uses the time-ordered origin times of earthquakes $T = [T_i] = T_{i+1} - T_i$:

$$R = [R_i] = (T_{i+1} - T_i)/(T_{i+1} - T_{i-1}). \qquad \text{(S17)}$$

Here, for a (quasi-)stationary Poissonian process, the PDF of interevent time ratios $p(R)$ is expected to follow a uniform distribution. Deviations from the uniform distribution suggest earthquake clustering and anti-clustering in time, which is expressed by peaks of the $p(R)$ close to R=0 and R=1. To measure the statistical significance, we compare whether the empirical distribution $p(R)$ fits into the confidence intervals estimated from multiple realization of data samples built upon input data (with the same number of events as the empirical catalog) that are randomly distributed over time, i.e. following the Poisson process (cf. Kwiatek et al., 2022b). Short-time, statistically significant temporal clustering or anticlustering of seismicity deviating from that expected from quasi-stationary Poisson process may reflect accelerations or decelerations in the seismic process.

Figures S1 and S2 present outcomes of the analysis for the 2018 catalog above the magnitude of completeness. The distribution of observed interevent time ratios (Fig. S2) is statistically indistinguishable from that expected from random distribution of events in time while considering 95% confidence intervals. In addition, the differences in the probability to observe a magnitude difference $M_{i+1} - M_i < \Delta m$ between observed catalog and its randomized versions are not significantly deviating from zero baseline while assuming 95% confidence interval for all Δm. This suggests that the observable magnitudes are behaving as randomly sampled from the G-R distribution. The calculated statistics suggest that the seismicity catalog associated with stimulation of OTN-3 well is displaying properties of quasi-stationary Poissonian processes for time and magnitude space. We note that in Kwiatek et al. (2019), the spatio-temporal clustering method following (Zaliapin et al., 2008; Zaliapin and Ben-Zion, 2013a) have been applied to the bottom-most cluster of induced seismicity revealing that this cluster display very limited clustering in time-distance-magnitude space. The separation of seismicity in time-distance-magnitude space following Baiesi and Paczuski (2004) metrics resulted in 88% background (explainable by Poissonian process) and 12%

clustered seismicity, which is significantly lower than that observed for tectonic seismicity (cf. Zaliapin and Ben-Zion, 2013a, 2013b).

Supplementary Figures

Figure S1: Probability density function (pdf) of interevent time ratios for a subset of the catalog (M_W>M_C, M_C=-0.5) associated with the 2018 stimulation (see Supplementary Text S2 for details). Magenta areas correspond to 68% and 95% confidence intervals calculated using synthetic catalogs containing the same number of events as the observed catalog, but with events randomly distributed in time. The observed catalog (black dots) does not present significant clustering or (anti-)clustering for M_W>M_C.

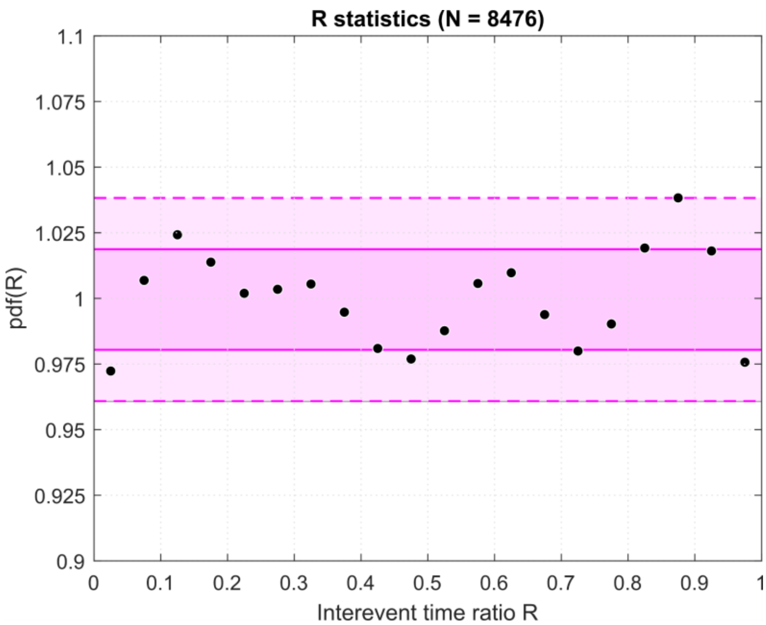

Figure S2: Differences in the probability to observe a magnitude difference $M_{i+1} - M_i < \Delta m$ between a selected subset of the seismic catalog (M_W>M_C, M_C=-0.5) associated with the 2018 stimulation and multiple realizations of its randomized versions which do not present signatures of magnitude correlations (see Supplementary Text S2 for details). The magenta areas correspond to 68% and 95% confidence intervals.

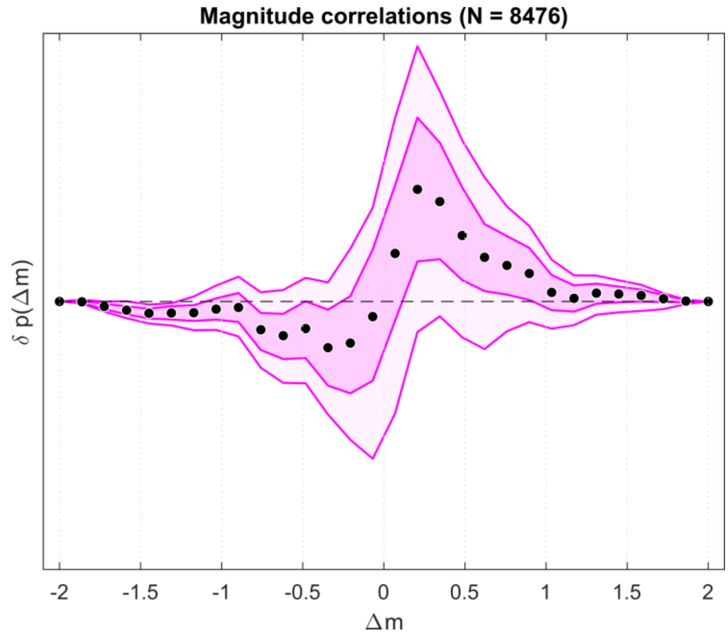

Figure S3. Temporal evolution of the Omori *p*-value(d,e) for the two stimulation campaigns. (a): 2018 stimulation; (b): 2020 stimulation. Black solid line shows the seismicity rate above M_C averaged over 2-hour bins. For 2018 stimulation, after the third week, the *p*-value starts decreasing and eventually converges down to *p*=1. However, the temporal evolution of the *p*-value for the 2020 stimulation exhibited the exactly opposite trend. The observable *p*-value is very low (below *p*=1) at the beginning of the stimulation, but after a week it starts to converge to *p*=1. Estimations of *p*-values are quite sensitive to the selected size of the binning window (e.g. 2 versus 4 hours), and thus the conclusions we can draw are limited.

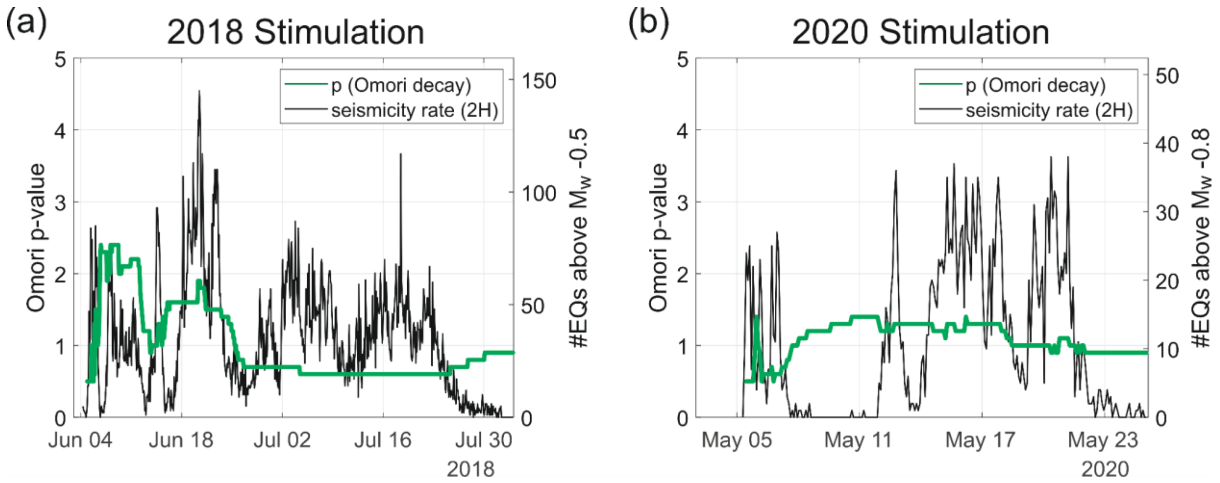

Figure S4. Hindcasted seismicity rates using the Seismogenic Index model in the magnitude-domain (Σ_{FV}, b_{FV}) and Omori's law (time-domain). Observed and simulated seismicity rates are shown with solid black and dashed dark magenta and red lines, respectively. Flow rate and well head pressure are shown with dotted blue and orange lines, respectively. Time period used for calibration of the model is shown with a gray background. (a): 2018 stimulation (staged injection), (b): 2020 stimulation (open hole injection), cf. Fig. 3.

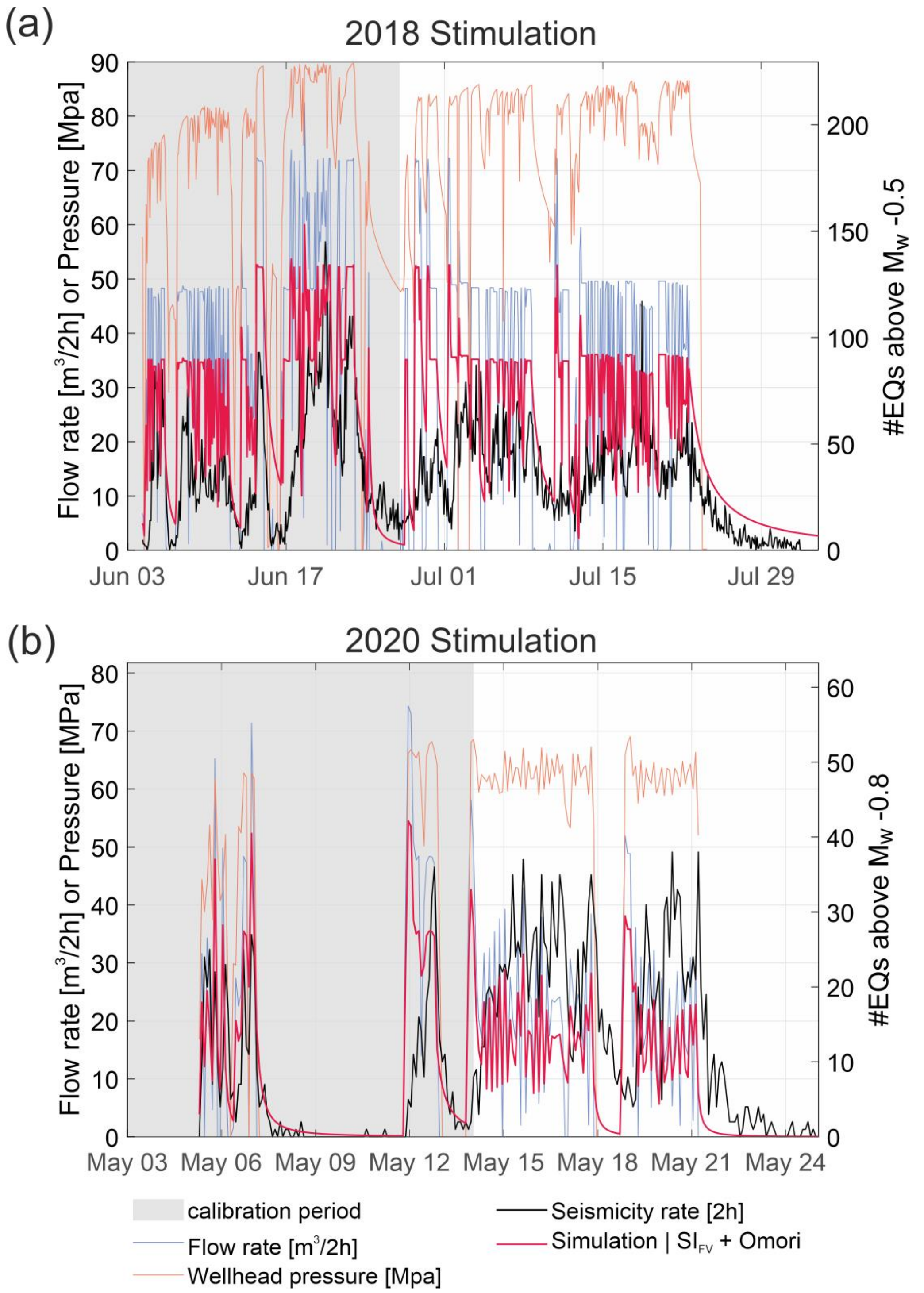

Figure S5: Example of magnitude-frequency distribution for 2020 without any bootstrapping applied to the cataloged magnitudes. Cumulative (solid black dots) and non-cumulative (open squares) distributions and G-R fits following Fitted models by Weichert (1980) and Shapiro et al. (2013) are shown with solid black line and dashed blue line, respectively.