



Design and implementation of a traffic light system for deep geothermal well stimulation in Finland

Thomas Ader · Michael Chendorain · Matthew Free · Tero Saarno · Pekka Heikkinen · Peter Eric Malin · Peter Leary · Grzegorz Kwiatek · Georg Dresen · Felix Bluemle · Tommi Vuorinen

Received: 11 April 2019 / Accepted: 12 July 2019 / Published online: 29 August 2019
© Springer Nature B.V. 2019

Abstract We describe the design and implementation of a traffic light system (TLS) created for the stimulation

T. Ader (✉) · M. Chendorain · M. Free
Arup, 13 Fitzroy Street, W1T 4BQ, London, UK
e-mail: thomas.ader@arup.com

M. Chendorain
e-mail: Michael.Chendorain@arup.com

M. Free
e-mail: Matthew.Free@arup.com

T. Saarno · P. Heikkinen
St1 Deep Heat Oy, Purotie 1/PL 100, 00381 Helsinki, Finland

P. E. Malin · P. Leary
Advanced Seismic Instrumentation & Research,
1311 Waterside, Dallas, TX 75218-4475, USA

G. Kwiatek · G. Dresen
Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Section 4.2: Geomechanics and Rheology, Potsdam, Germany

G. Kwiatek
Free University Berlin, Berlin, Germany

G. Dresen
University of Potsdam, Potsdam, Germany

F. Bluemle
fastloc GmbH, Berlin, Germany

T. Vuorinen
Institute of Seismology, University of Helsinki, Helsinki, Finland

phase of a 6.1-km-deep geothermal well in the Helsinki area, Finland. Because of the lack of local seismic data to calibrate the TLS, the TLS thresholds have been proposed on the basis of two parameters: acceptable ground motion levels and probabilities to reach these levels, which were then used to establish TLS magnitudes. A peak ground velocity (PGV) of 1 mm/s associated with a $M_L \geq 1$ event, or a $M_L \geq 1.2$ event alone, triggered an Amber alert. A PGV of 7.5 mm/s associated with a $M_L \geq 2.1$ event triggered a Red alert, where M_L was a local “Helsinki” magnitude. Specific thresholds based on PGV and peak ground acceleration were gathered for critical infrastructure sites and related to earthquake magnitudes in a probabilistic way. The implementation of the TLS during the stimulation showed that the selected thresholds were reasonably conservative. Some of the Amber events were reportedly felt or heard, but the public opinion remained very favourable to the project. Measured ground motions were too low to have any impact on the built environment and the stimulation was able to proceed without being impaired by the occurrence of a Red event. Seismic data acquired during the stimulation and lessons learnt are presented and used to revisit the TLS thresholds for potential future stimulations.

Keywords Geothermal · Stimulation · Traffic light system · Induced seismicity · GMPE

1 Introduction

St1 Deep Heat (St1 DH) is developing a geothermal doublet in Espoo, Finland, for the purpose of supplying deep geothermal heat to local district heat networks. As part of the project, the first well OTN-3 was completed in May 2018 with a true vertical depth of 6.1 km (Fig. 1). In June and July 2018, OTN-3 was stimulated during 7 weeks in order to improve rock permeability (Kwiatek et al. 2019). An adjacent 3.3-km-deep well is to be deepened to a similar depth of 6.1 km and stimulated in 2019.

Given that the stimulation took place beneath a densely populated area with multiple sensitive receptors, the City of Espoo's buildings department required that a seismic traffic light system (TLS) be developed and approved before granting permission for St1 DH to perform well stimulation activities.

TLSs are commonly used to reduce the potential seismic hazard due to induced seismicity and to mitigate the associated risk by modifying the fluid injection profile (e.g., Bommer et al. 2006; Bosman et al. 2016; Ellsworth 2013; Haering et al. 2008). TLSs were first proposed by Bommer et al. (2006) for the Berlín geothermal project in El Salvador (Bommer et al. 2006), an area with significant background seismicity. The approach has then been adopted on several geothermal projects (e.g., Diehl et al. 2017; Haering et al. 2008). In TLSs such as the one from Bommer et al. (2006), called traditional TLSs, the thresholds are usually defined ad hoc and primarily based on expert judgement (e.g., Wiemer et al. 2017). An alternative to traditional TLSs are the so-called adaptive TLSs (ATLSs). Although ATLSs are still under development, they already proved to be efficient at mitigating risk during some geothermal operations

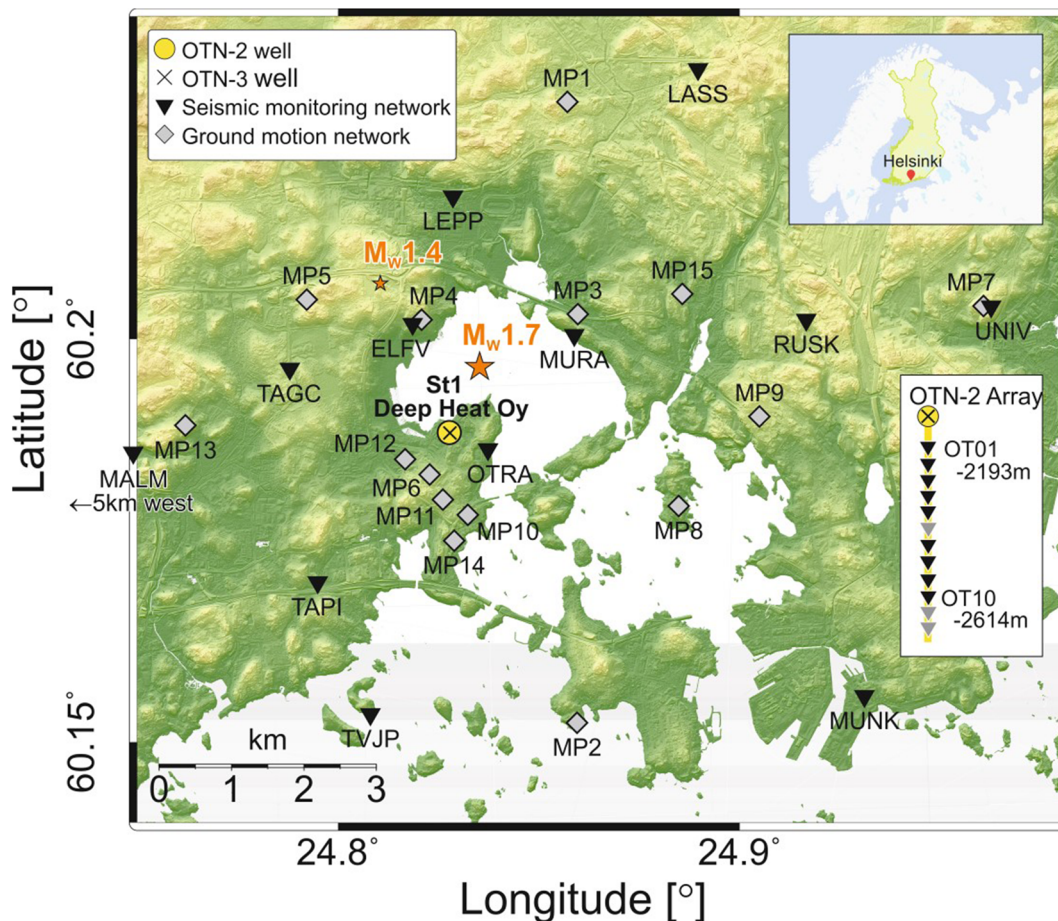


Fig. 1 Location of the St1 DH project site and of the different seismic networks used to monitor the stimulation campaign. Two natural earthquakes that occurred in 2011 in the proximity of project site are shown as orange stars (from Kwiatek et al. 2019)

(e.g., Broccardo et al. 2017; Gischig et al. 2014; Mignan et al. 2017). ATLSs have the benefit of being forward looking, probabilistic in their forecasts and adaptive, in the sense that the forecasted seismicity and resulting hazard are constantly updated. However, ATLSs are model dependent and they require sufficient data to be implemented.

The stimulation of OTN-3 was performed in an area with very little natural seismicity and there was no existing data on induced seismicity. Besides, the stimulation of OTN-3 was the first geothermal well stimulation ever performed in Scandinavia and the deepest geothermal well stimulation in the world to this date, so there was no guarantee that the models developed on other geothermal well stimulations elsewhere in the world would apply well in this case. Therefore, the traditional TLS remained the preferred solution due to the risk that ATLS models would not model well the expected induced seismicity.

In this study, we present the design, implementation, and post-stimulation analysis of the traditional TLS that was successfully implemented and used during the stimulation phase of the OTN-3 well. We first give an overview of the methodology used to define the TLS thresholds. In the following, we discuss different steps and measures taken in order to insure the successful implementation of the TLS during the stimulation of OTN-3. We present the impacts of stimulation-induced seismicity and use this new data to revisit the TLS thresholds. We study the performance of existing forward-looking models on the recorded induced seismicity in order to evaluate the potential of an ATLS for future stimulations in similar conditions.

2 Design of the TLS

2.1 TLS description

An effective TLS relies on a real-time seismic monitoring system and leverage this information to mitigate the risk of negative public response and the risk to the built environment. The developed TLS relied on the input of two independent seismic monitoring networks (Fig. 1). The first network was composed of 12 3-component seismometers installed in shallow boreholes between 300-m and 1.15-km depth (referred to as the satellite network). This network was completed

with a vertical array of 12 3-component sensors located in a well located 10 m away from OTN-3, at depths from 2200 to 2630 m (for details, see Kwiatek et al. 2019). The second network was composed of 14 surface geophones (referred to as the surface network) placed at strategic surface locations, such as nearby critical infrastructure sites (see Section 2.4). The satellite network was used to provide hypocentral locations and magnitudes of seismic events (see Kwiatek et al. 2019 for details), while the surface network measured the actual amplitude of ground motions for the purpose of TLS operations.

The installation and maintenance of the satellite network was performed by Advanced Seismic Instrumentation & Research (ASIR), while the localization of the seismic events and computation of source parameters in near-real time was undertaken by fastloc GmbH. Arup was in charge of the design and implementation of the TLS (Fig. 2).

The TLS comprised the following elements:

1. Provision for baseline monitoring of TLS networks for a period of 1 month prior to stimulation activities;
2. Provision for post-stimulation monitoring for a period of at least 6 months consisting of a subset of the satellite network operated remotely, and the surface network operated for only the first 2 months of the post-stimulation monitoring;
3. TLS thresholds associated with **Green**, **Amber**, and **Red** events:
 - (a) **Green** conditions allowed for stimulation activities to proceed as planned. During **Green** conditions, TLS activities included network monitoring and confirmation that monitoring stations were operational and transmitting data;
 - (b) **Amber** conditions indicated that a TLS exceedance had occurred, which triggered notification, documentation, and evaluation of whether any mitigation of seismic risk was required. While the trigger level for an **Amber** event might result in surface vibrations which would be felt in the vicinity of the event, no cosmetic or structural impacts on the built environment were expected;
 - (c) **Red** conditions indicated that a TLS exceedance had occurred. This triggered an immediate stop of stimulation activities with

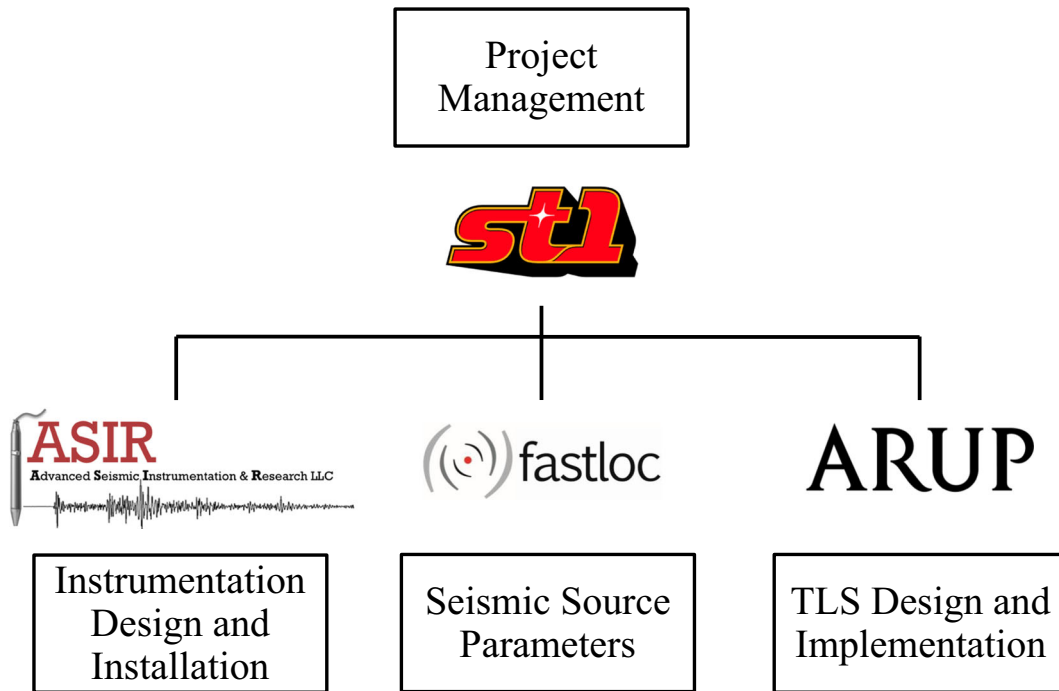


Fig. 2 Role of the different parties involved in the TLS

- a well bleed-off option, a notification of the earthquake event, and confirmation that activities had stopped. In addition, the event and necessary mitigation measures had to be documented. Stimulation activities following a **Red** event might only proceed with the approval of local authorities. Similar to the **Amber** threshold, the **Red** threshold corresponded to a level of shaking where no cosmetic or structural impacts were expected.
4. A detailed communication plan, or communication tree, including communication delays, parties to be informed, and template of information to be communicated, in case of an **Amber** or **Red** alert:
 - (a) Provision for communication by phone to the relevant parties within 20 min of an **Amber** or **Red** event;
 - (b) Provision for communication by email to the relevant parties within 30 min of an **Amber** or **Red** event; and
 - (c) Provision for reporting following exceedance of thresholds (known as exceedance reports) within 4 days of the event.

5. Actions and mitigation measures following an **Amber** event or a **Red** event;
6. Constraints on allowing stimulation to proceed during period of monitoring station outages; and
7. Provision for a final summary report to conclude the use of the TLS.

2.2 Methodology to establish TLS thresholds

The goal of an effective TLS is to reduce the risk of negative public response and infrastructural damage to a level as low as reasonably possible. The developed methodology considered the surface expression of the seismicity (ground motions) induced by the geothermal stimulation. The surface expression controls the public response and the hazard to the built environment located at the surface in the vicinity of the project. The TLS was to be applied in an environment where natural seismicity is practically non-existent. Therefore, limited data and no seismic norms were available to calibrate the potential effects of induced seismicity directly in terms of earthquake magnitudes. However, many countries have regulations and best practices for ground-borne vibrations, such as blasting or other vibration sources, and it is therefore much

easier and more practical to find a consensus on a level of surface vibrations with a regulator rather than directly on event magnitudes.

For the stimulation of OTN-3, peak ground velocity (PGV) was selected as the measure of ground motion. PGV is a simple parameter that has many engineering applications and it is meant to be a better indicator of the potential damage due to ground motions than peak ground acceleration (PGA) (Bommer 2017). The PGV levels were developed in accordance with Finnish Building Code (RIL 253-2010 2010), British Standards on surface vibrations (BSI 1993, 2008), and various publications illustrating the relationship between PGV and impacts on human perception and the built environment (Bommer 2017; Westaway and Younger 2014). The PGV thresholds selected were as follows:

1. For **Green** conditions, a low level of 0.3 mm/s or 0.13 mm/s was set in surface-network stations to generate low-level alarms, in the form of text messages to a dedicated site phone. The value of 0.3 mm/s was the default value and the threshold was lowered to 0.13 mm/s for a few stations in low background noise level areas. This low level condition was aimed at providing sufficient feedback that surface network stations remained operational. This threshold was selected as it was below the general thresholds of human perception (Bommer 2017; BSI 2008; Westaway and Younger 2014).
2. For **Amber** conditions, a threshold of 1 mm/s was selected as the PGV threshold. This level correlated with the lower threshold of human perception proposed in Bommer (2017) and with the maximum satisfactory magnitudes of vibrations in residential buildings at night, with respect to human response, according to British Standards (BSI 2008). An additional factor of safety of two was applied to reach the PGV threshold of 1 mm/s. While a PGV of 1 mm/s might be noticed by the local community, no credible impacts would be expected.
3. For **Red** conditions, a threshold of 7.5 mm/s was selected as the PGV threshold. This level correlated with the lower threshold of potential cosmetic damage for buildings according to British Standards (BSI 2008) and Bommer (2017), with an additional factor of safety of two. This threshold was considered to be conservatively

reasonable since it implied that stimulation activities would be halted prior to any cosmetic or other (e.g., structural) impacts.

These PGV thresholds proved to be a reasonable and practical basis for the design of the TLS. However, the TLS based on PGV required addressing potential problems including:

1. **False positives** (i.e., false alerts): high levels of vibrations could be recorded that would not have been caused by induced earthquakes at the geothermal well. The levels of vibrations could be caused by blasting at the surface, road traffic, equipment malfunction, distant natural earthquakes, or other unanticipated sources; and
2. **False negatives** (i.e., no alert when there should be one): by nature, the PGV can only be measured at the specific locations where instruments have been installed. PGV thresholds could be exceeded at locations where no instrument was recording, and would therefore be missed.

In order to circumvent these issues, two types of alerts were considered:

1. **Joint PGV-magnitude alert** When a measured PGV exceeded one of the TLS thresholds, a seismic event at the production site had to also be detected with magnitude large enough to trigger such a level of ground vibrations. This process was aimed at reducing the risk of false alerts.
2. **Magnitude-only alerts** If an earthquake was detected during operations, which had enough potential to generate PGVs exceeding one of the PGV thresholds, an alert would be triggered. This would happen even if no PGV was measured at any of the surface geophones above any of TLS thresholds. This type of alert was aimed at reducing the risk of false negative (i.e., missed alerts).

The PGV thresholds were therefore related to corresponding magnitudes of events that would occur at the depth of injection. For that purpose, the empirical ground motion prediction equations (GMPEs) by Douglas et al. (2013) and a GMPE by the Institute of Seismology of the University of Helsinki (ISUH 2017, personal communication) were used. More details are provided in Ader et al. (2019).

The magnitude scale used in the TLS was the local “Helsinki” magnitude M_{LHEL} (Lund et al. 2015; Uski and Tuppurainen 1996). In this study, the designations

M_L and M_{LHEL} are used interchangeably to designate this local “Helsinki” magnitude.

GMPEs provide median estimations of ground motions and uncertainties on these median estimations, the parameter σ , which is the standard deviation on the logarithm of the median value. In practice, for a seismic event of given magnitude at a known distance, the median ground motion predicted by the GMPE has 50% chance to be exceeded. The GMPE uncertainties can be used to compute the ground motions that have different probabilities to be exceeded by a given event: for instance, the median value of the ground motion plus one sigma has about 16% chances to be exceeded.

The magnitudes to be used in the TLS thresholds were therefore based on agreed probabilities to exceed PGV thresholds at the epicentre of a seismic event. These probabilities were then related to an event magnitude by taking into account the median ground motions predicted by the GMPEs and the GMPEs’ uncertainties on this median prediction (Ader et al. 2019). More specifically, two probabilities were agreed for the two types of thresholds considered for the **Amber** alert:

1. Magnitudes associated to joint PGV-magnitude alerts were selected based on a 2% probability that the seismic event’s magnitude would cause a PGV exceedance at the epicentre; and
2. Magnitudes associated to magnitude-only alerts were initially selected based on a 10% probability that the magnitude would result in a PGV exceedance at the epicentre.

In order to simplify the **Red** trigger of the TLS, it was agreed that a TLS alert would be conservatively triggered by the occurrence of a seismic event, which magnitude had a probability of at least 2% to cause a **Red** PGV exceedance at the epicentre. Besides, in the case of a **Red** event, any PGV measured above the **Amber** and or **Red** PGV thresholds would have to be reported.

Details of the calculations and negotiations with the regulator are given in Ader et al. (2019). The resulting **Amber** and **Red** thresholds of the TLS were selected as follows:

Amber alert:

PGV ≥ 1 mm/s detected at one or more surface stations and $M_L \geq 1.0$; or
 $M_L \geq 1.2$.

Red alert:

$M_L \geq 2.1$.

The method provided the benefit that only two parameters had to be agreed with the regulator: the PGV thresholds and the acceptable levels of probability to reach these levels. These proved to be practical parameters that both technical and non-technical parties could relate to and which could be backed up by factual standards and publications. From there, the magnitudes could be deduced with the data available, through a process taking into account the uncertainties on the data. Importantly, once more data would become available and the uncertainties reduced, the magnitude and PGV thresholds could be revisited using the same methodology.

2.3 Spatial domain of application of the TLS

In areas of natural seismicity, it is difficult to determine whether a particular seismic event occurring in the vicinity of stimulation site is a result of stimulation.

However, the Helsinki area is characterized by very low seismicity, and the occurrence of a natural $M_L \geq 1.2$ during the stimulation was considered very unlikely. For these two reasons, the TLS did not distinguish between natural and induced events in the vicinity of the injection point. In previous EGS projects, induced seismicity had been well contained within 2 km from the injection point in all directions (e.g., Cuenot et al. 2008; GEISER 2013; Halldorsson et al. 2012; Wiemer et al. 2017). For the purpose of the TLS, it was decided that any detected event that met the following criteria, regardless of whether it was induced by stimulation or coincidentally natural, would trigger a TLS threshold breach:

- Radius of 5 km from the water injection point in horizontal direction; and
- Depth of events less than 10 km and more than 1 km (in order to exclude surface blasts).

These conservative 5-km radius and 10-km depth limit largely compensated for the uncertainties in earthquake locations (Kwiatek et al. 2019). Again, once more data became available and location uncertainty was reduced, then these location thresholds could be readjusted.

2.4 Sensitive receptors

Sensitive receptors are critical infrastructure sites, such as hospitals, schools, and laboratories, where the occupants or equipment might be more susceptible to the effects of vibrations than normal residential or commercial buildings.

A survey of local receptors was performed by the St1 DH team, who identified six sensitive receptors within 15 km from the injection point:

- Micronova Company of the Technical Research of Finland (1.7 km);
- Mikes Metrology Building (2.3 km);
- Meilahti Hospital (3.5 km);
- Accelerator Lab at the University of Helsinki, Kumpula Campus (10 km);
- IT Centre for Science (3 km); and
- Otaniemi drilling site (1 km).

One surface station was placed at each of these sensitive receptors in order to specifically monitor ground motions. Specific TLS thresholds were established independently for each of the sensitive receptors, by investigating acceptable levels of ground motion for each sensitive receptor. The ground motions were related to event magnitudes through the same methodology as previously described. When ground motion requirements were defined in terms of PGA rather than PGV, the GMPEs used were the empirical PGA GMPE by Douglas et al. (2013) and a PGA GMPE provided by ISUH (2017, personal communication).

When the general TLS thresholds were more conservative than the requirement at a specific site, the general TLS thresholds were selected. This turned out to be the case for four out of the six sensitive receptors. The last two receptors had requirements in

terms of PGA and specific thresholds were therefore maintained. It should be noted that the magnitudes related to these specific thresholds were higher than the magnitudes related to the general TLS thresholds, highlighting the conservativeness of the general TLS thresholds.

When a TLS alert occurred, the PGV and PGA measured at the sensitive receptors were systematically reported in the exceedance report, whether or not they exceeded one of the specific thresholds.

3 Implementation of the TLS

3.1 OTN-3 stimulation

The stimulation of well OTN-3 started on 4 June 2018 and ended on 22 July 2018 (49 days). A total volume of 18,160 m³ of drinking quality water was injected through five stages located along the open-hole section of the well, inducing 1357 $M_L \geq 0$ events. The injection rate ranged generally between 400 and 600 l/min. It was increased to about 800 l/min for a couple of hours during stage 2, leading to a rapid acceleration of seismic activity (cf. Kwiatek et al. 2019). The start dates of each stage, net injected volumes, max injection pressures, mean depth, and number of events are detailed in Table 1.

The spatial arrangement of $M_L \geq -1$ seismic events used for the TLS implementation is plotted in Figs. 3 and 4. The seismicity has been relocated after the stimulation (Kwiatek et al. 2019), but we only present the seismicity originally provided to the TLS here.

The event hypocenters fed into the TLS during stimulation appear to be slightly shifted horizontally

Table 1 Details of the five stimulation stages

Stage	Start date	Net injected volume (m ³)	Max injection pressure (bars)	Mean depth (m)	Number of events			
					$M_L \geq 0$	$M_L \geq 1$	$M_L \geq 1.2$	$M_L \geq 2.1$
1	4 June 2018	3783	890	6056	173	5	3	0
2	16 June 2018	4023	900	5973	280	11	6	0
3	23 June 2018	1537	850	5884	131	8	4	0
4	1 July 2018	4411	860	5771	361	14	9	0
5	12 July 2018	4406	870	5625	412	18	14	0
Total	22 July 2018	18,160	–	–	1357	56	36	0

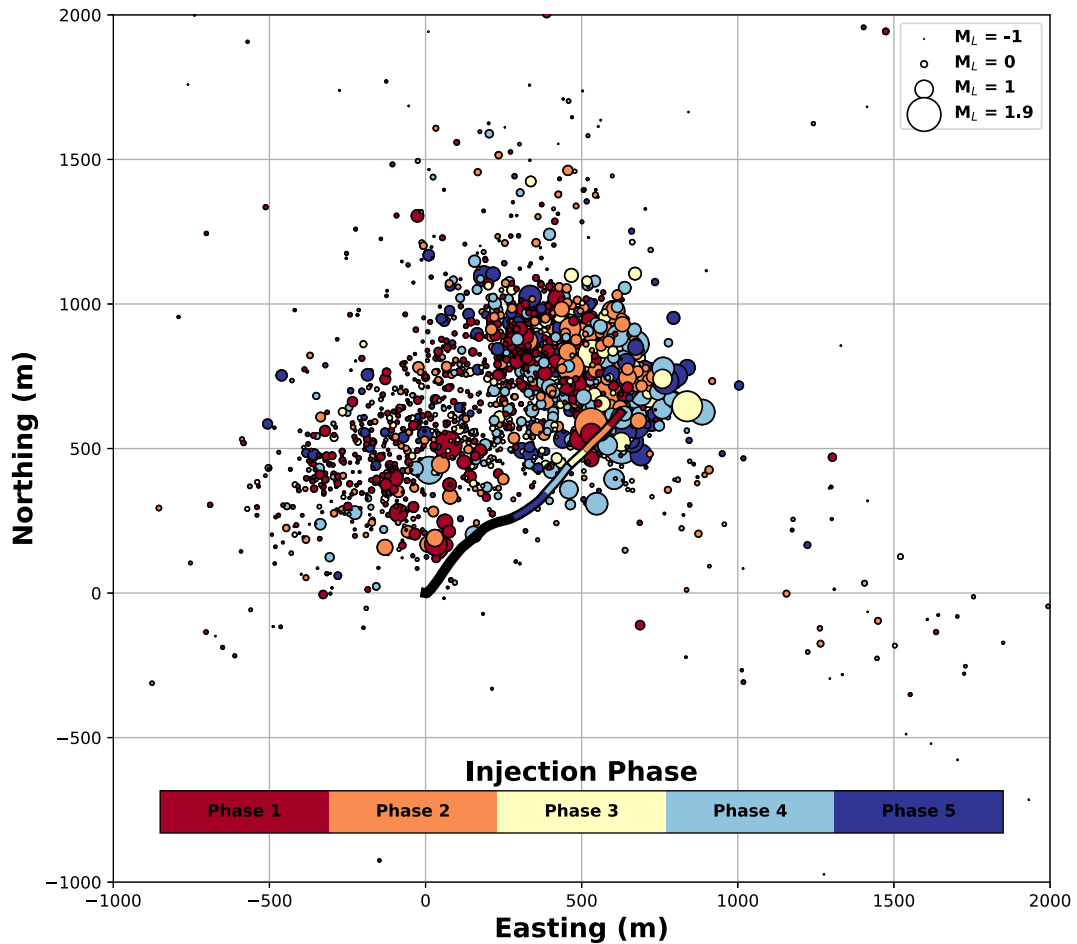


Fig. 3 Map view of the $M_L \geq -1$ stimulation-induced events used for TLS implementation. Events are colour-coded depending on the stimulation phase during which they occurred, while event sizes represent their magnitudes

and vertically compared with the injection point and the well path. This has been identified to be related to the velocity model used, and it was addressed in post-processing (cf. Kwiatek et al. 2019). Despite this apparent shift, the events were concentrated within 1000 m from the injection points, at depths between 5 and 7 km. This was confirmed for $M_L \geq 0$ events and this was even clearer for Amber events, which were all located within 500 m horizontally and 1000 m vertically from the injection point. The events were therefore located well within the volume considered for the TLS (Section 2.3). Finally, the observed shift did not influence significantly the calculated magnitudes provided to the TLS operator.

The magnitude-frequency Gutenberg-Richter (GR) distribution of events recorded during stimulation is

plotted in Fig. 5. It is clearly observed that event detection and location limit were between $M_L -1$ and 0, which was sufficient for the implementation of the TLS and its Amber level starting at $M_L = 1$. The GR fit to this distribution has a b value of 1.32 ± 0.03 . Such b values greater than one are usual for geothermal induced seismicity (e.g., Grünthal 2014).

On the upper end of magnitudes, the distribution of events followed the Gutenberg-Richter distribution up to approximately $M_L 1.5$. For greater magnitudes, the frequency of events dropped off. Figure 6 shows the GR distribution of the events in each injection phase and suggests that this reduction of larger events only occurred during phase 5 (the shallowest stimulation phase). The implications of this observation are discussed in Section 5.2.

Fig. 4 Cross sections of $M_L \geq -1$ stimulation-induced events used in the TLS, in the cross-sectional plane of the well (upper) and in the cross-sectional plane normal to the plane of the well (lower)

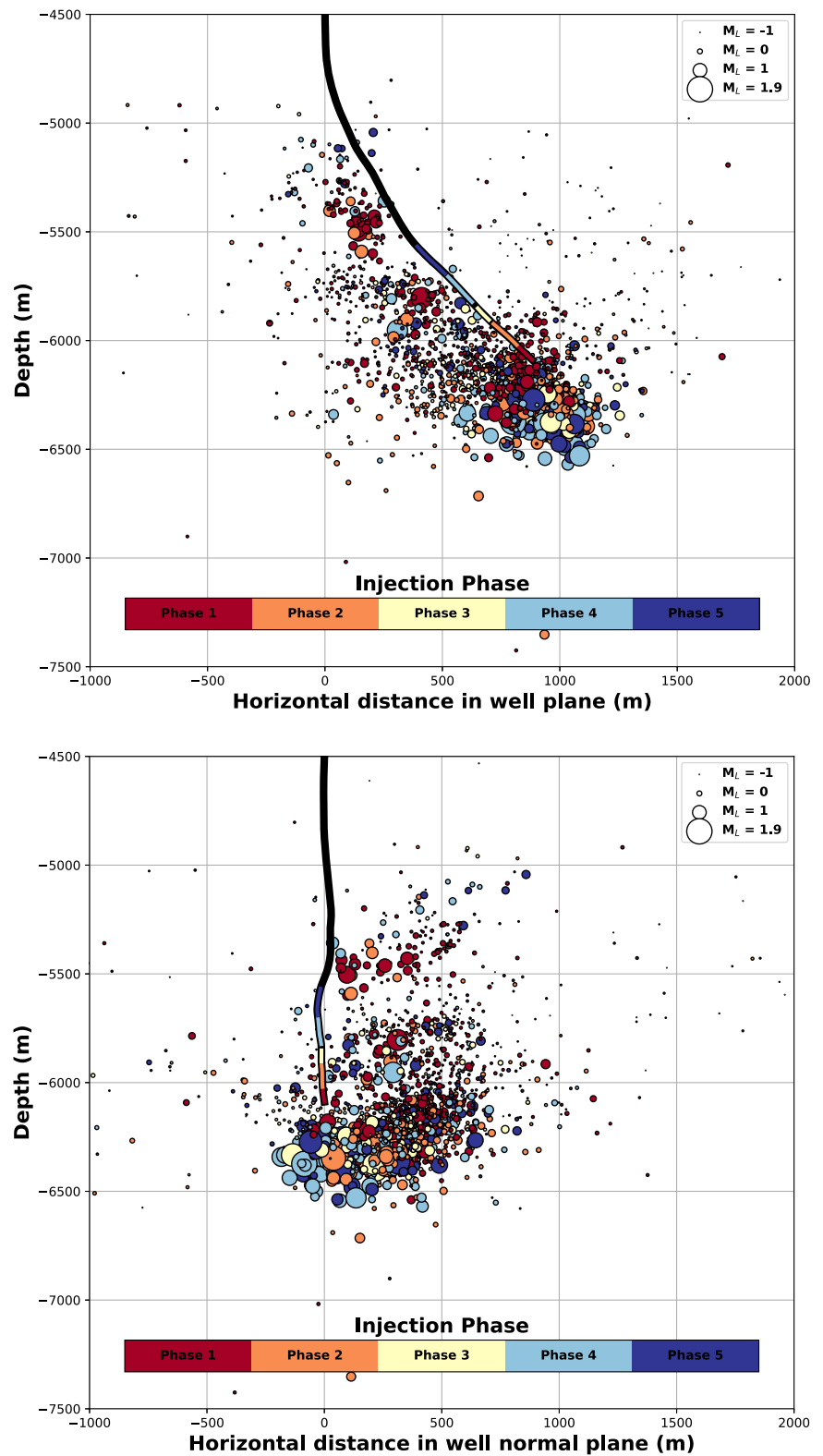
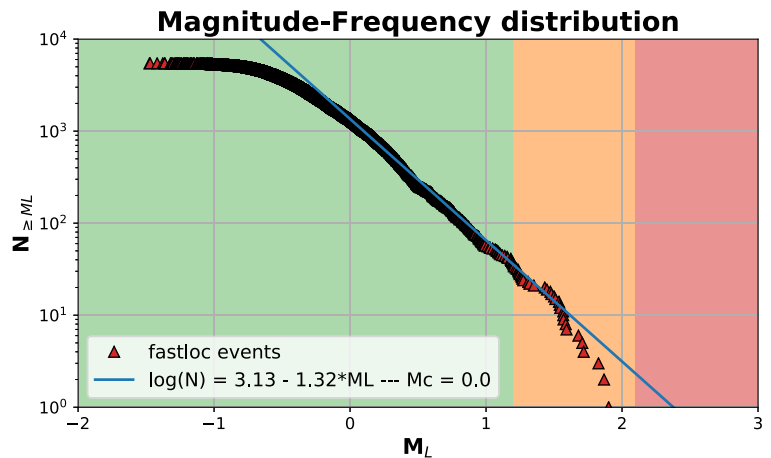


Fig. 5 Magnitude-frequency distribution of events detected during the stimulation and used in the TLS with Gutenberg-Richter fit to events with $M_L \geq 0$



3.2 TLS alerts

A total of 37 alerts were triggered during the stimulation and are detailed in Table 2. The first alert triggered was actually a **Green** event (albeit close to exceeding the **Amber** threshold) but St1 DH decided to carry on the alert process as an in situ practice exercise.

The following 36 alerts triggered were all **Amber** alerts and the largest magnitude recorded was $M_L = 1.9$. No **Red** event occurred during stimulation. Three of the 36 **Amber** alerts happened after the end of the stimulation activities and the latest **Amber** was a $M_L = 1.21$ event, which occurred about 30 h after the end of stimulation.

Only six of the **Amber** events resulted in surface motion with $PGV \geq 1$ mm/s, i.e., the **PGV Amber** threshold (Fig. 7). The lowest magnitude event generating a $PGV \geq 1$ mm/s was a $M_L 1.55$ event, which

caused a maximum PGV of 1.19 mm/s. The greatest PGV recorded during stimulation was 2.99 mm/s, following a $M_L 1.87$ event. Generally, Fig. 7 shows that events started to trigger **Amber** alerts at magnitudes lower than the ones required to exceed the **Amber** PGV threshold of 1 mm/s. This suggests that the magnitudes associated to the PGV thresholds were indeed reasonably conservative.

3.3 Procedures for TLS alerts

In order to implement the TLS alert actions, an on-site 24h-a-day presence was required and St1 DH hired four so-called TLS managers, who worked 12-hour shifts during stimulation. The TLS managers underwent training before the start of the stimulation activities and a user manual was put together to assist them through their task, which covered items such as

Fig. 6 Magnitude-frequency GR distribution of events detected during each phase of the stimulation with power-law GR fit to events with $M_L \geq 0$

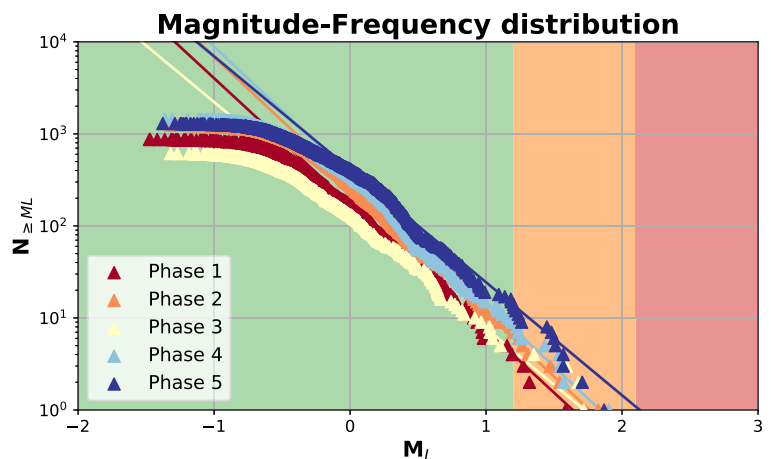


Table 2 List of alerts triggered, with maximum PGV and station at which it was measured

	Time (Helsinki)	M_L	Max PGV (mm/s)	Post event decisions and other notes
STAGE 1	09/06/2018 02:42	1.18	0.13	Green alert, done for practice
	09/06/2018 08:42	1.27	0.48	–
	11/06/2018 08:26	1.59	0.81	–
	13/06/2018 06:11	1.32	0.34	–
STAGE 2	17/06/2018 12:50	1.31	0.18	–
	20/06/2018 03:13	1.47	0.75	–
	20/06/2018 20:39	1.27	0.45	Decided to stop if further events were to occur
	21/06/2018 02:27	1.83	1.45	Stopped pumping for 4 h
	21/06/2018 20:56	1.54	0.57	–
STAGE 3	22/06/2018 21:13	1.24	0.44	–
	29/06/2018 02:31	1.35	0.60	High seismicity rate: stopped injection
	29/06/2018 07:02	1.72	0.76	–
	29/06/2018 12:42	1.59	0.63	–
STAGE 4	30/06/2018 09:53	1.68	1.24	Ended stimulation stage
	01/07/2018 23:33	1.20	0.38	–
	03/07/2018 23:11	1.26	0.23	–
	03/07/2018 23:30	1.43	0.47	Stopped pumping for 12 h
	05/07/2018 10:01	1.57	0.78	–
	06/07/2018 11:49	1.54	0.57	–
	07/07/2018 20:32	1.55	1.19	–
	07/07/2018 20:55	1.24	0.34	Hold pumping if other Amber event during the night
	08/07/2018 20:36	1.90	1.72	Stopped pumping for 48 h
	09/07/2018 05:09	1.23	0.28	No pumping
STAGE 5	12/07/2018 17:25	1.57	0.56	–
	13/07/2018 16:35	1.51	0.65	–
	16/07/2018 20:26	1.87	2.99	Stopped pumping for 5 h
	18/07/2018 13:43	1.50	0.67	–
	18/07/2018 20:09	1.22	–	–
	19/07/2018 01:45	1.48	0.41	Stopped pumping for 3 h
	19/07/2018 11:27	1.20	0.18	–
	19/07/2018 13:09	1.25	–	Pumping was stopped since 13:00
	19/07/2018 13:55	1.57	0.66	Extended pumping pause to 12 h
	19/07/2018 13:55	1.71	1.03	Same as above
21/07/2018 16:46	1.26	0.49	–	
22/07/2018 20:15	1.24	0.23	Stimulation ended 1 h earlier	
24/07/2018 01:03	1.45	0.93	–	
24/07/2018 02:48	1.21	0.28	–	

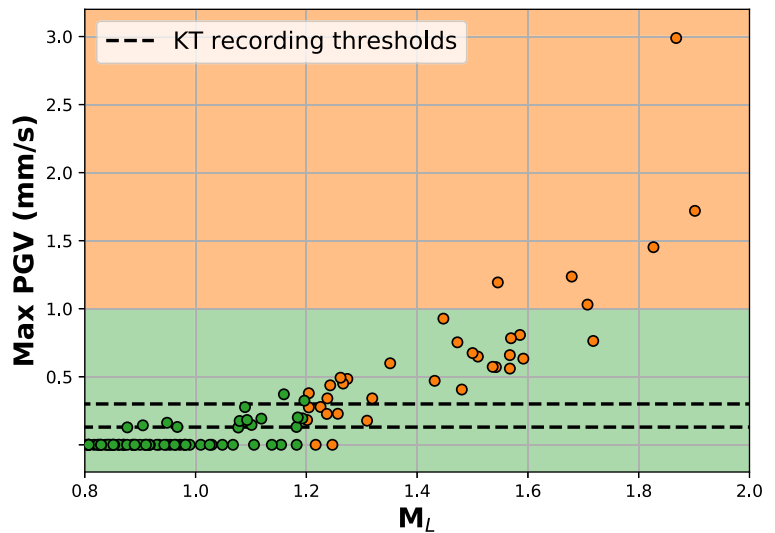
a summary of the TLS, the description of the different monitoring tools available to the TLS manager, and a description of the detailed tasks of the TLS manager.

The main role of the TLS manager was to monitor seismicity and to give the alert in case of a TLS exceedance. A communication tree was designed to

ensure that the different parties were alerted and provided with the right information in due time. Within 20 min of the event, the TLS manager had to ensure that the following parties were alerted by phone:

- The pumping operator;

Fig. 7 Maximum PGV recorded for $M_L \geq 0.8$ events during stimulation. The dashed lines show the 0.13 mm/s and 0.3 mm/s thresholds of recording at the surface stations



- The St1 project manager;
- St1 public affairs;
- The regulator;
- The rescue department; and
- The City of Espoo.

Within 30 min of the event, the same parties were all alerted by email, following a set template. A more comprehensive report on the event and possible mitigation measures taken by St1 DH was sent out within 4 days of the event.

In order to minimize further the risk of false alert, an automated alert was sent to all members of the operating team through the Pushover app, when an event of magnitude greater than 1 was detected within the TLS volume, within 5 min of the event. This procedure ensured that the TLS manager had initiated the TLS procedure and that the event magnitude and location was being manually checked if needed.

3.4 Response to induced seismicity

Several measures were implemented by St1 DH following **Amber** events. Engineering measures were typically implemented when the frequency of events was qualitatively observed to increase, when the background seismicity rate seemed to increase, and/or when the magnitude of **Amber** events was approaching a **Red** threshold:

- Pause injection for some period of time;
- Terminate the current injection stage (i.e., stop injection and then start bleeding off the well);

- Decrease the injection rate; and/or
- Decrease the injection pressure.

As expected, following a pause in injection, the seismicity rate decreased (Kwiatek et al. 2019). When injection resumed after a pause, there was usually a short time lag before the seismicity rate resumed. Based on this, St1 DH controlled the seismic hazard during the stimulation by reducing water injection and well-head pressure when the seismicity rate appeared to increase. Table 2 details the instances where stimulation was stopped because of the observed seismicity. More rationale behind the measures taken to mitigate the seismic hazard during injection are detailed in Kwiatek et al. (2019).

4 Impacts of stimulation-induced seismicity

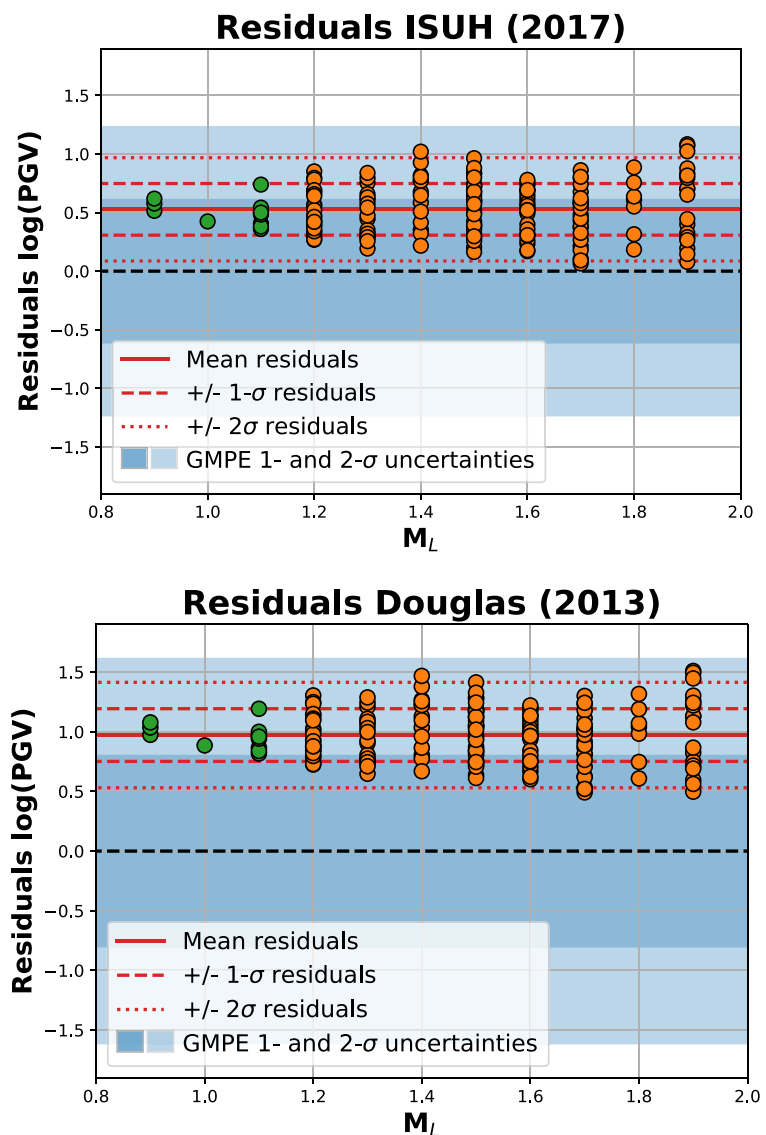
4.1 Peak ground velocity

Ground motions were recorded at several stations from the surface network for events of magnitudes ranging from $M_L 0.9$ to 1.9. The different PGVs recorded by the surface stations were compared with the predictions from the different GMPEs used in the TLS design: the empirical GMPE by Douglas et al. (2013) and the GMPE provided by ISUH.

Figures 8 and 9 show the residuals on the PGV, defined as:

$$\text{res}(\log \text{PGV}) = \log \text{PGV}_{\text{mes}} - \log \text{PGV}_{\text{GMPE}}, \quad (1)$$

Fig. 8 Residuals of the $\log(\text{PGV})$ for the two GMPEs used to design the TLS thresholds (upper plot: ISUH (2017, personal communication); lower plot: Douglas et al. 2013), as a function of the event magnitude. The mean of the residuals and the 1 and 2 standard deviations of the residuals are indicated in red. The 1- and 2- σ uncertainties on the GMPE are indicated in shades of blue. The data points are coloured according to their TLS category (Green or Amber)



where $\log \text{PGV}_{\text{mes}}$ is the decimal logarithm of the measured PGV at the surface stations, and $\log \text{PGV}_{\text{GMPE}}$ is the decimal logarithm of the PGV predicted by the GMPE (median value). Note that, with this definition of the residuals, positive residuals indicate that the measured PGV is greater than the GMPE estimates.

The median predictions of both GMPEs consistently underestimate the measured PGVs: for both GMPEs, the residuals are all positive. However, owing to the large uncertainties on both GMPEs, all of the residuals fall within the 2- σ prediction range from the GMPEs. In terms of the TLS design, the methodology used to calculate the TLS thresholds was accounting

for the GMPE uncertainties, so that the TLS thresholds computed were still conservative, even though the median predictions of the GMPEs turned out to be under-conservative. This shows that the method to establish the TLS thresholds is robust with respect to the GMPEs used, as long as the uncertainties on the GMPEs are well estimated.

Quantitatively, Figs. 8 and 9 suggest that the median predictions of the GMPE by Douglas et al. (2013) under-predict the $\log(\text{PGV})$ by about 1.0, while the GMPE by ISUH (2017, personal communication) under-predicts the $\log(\text{PGV})$ by about 0.5, i.e., by factors of respectively 10 and 3.

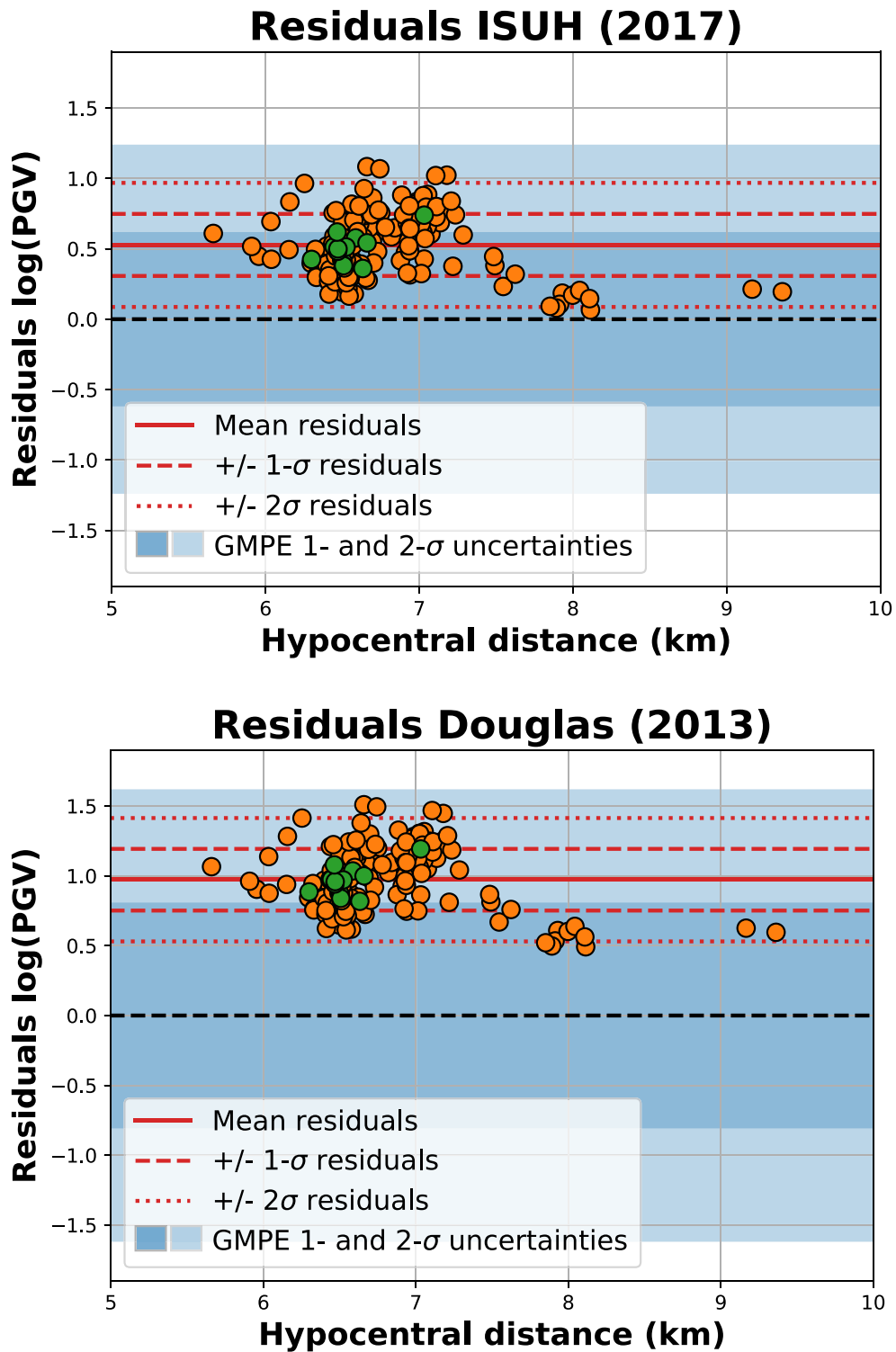


Fig. 9 Residuals of the log(PGV) for the two GMPEs used to design the TLS thresholds (upper plot: ISUH (2017, personal communication); lower plot: Douglas et al. 2013), as a function of the hypocentral distance of the event. The mean of the

residuals and the 1 and 2 standard deviations of the residuals are indicated in red. The 1- and 2- σ uncertainties on the GMPE are indicated in shades of blue. The data points are coloured according to their TLS category (Green or Amber)

Figure 8 shows how the residuals depend on event magnitude and suggests that there is no magnitude bias, as the residuals have the same offset at all magnitudes. Figure 9 shows the variation of residuals with hypocentral distance and might suggest that the predictions of the GMPEs are closer to the measured GMPEs at hypocentral distances greater than 7.5 km. However, there are too few data points at hypocentral distances greater than 7.5 km to confidently determine this. The standard deviation of the residuals in Figs. 8 and 9 shows the spread of the measured values around their own median value. This standard deviation is only 0.22, significantly smaller than the $1\text{-}\sigma$ uncertainties of 0.81 for Douglas et al. (2013) and 0.62 for ISUH (2017, personal communication). Using the stimulation data available, a site-specific GMPE was generated based on the ISUH GMPE, but corrected by adding the mean value of the residuals, resulting in the following corrected GMPE for PGV:

$$\log \text{PGV}_{\text{corrected}} = \log \text{PGV}_{\text{ISUH}} + 0.51. \quad (2)$$

The $1\text{-}\sigma$ uncertainty for this corrected $\log \text{PGV}$ simply became the standard deviation of the residuals:

$$\sigma(\log \text{PGV})_{\text{corrected}} = 0.22. \quad (3)$$

The resulting GMPE is shown in Fig. 10, with 1- and $2\text{-}\sigma$ uncertainties represented, for some selected event magnitudes. Figure 10 suggests that this corrected GMPE represents the observed PGV well: the median predictions of this corrected GMPE seem to match the median recorded PGV and the uncertainties on the GMPEs seem to cover the spread of the measured PGVs.

Finally, it is important to note again here that the surface stations generally had a triggering level at 0.3 mm/s, or 0.13 mm/s for a few stations in low background noise level areas. PGVs lower than the triggering levels are therefore missing from the dataset, and it could therefore be argued that the dataset might be biased towards higher PGVs. However, Fig. 10 shows that, for $M_L \geq 1.5$, there is not a strong cut-off of data points at $\text{PGV} = 0.13$ mm/s. This suggests that, at least for $M_L \geq 1.5$, the consistently high PGVs recorded are not a measurement artefact.

4.2 Public perception

Prior to stimulation, ISUH had indicated that small seismic events in Finland, such as $M_L \leq 2$ events,

had previously resulted in several notifications from concerned citizens, in areas much less populated than Helsinki. These events were usually much shallower than the depth of stimulation of OTN-3, but the public sensitivity to potential induced events during OTN-3 stimulation was an important point of focus for St1 DH.

In order to make the local community aware of its plan to stimulate the geothermal doublet, St1 DH developed a communication plan. Advertisements were made in local newspapers and through mailers to the local community. A website was launched and maintained by St1 DH (www.st1.fi/geolampo for the Finnish version and www.st1.eu/geothermal-heat for the English version). The website provided general project information and was updated regularly. The blog within the website provided a mechanism for project updates, including information on the TLS during stimulation. The website gave people the possibility to report any heard or felt seismicity during the stimulation period, either through the website or through a phone number. A media event and open house were held before the start of stimulation activities, including a site visit and presentations by the St1 DH team. The open house allowed locals to raise questions and concerns about the project. Following the media event, additional presentations by the St1 DH team were given at Aalto University in Espoo, close to the location of St1 DH project.

ISUH is usually the institution in Finland who collects reports of felt seismicity in the country and their service was also working during OTN-3 stimulation.

During the 7 weeks of stimulation, St1 DH received a total of 25 notifications and ISUH a total number of 173 notifications from the public. These numbers were proportionally very low, in comparison with the numbers of reports that ISUH can receive from much less populated areas than Helsinki following a small natural earthquake. The total number of responses recorded by ISUH and St1 DH for each event is plotted as a function of the event magnitude in Fig. 11. ISUH indicated that their numbers were calculated from raw data and were therefore preliminary, as they had not been checked for possible duplicates or invalid reports. The $M_L \geq 1.8$ events were the only events for which there were more than ten reports in total. The event for which both St1 DH and ISUH received the largest number of public responses was the largest

Fig. 10 PGVs recorded for M_L 1.2, 1.5 and 1.9 events. The recorded PGVs are compared with the predictions of the corrected GMPE (in red)

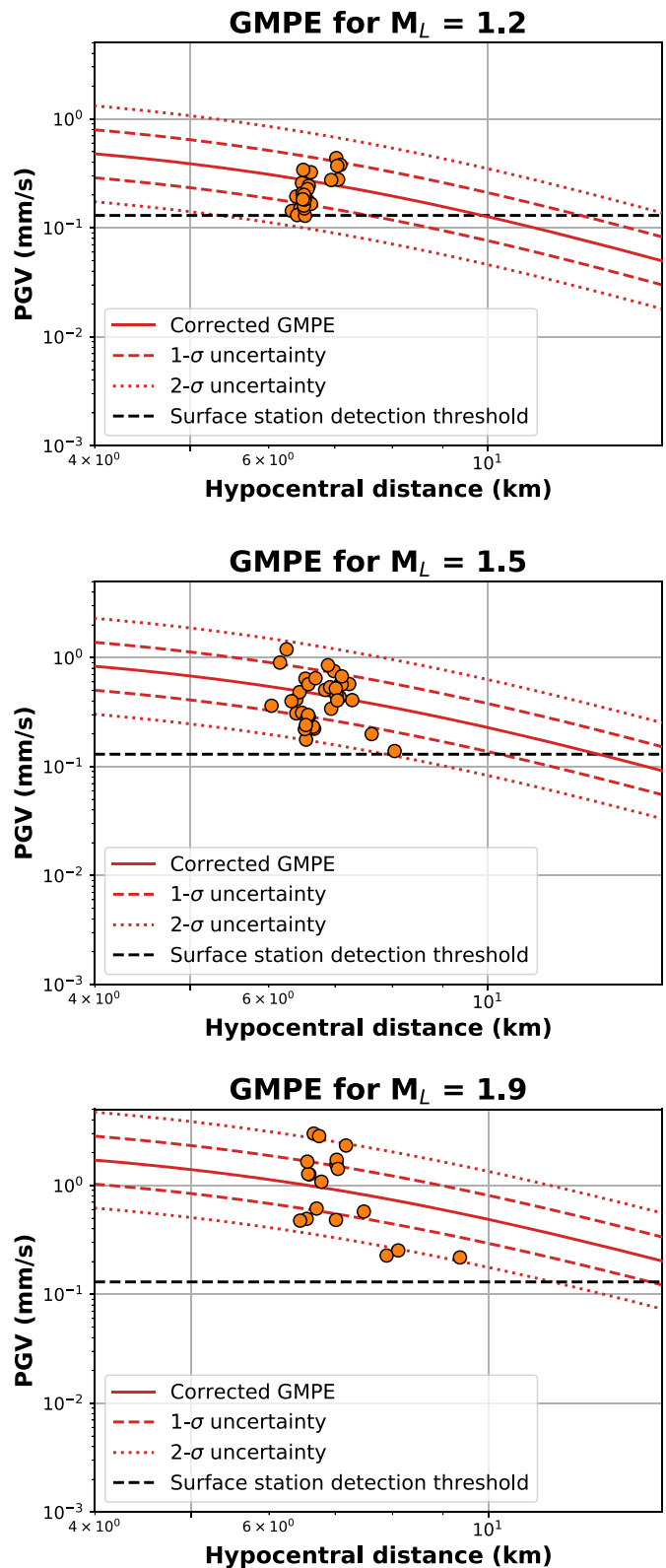
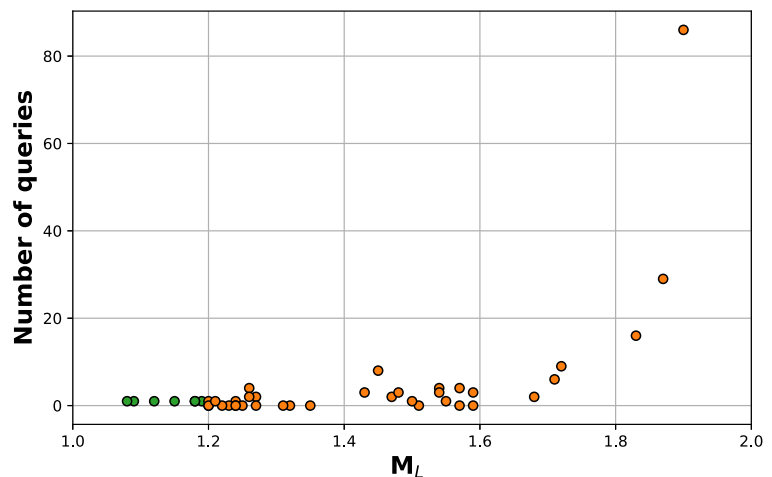


Fig. 11 Combined number of public responses recorded by ISUH and by St1 DH for **Amber** events and some of the largest **Green** events, with respect to the event magnitudes



event recorded, i.e., the M_L 1.9 event, which generated eight reports to St1 DH and 78 reports to ISUH.

Some events were reported for magnitudes as low as M_L 0.9, and the reports mentioned that events were usually heard rather than felt and compared them to distant thunder. Vibrations were only reported as felt (although very rarely) for events with magnitudes greater than M_L 1.5. The fact that such low-magnitude events could be heard was unexpected, although the occurrence of audible signal was clearly related to the occurrence of larger events. It has been speculated that the audible signal is generated by a water table in the nearby shallow bay aquifer, due to elastic waves resonating in shallow sedimentary layer, and the origin of audible signals is currently under investigation.

In addition to the general public response, the $M_L \geq 1.8$ events received some press attention. The content of the press releases was factual and without any apparent negative bias against St1 DH project. The press reports also described the events as being heard and sounding like thunder and sometimes observing some window vibrations. The reports indicated that the level of vibrations generated by the events was much too low to create any damage to structures and could not pose a threat, showing that the public and media understanding on this practical point was clear.

Levels and thresholds of public perception show that with the methodology adopted to design the TLS, the thresholds were set at levels that were both conservative enough to protect the population and the built environment, but also practical enough to enable St1 DH to proceed with their operations in a productive way.

5 Post-stimulation TLS revisions

The implementation of the TLS yielded a large amount of new and site-specific data along with a range of lessons learnt. This section details how this information can be used to refine the TLS for future stimulations.

5.1 Proposed new thresholds

The corrected GMPE presented in Section 4.1 is site-specific and has significantly smaller uncertainties than the GMPEs that were used when designing the TLS before the stimulation, so that surface expression of induced seismicity can be predicted with improved confidence for future stimulations.

As highlighted in Section 2.2, the philosophy behind the TLS magnitude thresholds was to correlate the magnitude thresholds with the probabilities of exceeding the different PGV thresholds and was approved by the regulator. Given the precedence and successful implementation, this approach is maintained, but the thresholds can now be refined with the new site-specific GMPE. The probabilities of exceedance for the different TLS triggers can be maintained as the same levels as for the original TLS (Section 2.2):

- 2% for the **Amber** scenario where the TLS alert is based on the combination of a PGV exceedance and an associated magnitude; and
- 10% for the **Amber** scenario where the TLS alert is solely based on magnitude; and
- 2% for the **Red** scenario.

Given the improved confidence in ground motion prediction, it could be argued that the first condition could be omitted in order to keep a TLS solely based on magnitude, which would be more practical to implement.

Figure 12 shows the probability of exceedance of the three PGV thresholds (0.3, 1, and 7.5 mm/s) at the epicentre of events located at 6-km depth, as a function of their magnitude. The 10% probability for the **Amber** curve and 2% probability for the **Red** curve would lead the following new TLS thresholds:

Green $M_L < 1.6$;

Amber $1.6 \leq M_L < 2.5$, while $PGV \geq 1$ mm/s should be reported; and

Red $M_L \geq 2.5$, while $PGV \geq 7.5$ mm/s should be reported.

Although the corrected GMPE predicts larger median ground motions than the GMPEs used in the pre-stimulation design of the TLS, the uncertainties of the corrected GMPE are significantly reduced. Therefore, the approach based on low probabilities of exceedance still yields slightly higher magnitudes for the TLS thresholds.

The new thresholds, computed using the same methodology used to design the TLS before stimulation, seem consistent with the different observations during the stimulation:

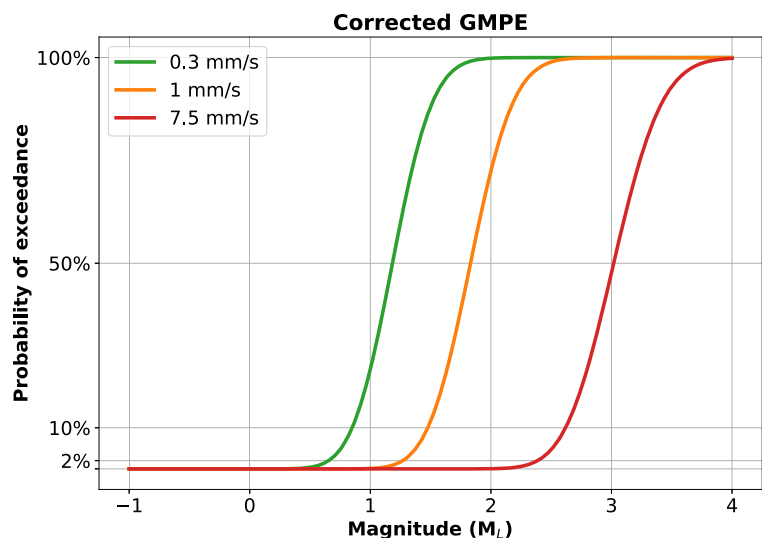
- Except for one $M_L 1.55$ event, only $M_L \geq 1.68$ events resulted in $PGV \geq 1$ mm/s during OTN-3 stimulation;
- Out of ten events with $1.5 \leq M_L \leq 1.6$, only one resulted in a $PGV \geq 1$ mm/s;
- Only $M_L \geq 1.8$ events resulted in more than ten public reports during OTN-3 stimulation; and
- The PGVs measured and the rate of public response for events up to $M_L 1.9$ were low.

These observations highlight the fact that the methodology used to quantitatively determine the TLS threshold appears to be robust with respect to qualitative observations during the stimulation of OTN-3.

5.2 Forward-looking models

Introduction: The TLS has been designed based on conservative thresholds and associated mitigation measures. As detailed in the introduction (Section 1), such TLSs have been shown to be effective at reducing the risk and quite simple to implement (e.g., Wiemer et al. 2017). They are, however, based on a reactive scheme: measures are implemented only after an event has occurred, which is why the thresholds must be conservative. Recent “forward-looking” models have been developed based on physical or statistical models, which aim at predicting the maximum magnitude of induced seismicity based on the past seismicity. Three such predictive models prevail in the literature (in chronological order):

Fig. 12 Probability of exceedance of the three PGV thresholds of the TLS at the epicentre of events at 6-km depth, as a function of the event magnitude, for the corrected GMPE



- The McGarr (2014) model;
- The Van der Elst et al. (2016) model, based on the Shapiro et al. (2010) seismogenic-index equation; and
- The Galis et al. (2017) model.

Although these models were not officially incorporated in the TLS, the predictions of the Van der Elst et al. (2016) and the Galis et al. (2017) models have been examined during pending stimulation of OTN-3 to estimate what would be the maximum magnitude observed by the end of the stimulation (Kwiatek et al. 2019).

In the following sections, we therefore provide a discussion on these different forward-looking models and show preliminary results on how they apply to the stimulation of OTN-3. We show that the predictions from these models overestimated the final maximum magnitude observed during the stimulation. The following three sections introduce the models, followed by a comparison of their use for the OTN-3 stimulation.

McGarr 2014 model: While the assumptions of the McGarr (2014) model are invalid for this application, we have nevertheless included a review of its use as it is commonly used for maximum magnitude estimation. The McGarr (2014) model is a blend of a physical and statistical model, which assumes that all the strain caused by the injection of water is released seismically. The seismicity is assumed to follow a GR distribution with a b value of 1. Based on these assumptions, McGarr (2014) proposed that the seismic moment of the largest earthquake \mathcal{M}_o^{\max} was:

$$\mathcal{M}_o^{\max} = G \times V, \tag{4}$$

where G is the shear modulus (or modulus of rigidity) of the rock and V is the net injected volume of fluids in the rock. The moment magnitude of the largest event M_{\max} is related to its moment through the Hanks and Kanamori (1979) moment-magnitude formula:

$$M_{\max} = \frac{2}{3} (\log \mathcal{M}_o^{\max} - 9.1). \tag{5}$$

Although this model is regularly cited and referred to, it should not apply in the case of the stimulation of OTN-3 for the three following reasons:

- Recent experiments of water injected in faults (e.g., Guglielmi et al. 2015) have shown that the

the strain caused by fluid injection may be accommodated by aseismic creep on the fault rather than by seismic events, which would be in direct contradiction with the main assumption of the McGarr (2014) model that all the strain caused by the injection of water is released seismically;

- The McGarr (2014) model predicts that the possible maximum magnitude is the same for all sites, regardless of their geology, seismotectonic setting, or background seismicity; and
- Finally, Eq. 4 assumes a b value of one. More generally, the McGarr (2014) model can only handle b value up to 1.5: the model yields negative seismic moments (i.e., imaginary magnitudes) for $b \geq 1.5$, while in practice such b values greater than 1.5 are quite common in geothermal induced seismicity (e.g., Bachmann 2012; Grünthal 2014).

Hallo et al. (2014) proposed further modifications of the McGarr (2014) model, such as including a seismic-efficiency prefactor to Eq. 4. This would indeed solve the first issue with the model, and potentially the second issue, but the third issue would remain.

Finally, it is worthwhile to note that Mignan et al. (2019) showed that the recent Pohang earthquake is far above the predicted limit by the McGarr (2014) model (Mignan et al. 2019). This is due to the fact that the Pohang earthquake released previously stored elastic energy, while the McGarr (2014) model locks the budget of seismic energy to the injected hydraulic energy.

Van der Elst et al. 2016 model: The Van der Elst et al. (2016) model is a purely statistical model based on the Shapiro et al. (2010) formula, which states that induced seismicity follows a usual GR magnitude-frequency distribution, where the a value depends on the net volume V of fluids injected and a seismogenic index, Σ (Shapiro et al. 2010; Van der Elst et al. 2016):

$$\log N_{\geq M_L} = a - bM_L, \tag{6}$$

$$a = \log V + \Sigma, \tag{7}$$

where $N_{\geq M_L}$ is the number of events with magnitude greater than or equal to M_L ; V is the net injected volume of fluids, in cubic meters; and Σ is the seismogenic index. Van der Elst et al. (2016) used this formula to compute the expected maximum magnitude

during injection, together with confidence intervals:

$$M_{\max} = \frac{\log V + \Sigma}{b}. \tag{8}$$

The Van der Elst et al. (2016) model is a parameter-dependent model, as the seismogenic index Σ and the b value must be estimated to compute the maximum expected magnitude at a given injected volume. These two parameters can be estimated after a sufficient volume of fluid has been injected so that enough events have been recorded to compute the GR model parameters. The main assumptions in this model are as follows:

- Both the seismogenic index Σ and the b value are constant throughout fluid injection or at least vary over a time period much larger than the prediction window; and
- The seismicity follows a GR magnitude-frequency distribution.

These assumptions are a priori reasonable, although we will see later that they might not have been entirely satisfied during the stimulation of OTN-3.

Galis et al. 2017 model: The Galis et al. (2017) model is a physical model based on rupture physics, which assumes that ruptures on faults are self-arrested, i.e., do not propagate across the entire fault (Galis et al. 2017). Based on these assumptions, Galis et al. (2017) propose that the seismic moment of the largest earthquake \mathcal{M}_o^{\max} is:

$$\mathcal{M}_o^{\max} = \gamma \times V^{2/3}, \tag{9}$$

where γ is a reservoir-dependent parameter and V is the net injected volume of fluids in the rock. Similarly to the seismogenic index Σ in the Van der Elst et al. (2016) model, the parameter γ in the Galis et al. (2017) model can be estimated after an initial volume of fluids has been injected and some seismicity has been recorded. In the case where the b value is equal to 1, there is actually a direct relationship between the Galis et al. (2017) γ parameter and the Shapiro et al. (2010) seismogenic index (Galis et al. 2017):

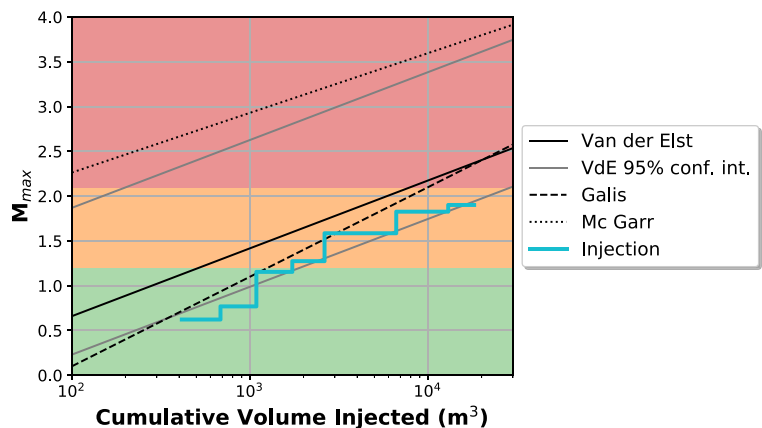
$$\Sigma = \frac{2}{3} (\log \gamma - 9.1). \tag{10}$$

Application to OTN-3 stimulation: Figure 13 shows the maximum magnitudes predicted by the three forward-looking models presented above as a function of the net injected volume of water during the stimulation of the well OTN-3. The evolution of the maximum magnitude observed as the injection progressed is also plotted in Fig. 13.

The parameters for the Van der Elst et al. (2016) model have been computed from the GR parameters in Fig. 5 and based on a total injected volume of fluids of 18,160 m³. This yields a b value of 1.32 and a seismogenic index value, Σ , of -1.13 . The γ value in the Galis et al. (2017) model has been taken as $\gamma = 1.8 \times 10^6$, based on the maximum magnitude recorded during the first 3000 m³ of water injected.

Figure 13 suggests that, as expected, the McGarr (2014) model largely over-predicted the maximum magnitude for the OTN-3 stimulation. For example, the McGarr (2014) method predicted a magnitude $M_L 2.3$ event after only 100 m³ of injection and $M_L 3.7$ by the end of the stimulation. As noted above, in this

Fig. 13 Predictions of the maximum magnitude with respect to the net injected volume, according to the McGarr (2014), Van der Elst et al. (2016), and Galis et al. (2017) models, compared with the maximum magnitude observed during the injection. The 95% confidence intervals for the Van der Elst et al. (2016) model are indicated in grey



form, the McGarr (2014) model is not appropriate for the geologic setting at this project site.

Both the Van der Elst et al. (2016) and the Galis et al. (2017) models predicted a maximum magnitude of $M_L 2.4$ by the end of the stimulation, after 18,160 m³ of water would be injected. The probabilistic nature of the Van der Elst et al. (2016) model provides a tool to compute statistics on the maximum magnitude expected during the stimulation. Figure 14 shows the PDF of the maximum magnitude and the exceedance probabilities of different magnitudes by the end of the stimulation of OTN-3. The following predictions of the model should be highlighted:

- The Van der Elst et al. (2016) model predicts a 90% probability to have had a **Red** alert during the stimulation of OTN-3;

- The 95% confidence intervals on the maximum magnitudes are at 1.94 and 3.58; and
- The maximum magnitude observed of $M_L = 1.9$ lies right at the edge of the 95% confidence intervals and would have had a probability of 98.5% to be exceeded during the stimulation of OTN-3, according to the Van der Elst et al. (2016) model.

The fact that both the Van der Elst et al. (2016) and the Galis et al. (2017) models overestimated the maximum magnitude owes to the apparent non-linear deviation from the GR distribution for $M_L \geq 1.5$, as noted in Fig. 5. Both the Van der Elst et al. (2016) and the Galis et al. (2017) models implicitly assume uniform magnitude-frequency properties of the seismicity at all magnitudes and this is why they over-predict the maximum magnitude in this case. Besides, by

Fig. 14 Probability density function of the maximum magnitude expected and probability of exceedance of different event magnitudes during the stimulation of OTN-3, according to the Van der Elst et al. (2016) model

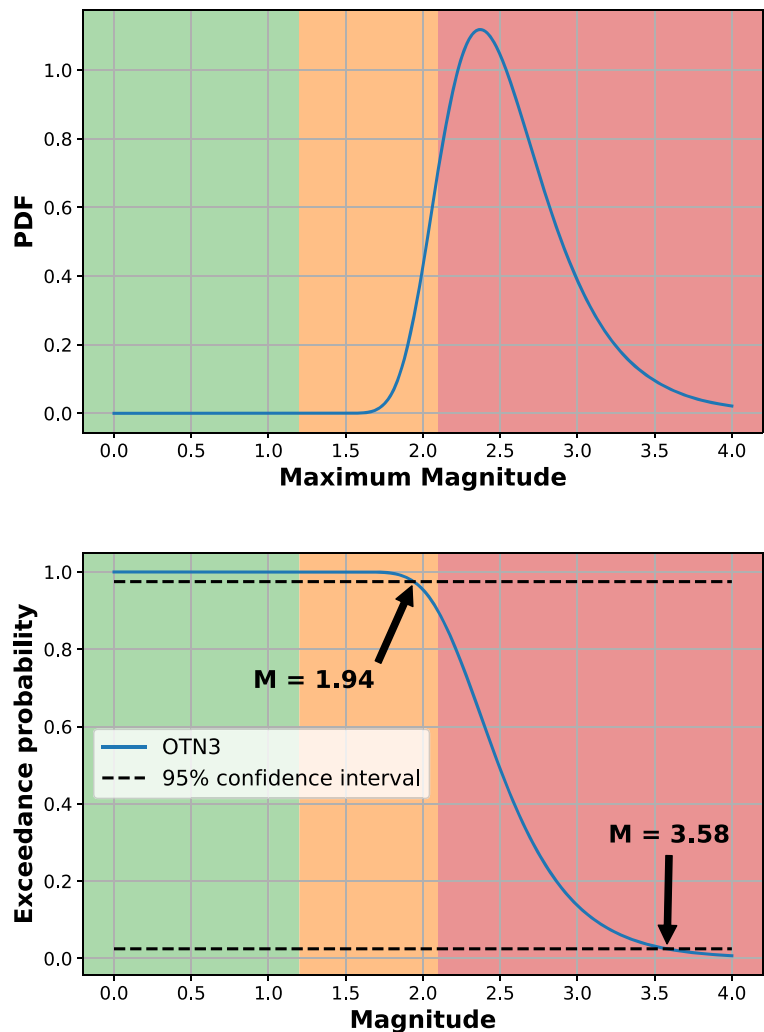
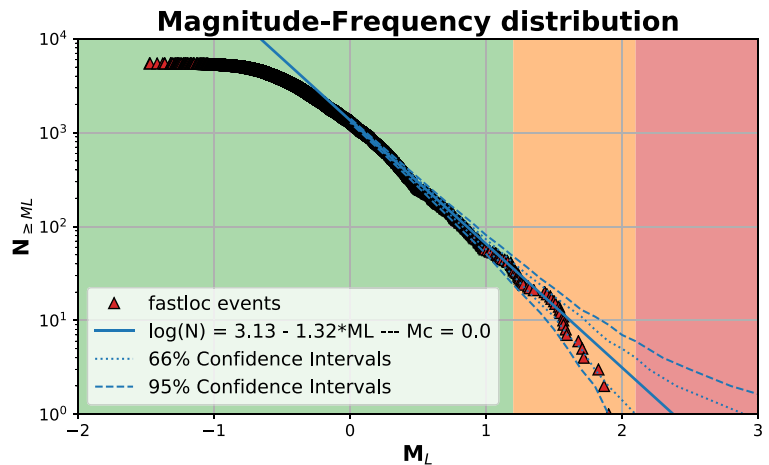


Fig. 15 Magnitude-frequency GR distribution of events as in Fig. 5 but with 66% and 95% confidence intervals to the Gutenberg-Richter distribution represented



only looking at the total net volume injected, these models do not account for the time history of the injection. In particular, they do not consider the time evolution of hydraulic energy build up and energy dissipation during periods of pumping and resting (Kwiatek et al. 2019).

To further assess whether the deficit in large events is a simple statistical realisation of a linear Gutenberg-Richter distribution or a true feature, we have added the 66% and 95% confidence intervals to the GR distribution in Fig. 15. These confidence intervals were computed by randomly generating 10,000 earthquakes catalogues of $1357 M_L \geq 0$ events, with magnitudes generated according to an exponential distribution calibrated on the GR parameters. Figure 15 shows that the larger magnitude end of the magnitude-frequency distribution of the events occurring during the stimulation falls right at the lower edge of the 95% confidence intervals of the GR distribution. This is consistent with the fact that the maximum magnitude observed is just outside the 95% confidence intervals of the predictions from the Van der Elst et al. (2016) model (Figs. 13 and 14). Indeed, the confidence intervals of the Van der Elst et al. (2016) model correspond to the confidence intervals at $N_{\geq M_L} = 1$ in Fig. 15. This suggests that the dip of the magnitude-frequency distribution at $M_L \geq 1.5$ has less than 5% probability to be a simple statistical deviation from the Gutenberg distribution and can therefore be considered unlikely to be a simple statistical effect. Similar results were obtained by Kwiatek et al. (2019) for the extended catalogue.

Similar deviations from the GR relationship at higher magnitudes have been observed at other deep geothermal stimulation projects globally, such as in Soultz-sous-Forêts, France (Shapiro et al. 2013). This deviation has also been observed in waste-water injection and hydraulic fracturing projects in Canada (Schultz 2018). However, such a deviation was not observed at the geothermal stimulation projects in Basel, Switzerland (Shapiro et al. 2013), Rosemanowes, Cornwall, UK (Bachmann 2012), or The Geysers, USA (Kwiatek et al. 2015).

Understanding the mechanisms which govern whether or not elevated seismic events follow the Gutenberg-Richter relationship is critical to the management of geothermal stimulation and the implementation of forward-looking models into more advanced types of TLSs. By understanding under what conditions higher-magnitude events will be less than that predicted by the Gutenberg-Richter distribution, greater clarity will be obtained in project risks during stimulation, and will lead to more effective mitigation and management measures. While understanding the physical mechanisms at play lies well outside the scope of this project, the geothermal research community should recognise the importance of this area of research.

6 Conclusion

The stimulation of OTN-3 was the deepest geothermal well stimulation in the world so far. A TLS was

designed in order to regulate and mitigate seismic hazard related to the fluid-induced seismicity. The validation of the TLS by the local regulator was contingent upon demonstrating fitness for purpose and robustness with respect to specific challenges:

- Owing to the low levels in seismicity in Finland, there was very little seismic data available for the calibration and design of the TLS;
- The stimulation took place beneath a large urban area, which meant a large and densely populated area with multiple sensitive receptors and high levels of vibration noise (especially from construction blasting), which posed a risk of false alerts; and
- The population was reportedly very sensitive to earthquakes, leading to the potential risk of bad public perception, which has been known to shut off geothermal projects in the past (Diehl et al. 2017; Giardini 2009).

On the operator side, TLS thresholds that would be too conservative might have represented a strong operational burden and might have impaired the financial viability of the project (Mignan et al. 2019).

In this paper, we detailed the elements of the TLS put in place to effectively mitigate the hazard and risk of induced seismicity during the stimulation. We present a methodology to establish the TLS thresholds, which can easily be transferred to other projects in different areas of the world, regardless of their depth. This methodology can be used in environments rich or poor in seismic data, and enables for straightforward updates of the TLS as data becomes available. The methodology would only require further adaptation if it were to be applied in a seismically active area, in order to better take into account the potential natural seismicity.

The design of the TLS thresholds relies on two parameters, which are usually quite practical to agree on with a regulator: the acceptable levels of surface ground motion and the probability to reach them. Reaching a consensus on acceptable levels of ground motion is usually facilitated by existing regulations and best practices, so that the discussion can be streamlined by factual evidence. In terms of the probabilities to reach agreed levels of ground motions, the range of 2 to 10% used in the presented design proved to be conservative enough to mitigate seismic hazard for the population and the

built environment, without impairing the stimulation operations.

The implementation of the TLS during the stimulation of OTN-3 progressed smoothly and the TLS thresholds turned out to be well adapted to mitigate seismic hazard. The data collected during the stimulation was used to compute a site-specific GMPE, with uncertainties reduced by a factor of three, compared with the more general GMPEs used in the initial design of the TLS. This site-specific GMPE, together with lessons learnt during the stimulation, was used to design a revised TLS.

In the future, it would be desirable to pair the traditional TLS with a more advanced ATLS by relying on forward-looking models. In the case of the stimulation of OTN-3, the existing forward-looking models would have over-predicted the level of hazard, which would have had a negative impact on the operations. The forward-looking models still appear to require some site-specific tuning before they can be used as a reliable tool for prediction of induced seismicity during a similar geothermal well stimulation project.

References

- Ader T, Chendorain M, Free M, Saarno T, Heikkinen P, Malin PE, Leary P, Valenzuela S, Kwiatek G, Dresen G, Bluemle F, Vuorinen T (2019) Setting up traffic light system thresholds for geothermal stimulation in Helsinki, Finland. In: 2019 SECED Conference, Greenwich, London, accepted
- Bachmann CE (2012) Induced seismicity in Rosemanowes, 16 pp. GEISER report D5.3 – Hazard models for specific test areas
- Bommer JJ, Oates S, Cepeda JM, Lindholm C, Bird J, Torres R, Marroquín G, Rivas J (2006) Control of hazard due to seismicity induced by a hot fractured rock geothermal project. *Eng Geol* 83(4):287–306. <https://doi.org/10.1016/j.enggeo.2005.11.002>
- Bommer JJ (2017) Predicting and monitoring ground motions induced by hydraulic fracturing. OGA-commissioned paper
- Bosman K, Baig A, Viegas G, Urbancic T (2016) Towards an improved understanding of induced seismicity associated with hydraulic fracturing. *First break* 34:61–66
- Broccardo M, Mignan A, Wiemer S, Stojadinovic B, Giardini D (2017) Hierarchical Bayesian modeling of fluid-induced seismicity. *Geophys Res Lett* 44:11,357–11,367. <https://doi.org/10.1002/2017GL075251>
- BSI (1993) Evaluation and measurements for vibration in buildings - Part 2: Guide to damage levels due to groundborne vibrations. BS 7385-2, British Standards Institute
- BSI (2008) Guide to evaluation of human response to vibration in buildings - Part 1: Vibration sources other than blasting. BS 6472-1, British Standards Institute

- Cuenot N, Dorbath C, Dorbath L (2008) Analysis of the Microseismicity induced by fluid injections at the EGS site of soultz-sous-forêts (Alsace, France): implications for the characterization of the geothermal reservoir properties. *Pure Appl Geophys* 165:797–828
- Diehl T, Kraft T, Kissling E, Wiemer S (2017) The induced earthquake sequence related to the St. Gallen deep geothermal project (Switzerland): fault reactivation and fluid interactions imaged by microseismicity. *J Geophys Res Solid Earth*, <https://doi.org/10.1002/2017JB014473>
- Douglas J, Edwards B, Convertito V, Sharma N, Tramelli A, Kraaijpoel D, Cabrera BM, Maercklin N, Troise C (2013) Predicting ground motion from induced earthquakes in geothermal areas. *Bull Seismol Soc Am* 103:1875–1897
- Ellsworth WL (2013) Injection-induced earthquakes. *Science* 341:1225942
- Galis M, Ampuero JP, Mai PM, Cappa F (2017) Induced seismicity provides insight into why earthquake ruptures stop. *Science Advances* 3: eaap7528. Available: <https://doi.org/10.1126/sciadv.aap7528>
- GEISER (2013) D5.3 - Hazard models for specific test areas, Report
- Giardini D (2009) Geothermal quake risks must be faced. *Nature* 462:848–849
- Gischig VS, Wiemer S, Alcolea A (2014) Balancing reservoir creation and seismic hazard in enhanced geothermal systems. *Geophys J Int* 198(3):1585–1598
- Grünthal G (2014) Induced seismicity related to geothermal projects versus natural tectonic earthquakes and other types of induced seismic events in Central Europe. *Geothermics*, <https://doi.org/10.1016/j.geothermics.2013.09.009>
- Guglielmi Y, Cappa F, Avouac JP, Henry P, Ellsworth D (2015) Seismicity triggered by fluid injection-induced aseismic slip. *Science* 348:1224–1226
- Haering MO, Schanz U, Ladner F, Dyer BC (2008) Characterisation of the Basel 1 enhanced geothermal system. *Geothermics* 37:469–495
- Halldorsson B, Ólafsson S, Snaebjörnsson JT, Sigurdsson SU, Rupakhety R, Sigbjörnsson R (2012) On the effects of induced earthquakes due to fluid injection at Hellisheidi geothermal power plant, Iceland. In: 15th world conference on earthquake engineering
- Hallo M, Oprsal I, Eisner L, Ali MY (2014) Prediction of magnitude of the largest potentially induced seismic event. *J Seismol* 18:421–431
- Hanks TC, Kanamori H (1979) A moment magnitude scale. *J Geophys Res* 84(B5):2348–2350. <https://doi.org/10.1029/JB084iB05p02348>
- Kwiatek G, Martínez-Garzón P, Dresen G, Bohnhoff M, Sone H, Hartline C (2015) Effects of long-term fluid injection on induced seismicity parameters and maximum magnitude in northwestern part of the Geysers geothermal field. *J Geophys Res Solid Earth* 120, <https://doi.org/10.1002/2015JB012362>
- Kwiatek G, Saarno T, Ader T, Bluemle F, Bohnhoff M, Chendorain M, Dresen G, Heikkinen P, Kukkonen I, Leary P, Leonhardt M, Malin P, Martínez-Garzón P, Passmore K, Passmore P, Valenzuela S, Wollin C (2019) Controlling fluid-induced seismicity during a 6.1-km-deep geothermal stimulation in Finland. *Sci Adv* 5(5):eaav7224. <https://doi.org/10.1126/sciadv.aav7224>
- McGarr A (2014) Maximum magnitude earthquakes induced by fluid injection. *J Geophys Res Solid Earth* 119:1008–1019. <https://doi.org/10.1002/2013JB010597>
- Mignan A, Broccardo M, Wiemer S, Giradini D (2017) Induced seismicity closed-form traffic light system for actuarial decision-making during deep fluid injections. *Sci Rep* 7:13607. <https://doi.org/10.1038/s41598-017-13585-9>
- Mignan A, Karvounis D, Broccardo M, Wiemer S, Giardini D (2019) Including seismic risk mitigation measures into the Levelized cost of electricity in enhanced geothermal systems for optimal siting. *Appl Energy* 238:831–850. ISSN 0306–2619, <https://doi.org/10.1016/j.apenergy.2019.01.109>
- RIL 253-2010 (2010) Rakentamisen aiheuttamat tärinät 2010. Helsinki: Suomen Rakennusinsinöörien Liitto
- Lund B, Malm MM, Mäntyniemi PB, Oinonen KJ, Tiira T et al (2015) Evaluating seismic hazard for the Hanhikivi nuclear power plant site, seismological characteristics of the source areas, attenuation of seismic signal, and probabilistic analysis of seismic hazard. F-Consult Ltd, 258 p Saari J (ed)
- Shapiro SA, Dinske C, Langenbruch C (2010) Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations. *Leading Edge* 29(3):304–309
- Shapiro SA, Krüger OS, Dinske C (2013) Probability of inducing given-magnitude earthquakes by perturbing finite volumes of rocks. *J Geophys Res Solid Earth* 118, <https://doi.org/10.1002/jgrb.50264>
- Schultz R (2018) Bridging gaps in induced seismicity hazard forecasting in Alberta, Canada, presentation at the 36th general assembly of the European seismological commission 6 september 2018, Valletta, Malta
- Uski M, Tuppurainen A (1996) A new local magnitude scale for the Finnish seismic network. *Seism Source Parameters Microearthquakes Large Events* 261:23–37
- Van der Elst NJ, Page MT, Weiser DA, Goebel THW, Hosseini SM (2016) Induced earthquake magnitudes are as large as (statistically) expected. *J Geophys Res Solid Earth* 121:4575–4590. <https://doi.org/10.1002/2016JB012818>
- Westaway R, Younger PL (2014) Quantification of potential macroseismic effects of the induced seismicity that might result from hydraulic fracturing for shale gas exploitation in the UK. *Q J Eng Geol Hydrogeol* 47(4):333–350. <https://doi.org/10.1144/qjegh2014-011>
- Wiemer S, Kraft T, Trutnevyte E, Roth P (2017) ‘Good practice’ guide for managing induced seismicity in deep geothermal energy projects in Switzerland. In: Proceedings from the 20th EGU General Assembly held 4-13 April, 2018, Vienna, Austria, pp 15021

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.