**An In-Depth Seismological Analysis Revealing a Causal Link Between the 2017 Mw 5.5 Pohang Earthquake and EGS Project**


1School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea, 2Division of Earth Environmental System Science, Pukyong National University, Busan, South Korea, 3Department of Structural Systems and Site Evaluation, Korea Institute of Nuclear Safety, Daejeon, South Korea, 4Faculty of Earth Systems and Environmental Sciences, Chonnam National University, Gwangju, South Korea, 5Swiss Seismological Service, ETH Zurich, Zurich, Switzerland, 6Stanford Center for Induced and Triggered Seismicity, Department of Geophysics, Stanford University, Stanford, CA, USA, 7Energy Science Center, Department of Earth Sciences, ETH Zurich, Zurich, Switzerland

**Abstract** Hydraulic injection by the Pohang enhanced geothermal systems (EGSs) has been suspected to trigger the 2017 moment magnitude (Mw) 5.5 Pohang earthquake in South Korea. The last stimulation experiment in the EGS was conducted only 2 months before the disaster, which has led to this suspicion. In this study, we conducted a seismic analysis on the earthquakes that have occurred around the EGS site in the past 10 years. The study included the construction of a velocity model, earthquake detection, the determination of hypocenters, magnitudes, focal mechanisms, and stress inversion, and a clustering analysis. No seismic activity was detected near the study area until November 2015 when there was a loss of a large quantity of heavy drilling mud. For three stimulations of a geothermal well, earthquakes sequentially migrated to the southwest along a fault plane, leading to the location of the mainshock. The delineated fault plane crossed the injection well at approximately 3,800 m, which corresponds to the borehole interval of not only the mud loss but also the breakage of the well’s casing due to the mainshock rupture. These findings can be treated as empirical evidence for the hypothesis that the 2017 Mw 5.5 Pohang earthquake was initiated on a critically stressed fault zone by the anthropogenic activity of fluid injection, consequentially releasing accumulated strain energy via tectonic loading.

1. **Introduction**

The 2017 Pohang earthquake of moment magnitude (Mw) 5.5 struck Pohang, South Korea, on 15 November 2017, at 05:29:31 UTC (hereinafter, the Pohang earthquake) (Grigoli et al., 2018; Kim et al., 2018; Lee et al., 2019). The magnitude of the earthquake is the second largest in the Korean Peninsula since official earthquake monitoring was first initiated in 1978 by the Korea Meteorological Administration (KMA). However, it is an earthquake that inflicted the most damage in the region (Choi et al., 2019; Kang et al., 2019). Concerned authorities suspected that the enhanced geothermal systems (EGSs) geothermal electricity project caused this earthquake (Ellsworth et al., 2019; Lee et al., 2019).

In the project, two geothermal wells, PX-1 and PX-2, with a depth of over 4 km were drilled to expedite water circulation through hot dry rock characterized by artificially enhanced permeability induced by hydraulic stimulation (Kim et al., 2018; Hofmann et al., 2019). Through these geothermal wells, the EGS project team conducted five hydraulic stimulation experiments (Kim et al., 2017). The Pohang earthquake occurred approximately 2 months after the completion of the final hydraulic stimulation experiment (i.e., 18 September 2017).

For various reasons, fluid injections can induce seismicity (Ellsworth, 2013). EGS projects, such as in the Pohang case, require fluid injection processes at high pressures through deep boreholes of a few kilometers (Zang et al., 2014). Such high pressure exerted on the surrounding rock during fluid injection causes cracks that increase the permeability of the rock mass, thus making it easy to extract heated fluid from deeper areas according to the geothermal gradient. This crack-forming process can induce seismicity (Ellsworth, 2013; Maxwell, 2014). Therefore, several studies have questioned the relationship between the occurrence of the 2017 Pohang earthquake and activities of the EGS project (Grigoli et al., 2018; Kim et al., 2018). However,
the magnitude of the Pohang earthquake was much larger than the expected magnitude estimated based on known relationships (McGarr, 2014; Galis et al., 2017; Van der Elst et al., 2016), which has provoked a heated debate (Geological Society of Korea, 2019; Lee et al., 2019). For reference, the two largest events related to the EGS project prior to the Pohang earthquake were the 2006 Mw 3.4 Basel earthquake in Switzerland and the 2003 M 3.7 Cooper basin earthquake in Australia, where the released seismic moments were only 0.07% and 0.20% of the seismic moment of the Pohang earthquake (Baisch et al., 2006; Deichmann & Giardini, 2009).

The purpose of this study is to provide a detailed analysis of the seismic activity that occurred near the Pohang EGS site and reveal the relationship between seismicity and injection activities. We have developed a one-dimensional (1-D) seismic velocity model for the EGS project site to better analyze the seismicity of the area. Furthermore, we introduce a series of precise analyses on the seismicity before and during the entire period of the hydraulic stimulation experiments and the Pohang earthquake. These analyses include the absolute and relative locations, magnitude determination, and focal mechanism inversion of each seismic event. The results of these analyses were used for the investigation of the spatiotemporal distribution of seismic activity in and around the EGS project site, as well as to infer the fault geometry responsible for the mainshock. Finally, the hypocenter distribution was described in terms of the state of the geothermal wells to scrutinize the possible relationship between the hydraulic stimulation experiments and seismic activity.

2. Materials and Methods

2.1. Geological Tectonic Setting

The epicenter of the Pohang earthquake was located in the Pohang basin, which is in the southeastern part of the Cretaceous Gyeongsang basin formed during the Miocene epoch. The geology of the epicentral area in the northeastern Pohang basin can be described as Quaternary alluvium overlying sandstone and mudstone of the Tertiary Yeonil Group (Figure 1a). The Pohang basin developed from transtensional subsidence related to NNW-SSE trending dextral shear movements that acted along the eastern coast of the Korean Peninsula with the opening of the East Sea (Sea of Japan). This was accompanied by the separation of the Japanese Island Arc from the far-eastern margin of the Eurasian continent, including the Korean Peninsula, in the Miocene (Son et al., 2015). Basement rock beneath the Pohang basin is composed of granitic rocks and a complex of gneisses and schists that form the Yeongnam massif (or Sobaeksan massif) (Geological Society of Korea, 1998; Lee & Cho, 2012).

Several dominant types of faults bound the Pohang basin, that is, NE to NNE trending normal faults and NW trending transfer faults to the west, NNW trending dextral strike-slip faults to the south, and NE to NNE trending sinistral strike-slip to dip-slip faults (Cheon et al., 2012; Son et al., 2002; Son et al., 2013). Among these faults, the Gokgang fault near the epicenter is known to be a normal fault with an NE strike and SE dip (Yun, 1994). Recent outcrop observations have shown that, although the antithetic faults have similar strikes with the Gokgang fault, the NW dips have formed a conjugate normal fault system in this area (Choi et al., 2019). Considering that the southeastern part of the Korean Peninsula is exposed to maximum compressional stress in the ENE-WSW direction (Chang et al., 2010), faults with a normal faulting sense related to the evolution of the Pohang basin can be reactivated in a reverse faulting scenario when exposed to the current stress field.

2.2. Data Acquisition

The purpose of this study is to investigate the mainshock and earthquakes that have occurred near the EGS site. We used the seismic waveform recorded at three different seismic networks to analyze earthquakes. The first network was a permanent network operated by the KMA, the Korea Institute of Geoscience and Mineral Resources (KIGAM), and the Korea Hydro and Nuclear Power (KHNP), which are denoted as KS, KG, and KN, respectively. This network consists of broadband velocity seismometers, short-period velocity seismometers, and accelerometers (Figure 1a). We mainly used stations located within approximately 50 km of the mainshock epicenter for the analysis but also used stations farther afield for the magnitude estimation of the mainshock and for the determination of the focal mechanism solutions of several events. The second network was a temporary network deployed as part of the EGS project, consisting of various instruments: (1) Eleven surface velocity seismometers (PH surface in Figure 1), (2) nine borehole seismometers installed at depths between 100 and 150 m (PH borehole in Figure 2), (3) a borehole geophone array deployed in PX-
2 at depths between 1,360 and 1,520 m, with an interstation distance of 10 m (17 sensors), which operated for 1 month from July 2017, (4) a vertical seismic profile (VSP) installed in PX-1 or PX-2 at depths between 1,350 and 1,550 m, which partly operated during the first, second, and third stimulations (from 27 January 2016 to 2 February 2016, from 24 December 2016 to 11 January 2017, and on 5 April 2017), and (5) a surface (POH01) and borehole seismometer (BH4) installed at a depth of approximately 2,300 m and located approximately 2 km SW of PX-2 (Figures 1b and 2a). The last network was a temporary network (PH short-period) installed after the third stimulation (Kim et al., 2018), which consisted of eight short-period velocity seismometers located within 3 km of the mainshock epicenter (Figure 1b). Detailed information on the seismic stations used in this study is summarized in Figure S1 and Table S1. In addition to the seismic waveforms, detailed information on the fluid injections from the EGS project team were collected to obtain the relationships between the earthquakes and injections (Figures 3 and S2–S6).

2.3. Construction of a Local 1-D Velocity Model

Seismic source parameters, such as hypocenters, focal mechanisms, seismic moments, and stress drops, were accurately estimated by constructing an appropriate seismic velocity model. As the Pohang EGS site is situated in the Pohang basin, which is expected to have significantly low-velocity layers with shallow depths, hypocentral parameters calculated via a regional velocity model that reflects the average medium properties may result in improper interpretations of the relationship between hydraulic stimulation and seismicity at the EGS site. Therefore, to avoid such misapprehensions, we must develop a local velocity model that represents the study area.

We developed a 1-D local velocity model that represents the EGS site based on geological strata observed from drilling cores of the stimulated wells, as well as seismological observations based on active and passive sources (Figure 2). In the layered model, local anisotropy, which can perturb the stress field and affect fault instability (Magnenet et al., 2017), was not introduced. Lithologic variations in core samples provide us with information on seismic discontinuities between layers where there are abrupt changes in the stiffness or
elastic moduli of the rock (Lee et al., 2015). The $P$ and $S$ wave velocities of each layer were determined from either seismic sources with assumed locations or well-logging data (Figures 1b and 2a). For depths below 4.5 km, with no geological data, we adopted the regional model representing the Gyeongsang basin reported in Kim et al. (2011). The constructed 1-D velocity model for the Pohang EGS site was composed of seven layers and a half-space. The final seismic velocity model and procedures to construct the model are summarized in Table 1 (see Text S1, available in the supporting information).

Figure 2. Local 1-D velocity model developed for two different depth ranges in this study. Blue and red lines represent the $P$ and $S$ wave velocity models, respectively. Vertical distributions of the borehole stations along the PX-2 well are illustrated in the right panel of (a). The PX-2 borehole chain stations are co-located with three VSP stations at depths of 1,360, 1,370, and 1,380 m but are operated for different stimulation periods from the VSP stations. The open section of PX-2 is colored in dark red. A simple geological model modified from Lee et al. (2015) is illustrated in the right panel of (a): S-MS (semi-consolidated mudstone), T (tuff), SS (sandstone), ST (siltstone), MS (mudstone), RL (rhylolite), AT (andesitic tuff), CT (crystal tuff), and GD (Granodiorite). The dashed line in (b) represents the lower depth limit of (a). Well-logging data used to determine the $P$ and $S$ wave velocities of the fifth layer are overlapped to (a).

Figure 3. Seismicity at the Pohang EGS site and daily cumulative injected volume. The amount of heavy mud loss from PX-2 at a depth of 3,800 m in 2015 is included in the total net injection volume. The beginning of heavy mud loss and the five stimulation periods are illustrated at the top of the graph. The events without magnitude measurements are denoted by circle symbols. The events whose magnitudes are estimated from either assumed or known locations are denoted by star symbols.
The topmost layer of our model had slow P and S wave velocities and a high Vp/Vs ratio compared with the other layers. The obtained seismic velocities and thickness of the shallowest layer were matched with semi-consolidated sedimentary rocks deposited from 200–400 m at the surface of sedimentary layers with volcanic intrusions. This was analyzed with a deep borehole drilling project implemented near the Pohang EGS site (Lee, 2003; Lee et al., 2015).

2.4. Earthquake Detection

Before analyzing the source parameters of the recent earthquakes, we detected the earthquakes that occurred near the EGS site before the occurrence of the mainshock using a template matching method (Kato et al., 2016; Shelly et al., 2007; Zhang & Wen, 2015). This method uses templates based on the observed waveforms to find similar waveforms using a waveform cross-correlation between the template and continuous waveforms recorded at an identical station. The advantage of this method is that it can be applied even when the signal-to-noise ratio (SNR) is not high. We selected the waveforms for 39 events as the templates (Table S2), including five immediate foreshocks that occurred before the mainshock, which we have clearly identified as earthquakes. As large events can easily be identified due to their high SNR, the main purpose of the template matching method is to detect small events. Therefore, we selected well-recorded waveforms as templates from small events.

To detect events that occurred before and during the stimulation, we used the continuous waveform recorded at PHA2 (or PHA; the official name of the seismic station had been changed but hereafter, for convenience, we only refer to it as “PHA2”) from 1 January 2009 until the mainshock occurrence. PHA2, which is operated by the KMA, is the closest station to PX-2 among the permanent stations operated over an extended period. PHA2 is located approximately 10 km north of PX-2. We used the short-period records with a sampling rate of 100 Hz. For comparison, HAK, which is operated by the KIGAM, is the second closest station to PX-2 but the distance is approximately 23 km, which is more than twice as far as PHA2. Therefore, to achieve consistent results, we only used data from PHA2 for earthquake detection.

The time window for the template waveform was defined as 4 s, from 1 s before the arrival time of the S wave. The bandpass filter between 5 and 20 Hz was simultaneously applied to the template and continuous waveforms before the waveform cross-correlation. On a daily basis, the correlograms calculated for the three seismogram components were averaged to increase the SNR and median absolute deviation (MAD) of the averaged measured correlogram. The criterion for the detection was set to 14 times the MAD, and we identified 3,547 event candidates. Candidate waveforms were visually inspected to remove false positives and overlapping events and to screen out seismicity near the EGS site. In this step, waveforms recorded at the other stations were also used to increase the confidence in the detection and categorization. The matched filter did not detect the mainshock because a large magnitude difference between the mainshock and template events break the waveform similarity and render template matching impossible. Finally, we created an initial catalog of 520 earthquakes after adding the mainshock. Figure 3 illustrates the temporal distribution of the events that were spatially correlated to the Pohang EGS site.
3. Seismic Source Parameter Analysis

3.1. Hypocenter Determination

Three steps were taken to accurately determine the hypocenters. The first and second steps were the single and double-difference methods, respectively, which are widely used in seismology to determine hypocenters. For convenience, we refer to the hypocenters determined by the first and second steps as the initial and relative locations, respectively. In the last step, we shifted the relative locations by considering the difference between the relative location and hypocenter as determined by an individual method for the one specific event that occurred on 13 August 2017, at 21:42:37 UTC (hereafter, referred to as the key event). We defined the shifted relative locations as the final locations.

3.1.1. Initial Locations

Classic absolute location methods are based on the minimization of time residuals between the theoretical and observed arrival times of body waves (i.e., generally the first $P$ and $S$ onsets) through iterative inversion algorithms. In this study, we used Hypoellipse (Lahr, 1999) to determine the initial locations. We collected all data recorded at operating seismic stations based on the origin times of 520 events and manually picked the $P$ and $S$ wave arrival times. We calculated the initial locations when more than four $P$ and $S$ wave arrival times were available. The stations located at depths below the bottom of the top layer in the velocity model (e.g., VSP stations, PX-2 borehole chain, and BH4 station) were not used as the arrival times could not be precisely adjusted using elevation corrections in Hypoellipse.

Among the earthquakes in the initial catalog, the 253 initial locations were calculated. These events were then divided into three groups in terms of their epicentral distances from PX-2, as well as their focal depths. The first group consisted of 140 events with epicentral distances larger than 5 km and a maximum epicentral distance of 62 km (i.e., red open circles in Figure 1). There were four events in the second group that occurred in the years 2013, 2015, 2016, and 2017 (i.e., blue open circles in Figure 1). The third group had 109 events with epicentral distances and focal depths less than 5 km and shallower than 10 km, respectively. Considering their large epicentral distances and deep focal depths, it is unlikely that events characterized as either the first or second group were directly related to stimulation of the EGS project. As the main motivation of this study is to clarify the relationship between hydraulic fluid injection and earthquake sequences at the EGS site, we focused on the third group and performed additional analyses for 109 events in that group.

3.1.2. Relative Locations

To relocate the relative locations from the initial locations, we used the double-difference method, which finds the hypocenters that minimize the residuals between the theoretical and observed travel time differences of event pairs observed at a common station. We determined the relative locations of 98 events using HypoDD (Waldhauser & Ellsworth, 2000) after excluding 11 events whose travel time measurements were less than eight or whose relative travel time measurements are less than nine, to increase the reliability of the results. The stations located below the first layer were not used again. To maximize the performance of the double-difference method, precise measurement of the travel time differences among the different events recorded at the same station is important. If two events recorded at the same station have similar hypocenters and focal mechanisms, the waveform cross-correlation can be used to precisely measure the differential travel times. We selected a 1-s time window (−0.5 to 0.5 s) from the manually picked arrival times of the $P$ and $S$ waves, increased the sampling rate to 1,000 Hz using cubic spline interpolation and applied bandpass filtering between 2 and 10 Hz before the cross-correlation. If relative travel time measurements using cross-correlation are not possible, the relative travel times can be simply calculated from the arrival times used to determine the initial locations. The damped least square QR algorithm was used for the inversion. The error associated with the relative locations was estimated with two individual methods, which are described as follows. First, the synthetic relative travel time data were reproduced by randomly sampling the residuals of the relative travel times calculated using the least square QR and then inverting them a second time. This process was repeated 200 times, such that the errors were estimated via a statistical analysis. Second, a linear equation was constructed using the residual of the relative travel times obtained from the last iteration in the inversion and the estimated least square error using the singular value decomposition method. The average 2σ errors of the relative locations in the east, north, and upward directions estimated using the two methods were 20, 13, and 25 m and 15, 10, and 19 m, respectively, showing that the difference in the error estimated using the two methods was not significant.
3.1.3. Key Event Hypocenter Determination and Final Locations

Among the relocated events, the key event on 13 August 2017 (MW 1.2) was the only event recorded by both the surface stations and a linear vertical array of 160 m that consisted of 17 three-component geophones (group interval of 10 m) deployed in the PX-2 well at a depth of 1,360 m (see supporting information for more details). The borehole array operated from 26 July 2017 to 23 August 2017, during stimulation operations in the PX-1 well. To process the data recorded with this tool, we must use an ad hoc technique. To locate the key event, we applied a combination of techniques based on array processing and polarization analysis (see supporting information for details). As a first processing step, we reoriented the borehole array using the approach proposed by Grigoli et al. (2012) and Krieger and Grigoli (2015). We then located the events using the principal component analysis technique to estimate the back azimuth and incidence angle (Eisermann et al., 2015; Noda et al., 2012). Finally, we used the PS time to obtain an estimation of the source-receiver distance that, combined with the azimuth and incidence angle, allowed us to obtain the absolute location of the seismic event. We further constrained the depth of the key event using the PS times of both the shallowest and deepest geophones and tube waves generated at the bottom of the cased well (see supporting information for further details). The latitude, longitude, focal depth, and origin time of the event were determined as 36.1117°, 129.3734°, 4.210 km, and 13 August 2017, at 21:42:37, respectively. As similar waveforms were recorded by the vertical array seismometers in the PX-2 wellbore, as well as the use of several independent data to evaluate the hypocenter (i.e., travel time, particle motion, and tube waves), the location of the key event was assumed to be the most reliable among the hypocenters of all events. Therefore, this result was used to determine the final locations. The key event determined from the records at the PX-2 borehole chain was located 3 m east, 430 m north, and 161 m vertically upward of its relative location. We shifted the relative locations of all events by the given values to determine the final locations. Most final epicenters were located NW of the PX-2 and distributed along NW-SE and NE-SW trends (Figure 4). The latitude, longitude, and depth of the mainshock were determined as 36.1061°, 129.3726°, and 4.270 km, respectively. All final focal depths were in the range of 3.5–4.5 km.

3.2. Magnitude Determination

We must consider the size distribution of the earthquakes (i.e., the magnitudes) to evaluate the seismic activity in a target region. Among the earthquakes identified by the template matching analysis, the local magnitude (ML) and MW of the events, which are considered to be near the EGS site (i.e., the hypocenters are within 5 km of PX-2 or the observed PS times at PHA2 are within 2 s), were determined to quantify the seismicity at the EGS site.

The magnitude scaling relationship proposed by Sheen et al. (2018) was applied to the vertical component to determine the ML. However, even the permanent seismic stations in South Korea, many of which are conventionally used to measure ML, show large deviations in terms of the individual station magnitudes of up to approximately 0.54 magnitude units in the horizontal component and approximately 0.29 magnitude units in the vertical component (Sheen et al., 2018). Therefore, we obtained the correction terms for the seismic stations before estimating the ML (see Text S3 for a detailed description of station corrections). With the station corrections, we estimated the local magnitudes of 40 events based on more than three observations, ranging from 0.15–5.33, with an average standard deviation of 0.14 magnitude units. Note that the ML of the Pohang earthquake was estimated at 5.33 ± 0.14 from 11 permanent stations within a distance of approximately 50 km from the epicenter. This is similar to 5.34 ± 0.18, which was independently estimated from 77 broadband seismic records from permanent seismic stations in South Korea.

The ML values of the 72 locatable events near the EGS site were not able to be determined from the synthesized Wood-Anderson displacements as single or double integration into the displacement increased the low-frequency noise or coda waves of the previous events. Instead, they were determined by the peak amplitudes of the S wave on the velocigrams and accelerograms of PHA2, which was the only station that had all ML values. For the events at unknown locations (due to low SNR on the seismograms), the ML values were determined using the empirical relationships between the ML values and peak amplitudes of the S wave for the 40 events at PHA2. A total of 96 ML values determined based on the peak amplitudes at PHA2 ranged from −1.30 to 2.03. The uncertainty of the resulting magnitude determinations due to location error and attenuation correction was likely on the order of ±0.2 magnitude units.
The moment magnitudes of 48 events were determined with a rock density of 2.7 g/cm³ and a P wave velocity of 6.0 km/s in the time and frequency domains. The P displacement pulse from the accelerometer at PHA2 was used to compute the seismic moment (Aki & Richards, 1980) in the time domain based on the methods reported in Tsuboi et al. (1995) and Prejean and Ellsworth (2001). We measured the moment magnitudes of 46 events in the time domain, which ranged from 0.58 to 2.72. However, it was not possible to compute the seismic moment of the largest two events, that is, ML 3.27 on 15 April 2017, and the Pohang earthquake, in the time domain due to the complexity and long duration of the P pulses. Therefore, the moment magnitudes of these events were estimated in the frequency domain.

The moment magnitude of the ML 3.27 event was measured with the spectral ratios of the P wave. We selected the M_L 3.27 event that occurred on 15 April 2017, which is the only event that had a mismatched origin time and location compared with the stimulation history, is denoted by a white circle. The location uncertainties of the events related to the PX-1 and PX-2 injections are shown in blue and red. The trajectories of the PX-1 and PX-2 wells are shown as two gray curves with two open sections in blue and red. The origin of the coordinate is set to the location of PX-2. The divisions of the geologic formations are from Song (2015).

### 3.3. Earthquake Size Distribution

Several previous studies on induced seismicity have discussed the frequency-magnitude distribution or the Gutenberg-Richter law (G-R law) as it represents the statistical characteristics of seismic activities in a specific region over a specific time period (Kraft et al., 2016; Zang et al., 2014). This distribution or law can be expressed with the following equation:
\[ \log_{10} N(\geq M) = a - bM, \]  

where \( N(\geq M) \) is the number of earthquakes that have a magnitude equal to or greater than \( M \) and the \( a \) and \( b \) values are the seismicity and frequency of large earthquakes compared with smaller earthquakes for a specific region and time, respectively (Gutenberg & Richter, 1954). We evaluated the \( a \) and \( b \) values from the estimated magnitude of the earthquakes near the EGS site based on the maximum likelihood method proposed by Aki (1965), with a magnitude bin size of 0.1. Before fitting the observed frequency size distribution to equation (1), we estimated the minimum magnitude of completeness (Mc) that holds equation (1) by maximizing the goodness-of-fit function proposed in Wiemer and Wyss (2000). The standard error of the \( b \) value was calculated using an equation proposed by Shi and Bolt (1982).

We investigated the behaviors of the frequency-magnitude distributions for the entire data set (A1) before the \( M_W 5.5 \) mainshock. Afterward, two subsets of events in the PX-1 (A2) and PX-2 stimulations periods (A3) were tested separately to compare the statistical parameters of the G-R law. Figure 5b shows the frequency-size distribution and fitting lines of equation (1) for the three different groups. The obtained G-R distributions for A1, A2, and A3 had nearly identical \( b \) values at 0.66 ± 0.08, 0.61 ± 0.09, and 0.65 ± 0.09, and \( a \) values of 2.08, 1.33, and 1.96, which are proportional to the number of events. The obtained \( b \) values were generally lower than commonly observed values in the Gyeongsang basin (Hong et al., 2016). The seismogenic indices \( (\Sigma = a - \log_{10} V; \text{Shapiro et al., 2007, 2018; Shapiro et al., 2010}) \) extracted from the obtained \( a \) values and cumulative injection volumes for A1, A2, and A3 were estimated at -1.69, -1.89, and -1.65, respectively.

### 3.4. Focal Mechanism Solutions

A distribution of the final locations can help delineate the geometry of the fault plane. However, we must have the focal mechanism solutions to understand the fault type of each event. We constrained the focal mechanism solutions from the \( P \) wave polarities (Lay & Wallace, 1995). We measured the \( P \) wave polarities from the vertical component seismograms. If the polarity was not clear due to low SNR, we determined the polarity using the polarities of the larger events that had similar waveforms. For certain borehole records, we speculate that the vertical sensor may or may not have been deployed correctly. In this case, we analyzed the waveforms of the event that occurred 37 km SW of PX-2, whose focal mechanism was well constrained by seismograms recorded at local seismographs near the epicenter. We then checked the polarities for suspicious borehole stations by comparing the observed and theoretical \( P \) wave polarities, performing
corrections when necessary. To determine the focal mechanism, we applied the HASH (Hardebeck & Shearer, 2002) software with all available records including those at the borehole stations beneath the first layer of our velocity model. We calculated the candidates for focal mechanism solutions by allowing one data point error for the polarities to determine the final focal mechanism solution for each event by averaging all possible candidates. The quality of the focal mechanism solutions was evaluated by following the criterion provided in Hardebeck and Shearer (2002). We determined 28 and 25 focal mechanisms with quality A and quality B, respectively. The strike, dip, and rake of the mainshock were determined to be 214°, 51°, and 128°, respectively.

We classified the fault types of the focal mechanism solutions based on the classification proposed Zoback (1992), identifying 14 strike-slip events, 22 thrust events, and 15 predominantly thrust events with strike-slip components. We did not find any normal faulting events. Only one event had a predominantly normal component with a strike-slip component. These results indicate that the strike-slip and thrust mechanisms are dominant in the study area (Figures 6a and 6b). The trend and plunge of the average P axis direction of the 53 events were 96° and 4°, which are consistent with the direction of maximum horizontal stress that occurs under the Korean Peninsula (Soh et al., 2018).

Previous studies have succinctly shown that the stress field in the region controls the faulting mechanism. Therefore, we inverted the focal mechanism solutions for a stress field in our study area using Michael method (Martínez-Garzón et al., 2014; Michael, 1984). To avoid the fault plane ambiguity problem, we adopted the concept of fault instability proposed by Vavryčuk (2014). The results of the stress inversion show that the trend and plunge of the maximum principal stress (σ1) were 276° and 7°, respectively, which are near the average P axis direction. In addition, the trends/plunges of σ2 and σ3 were determined to 18°/58° and 182°/31°, respectively.

3.5. Hierarchical Clustering Analysis

We applied two different hierarchical clustering analyses (Defays, 1977; Sibson, 1973) to understand the spatial and temporal characteristics of seismicity. The difference between the two methods lies in their definitions of a metric distance between two events. For the first analysis, we used the Kagan angle (Kagan, 1991) as a metric distance between two events for the 53 earthquakes with focal mechanism solutions. The Kagan angle indicates a minimum 3-D rotation angle required for one double-couple earthquake source mechanism and another one. The distance between two clusters is defined as the mean value of the Kagan angle for all possible pairs between the events of each cluster, such that five clusters were classified (i.e., A, B, C, D, and E) based on a criterion of 45° (Figures 7 and S7). For the second analysis, we used the waveform similarity between two events, which is useful when classifying events because additional information, such as the location or focal mechanism solutions, are not required to perform a hierarchical analysis (D’Alessandro et al., 2013; Son et al., 2018). We used the waveforms recorded at PHA2 again as it is the closest permanent station to PX-2. To simultaneously consider the P and S waveforms, we cut the S windows from 0.5 s before the P wave arrival times. Waveform similarity between two seismograms was estimated as the maximum cross-correlation coefficient (MaxCC), allowing a time shift of up to ±0.5 s for potential intrinsic picking errors. A band-pass filtering between 2 and 20 Hz, as well as a cubic spline interpolation with a sampling rate of 1,000 Hz, was applied before measuring the MaxCC. The metric distance between two events was defined as (1 − MaxCC), while the distance between two clusters was defined as the minimum distance between two events that belong to two different clusters. Exposed to a threshold of 0.25, we identified two clusters (i.e., CW-1 and CW-2) consisting of 24 and 80 earthquakes, respectively (Figure S8).

4. Discussion

4.1. Fault Reactivation Revealed by Hypocenter Distribution

To investigate the relationship between the hydraulic stimulations into the two wells (i.e., PX-1 and PX-2) and earthquake sequences near the Pohang EGS project site, we divided the 98 relocated events into two groups based on their origin times (Figures 3 and 8). The first and second groups (hereafter referred to as G1 and G2) consisted of earthquakes during the PX-1 and PX-2 stimulation periods, respectively. As the MW 0.9 event in the mud loss period, mainshock, and immediate foreshocks were adjacent to the preclassified G2 events, we additionally included the events in the G2 group.
The epicenters of G1 and G2 were clearly separated and distributed along the NW-SE and NE-SW trends, respectively, except for the MW 2.0 event on 11 September 2017, that occurred in the last stimulation at PX-2 (Figures 4a and S6). Considering that the experiment was conducted only 1 month prior to the fourth stimulation at PX-1 and that the event was located in the G1 area, we suggest that this event was a delayed seismic response to the fourth stimulation. The hypocenter distribution of G1 and G2 in Figure 4 reveals that G1 events formed a tube-like structure with deepening focal depths in an NW direction, whereas those for G2 appear to have occurred on a fault dipping in an NW direction. We evaluated the similarity of the spatial distribution’s shape for G1 or G2 to a planar structure using planarity: \(1 - \frac{L_3}{L_2}\), where \(L_2\) and \(L_3\) are the immediate and minimum eigenvalues of the covariance matrix for a collocation of \(x\), \(y\), and \(z\), respectively, that is, \(\text{Var}(xyz)\). The results show that the planarity values of the distribution are nonnegative and can be as high as 1. Here one corresponds to a perfect planar structure. The planarities of G1 and G2, with the exclusion of the MW 2.0 event at G2, were 0.44 and 0.71, respectively, which suggests that the G2 events were closer to a planar structure than those of G1. Using a conventional principal component analysis, we obtained

Figure 6. (a, b) Distribution of 53 focal mechanism solutions (a) in map view and (b) in a cross-sectional view along A1-A2. Colors of the beachball diagrams represent the fault types determined according to the classification proposed by Zoback (1992): strike slip (green), thrust (red), strike slip with thrust component (yellow), strike slip with normal component (blue), and unclassified faulting type (gray). The origin of the map coordinate is set to the location of the PX-2 well. (c, d) Distribution of the earthquake focal mechanisms in the PX-1 stimulation period and histogram of their Kagan angles. Distributions of \(P\) and \(T\) axes are denoted by red and white triangles, respectively. (e, f) Identical to (c) and (d) but for the PX-2 stimulation period. In (e), the nodal planes of the mainshock are in yellow. The projection of the best fitting plane estimated from the hypocenter distribution is denoted by a red line. Dashed lines in (d) and (f) represent the average value of the Kagan angles (Kagan, 1991).
the strike and dip of the approximate fault plane for G2 as N214° and 43° NW, respectively, which are similar to the focal mechanism of the Mw 5.5 mainshock (i.e., N214°/51°NW; see Figure 6e), the fault plane estimated from the aftershock distribution (Kim et al., 2018), and the focal mechanism obtained by a moment-tensor inversion (Grigoli et al., 2018). As the G2 fault plane shares the same strike and dip of the Pohang earthquake, as well as the immediate foreshocks, the mainshock appears to have initiated on a preexisting fault plane, with the occurrence of earthquakes induced by fluid injections in PX-2.

Topographical variations perpendicular to the epicenter distribution of G2 have been observed. The overall altitude of the western region is lower than that of the eastern region (Figure 4a). Unconsolidated alluvial sediments are prevalent on the western side of G2, while the sedimentary layer of sandstone and mudstone are mapped on the eastern side (Song, 2015). Therefore, the fault, delineated by the spatial distribution of G2, may represent the reactivation of a preexisting normal fault in the basin structure that developed in an extensional tectonic setting. Including the Gokgang fault, NE striking normal faults that dip in the SE/NW direction can be found throughout the Pohang basin, such that stress inversion from extension to compression resulted in the reactivation of the normal faults (Yoon et al., 2014).

4.2. Breakage at the Crossing Point of the Fault and PX-2

When the best fitting fault plane, which was obtained from the spatial distribution of G2, was extended toward the PX-2 well, it crossed the well at a depth of approximately 3,800 m (Figure 8). This depth nearly coincides with a depth of 3,783 m, which has been inferred as the breakage of the borehole. The image logging test was performed nine months after the Mw 5.5 mainshock but an image logging device was unable to descend any further. (Geological Society of Korea, 2019). Therefore, it is reasonable to infer that the G2 events, including the mainshock, occurred on the same fault plane with similar focal mechanisms and rupture propagation of the mainshock broke the casing of the PX-2 well.

Nearly 2 months before the first stimulation of PX-2, 652 m3 of heavy mud (1.6 g·cm$^{-3}$) was lost during the well drilling process (Geological Society of Korea, 2019). Most of the loss occurred at a depth range of 3.8–4.0 km, which matches with the crossing point between the best fitting fault plane for G2 and PX-2 well trajectory. The largest event (MW 0.9) in the prestimulation period was also on the approximate fault plane. Therefore, it is plausible to assume that the mud was lost along the fault fracture zone, perturbing the stress field due to the inflow of mud into the hydraulic medium and generating microearthquakes on the fault in the prestimulation period. Induced seismicity during injection well drilling has also been reported in the Schlattingen geothermal project (Kraft et al., 2016) and Hellisheiði Power Plant (Ágústsson et al., 2015). Reactivation of the fault plane for G2 can be attributed to the mud loss during drilling, along with the three stimulations in PX-2.

4.3. Seismic Responses of the PX-1 Hydraulic Experiments

Based on the spatial distribution and diversity of the focal mechanisms, the G1 events are considered to have occurred on optimally oriented faults as a response to the local stress field unlike the seismicity of G2 with similar fault types (Figures 6c–6f and 7). Therefore, in
contrast to the seismicity of PX-2, which delineates a preexisting fault plane that deviates from PX-2 (e.g., Lei et al., 2017; Schultz et al., 2017; Yeck et al., 2016), the earthquakes related to the PX-1 hydraulic stimulations should be clustered near the open section of PX-1 as a response to the elevated pore pressure in the proximity of the PX-1 injection interval, possibly due to pore pressure diffusion (Ogvardi & Horton, 2016) or poroelastic responses as suggested by Cornet (2016). However, the overall hypocenters of G1 deviate from PX-1 to the NE by approximately 200 m (Figure 4a).

As the final locations were determined by shifting the relocated events to the absolute location of the key event, the accuracy of the final hypocenters was entirely dependent on the estimated absolute location of the key event. To locate the key event, we used the particle motions-recorded PX-2 borehole chain, where the seismometers are arranged nearly vertically along PX-2. High-resolution phase arrivals measured from 1,000 Hz seismograms with high SNR resolved the focal depth of the event with an uncertainty of 0.1 km. These estimates were cross-validated by two independent estimations from differential PS times at the two farthest sensors and from tube waves generated in the encased wellbores (see Text S2 in the supporting information). The azimuth of the key event was also accurately measured from the stacked waveforms of seismograms with high coherency. However, the absolute orientations of the wellbore sensors estimated by the teleseismic events may have intrinsic uncertainties due to the lateral fractions of ray paths (Laske, 1995). Inaccurate estimation of the station orientation can result in horizontal uncertainties for the determined hypocenters regardless of the accuracy of the estimated azimuth for the key event.

Assuming that the perturbed pore pressure at the open section of PX-1 governed the seismicity in G1, as well as that the optimally oriented fractures ruptured in response to the current tectonic stress, earthquakes that were expected during the stimulations at PX-1 were clustered at the open section of PX-1. If we assume that the orientation error of the borehole sensor is 20° and rotates the absolute location of the key event anticlockwise by 20°, the events in G1 may be shifted to near the open hole interval of PX-1. The shifted final hypocenters do not affect the main observed features discussed above. The best fitting fault plane for G2 still crosses the PX-2 well at a slightly shallower depth of approximately 3.77 km. The crossing point and the projection of the PX-2 open section onto the best fitting fault plane of G2 are displaced to the NE, such that the migration pattern clearly appears to have initiated near the projection of the open hole interval at PX-2.

4.4. Spatiotemporal Variations in Seismicity for Stimulation Periods

We investigated the detailed spatial and temporal distribution of earthquakes that occurred during or after five stimulations at PX-1 and PX-2 by subdividing G1 and G2 into seven subgroups based on the stimulation periods. G1 was divided into G1-1 and G1-2 for the second and fourth stimulations performed at PX-1, respectively, while G2 was divided into G2-1, G2-2, and G2-3 for the first, third, and fifth stimulations at PX-2, respectively (Figure 4). The prestimulation earthquakes that occurred due to heavy mud loss in PX-2 and the Mw 5.5 mainshock sequences were set to G2-0 and G2-M, respectively.

We observed no spatial variation between G1-1 and G1-2 despite a lack of seismicity at G1-1. Compared with the hypocenter distribution demonstrated in Hofmann et al. (2019), who analyzed 52 events in the fourth stimulation (including the key event) using the PX-2 borehole geophone array, the relative locations of the 52 events to the key event also overlapped with our results for G1-1. In contrast, G-2 seismicity appears to have sequentially migrated SW as the stimulation experiments were iterated at PX-2 (Figure 9). Moreover, the only relocated event of Mw 0.9 in the prestimulation period (G2-0) is situated on the NE edge of the events in G2-1. Compared with the events in G2-1, the events of G2-2 slightly expanded to the SW by 100
m and in the downdip direction by 200 m. One day after the end of the third stimulation, the MW 3.2 event occurred at the SW periphery in the G2-2 seismicity. Its aftershocks, which are thought to have been triggered by stress transfer or postseismic stress relaxation, are distributed over a wider range of approximately 1 km to both the SW and NE. Barring the aftershocks of the MW 3.2 event, the G2-3 events apparently migrated SW and in downdip directions by approximately 100 m compared with the G2-2 events. Immediate foreshocks and the mainshock, which occurred 2 months after the end of the last stimulation at PX-2, are located 200 m to the SW compared with the events in G2-3.

Although a matched-template approach was used to maintain a constant detection power, the observed seismic activity with respect to the injection rates or wellhead pressure varied for each stimulation period (Figures S2–S6). For PX-1, the first detected event of ML −0.3 at the second stimulation occurred approximately 3 days after the onset of its operation, whereas, for the fourth stimulation, it took more than twice as long, that is, approximately 7 days (Figures S3 and S5). In addition, the number of events in G1-1 outnumbered that in G1-2 by over an order of magnitude. The seismic quiescence at the beginning of the later stimulation may represent the Kaiser effect, which indicates a lack of seismicity until the current stress level exceeds the previous loading (Baisch et al., 2002; Hofmann et al., 2019; Kwiatek et al., 2019). In the second stimulation, 3,907 m³ of water was injected with a maximum wellhead pressure of 27.7 MPa (Hofmann et al., 2019). In contrast, the fourth stimulation applied a cyclic soft stimulation, which allows alternating phases in the flow rate to control the wellhead pressure and maximum event magnitude, such that 1,756 m³ of water was injected with a maximum wellhead pressure of 25.2 MPa without allowing any shut-in period (Hofmann et al., 2019). A large amount of bleed-off in the fourth stimulation resulted in a decreased net injection volume for PX-1. Therefore, scarce seismicity in the fourth stimulation may have resulted from the low injection rate, smaller injection volume, and the use of a controlled stimulation procedure (Figures 3 and S5).

In the three stimulations at PX-2, differences were also observed in the time delay between the first observed event and onset of each stimulation. The first detected ML 0.0 event in the first stimulation...
occurred 3 days after the onset of injection, whereas there was a 2-week gap between the seismic activity and onset of periodic injection in the third and fifth stimulations. As the injected fluid perturbs local stress fields across a large area over time due to pore pressure diffusion and the poroelastic effect (e.g., Goebel et al., 2017), the time delay in seismicity initiation for the two later stimulations may account for the expansion of the area with seismic activities on the best fitting fault plane of PX-2. The lack of seismicity at the previously activated region may indicate that the loaded stress on the area was far less than the critical stress due to stress release through previous seismic activity. The observed unilateral migration pattern, which has also been observed for the large-scale wastewater disposal operation in Oklahoma and southern Kansas (Schoenball & Ellsworth, 2017), hydraulic fracking in Canada (Woo et al., 2017), and at the Yamagata-Fukushima border (Yoshida & Hasegawa, 2018), may have resulted from heterogeneities in the hydraulic parameters, frictional properties, and stress state on the PX-2 fault. Further migration of G2-M to the SW also matches with a 2-month gap of seismicity from the onset of the last stimulation. This phenomenon shows that redistribution of the stress field over large areas takes time.

All earthquakes with \( M_L > 1 \) occurred during periods characterized by bleed-off or shut-in rather than fluid injection. While the maximum magnitude of the earthquakes that occurred during fluid injection was only 0.905, seven events with \( M_L > 2 \) (one event in G1-1, three events in G2-2, and three events in G2-M) occurred immediately after each stimulation period or in the foreshock sequences. This is not an unusual occurrence as large-magnitude earthquakes have often been observed in other EGS sites such as Cooper Basin, Soultz, and Basel during shut-in periods or after bleeding-off (Häring et al., 2008; Majer et al., 2007; Mukuhira et al., 2017).

4.5. Deviation Between Physics-Derived Scaling Equations for Maximum Seismic Moment and Net Injection Volume for the \( M_W \) 5.5 Mainshock

Based on the derived equation reported in McGarr (2014) for the relationship between the maximum seismic moment \( (M_{\text{max}}) \) and volume \( (V) \) of the injected fluid, inducing the \( M_W \) 5.5 Pohang earthquake required at least a 500-fold-greater fluid injection volume. Despite the inconsistency between the observed and expected maximum seismic moments, the \( M_W \) 5.5 event is suspected to have had an anthropogenic origin due to its spatiotemporal correlation with the injection workflow (Grigoli et al., 2018; Kim et al., 2018). To closely analyze the relationship between \( M_{\text{max}} \) and \( V \) at the Pohang EGS site, we compared the seismic moment of the Pohang earthquake and largest prior events with the net injected volumes at their origin times for PX-1, PX-2, and both wells (Figure 10). We observed two main characteristics in the correlation between the two parameters. First, most of the detected earthquakes had moment magnitudes less than the expected \( M_{\text{max}} \) when using the \( V \) except for the \( M_W \) 5.5 mainshock. Based on the three scaling equations in Figure 10, the \( M_W \) 3.2 event, which was the largest event before the occurrence of the \( M_W \) 5.5 event, was already near the upper limit of the maximum magnitude responsible for the net inflow at the moment. Based on Galis et al. (2017), the \( M_W \) 5.5 event can be classified as a runaway rupture or self-sustained rupture, where the majority of the seismic energy was tectonically released along critically stressed faults regardless of the initial rupture process. This is consistent with the modeled stress state and static friction of the fracture obtained from focal mechanism inversions, drilled boreholes, and measured wellhead pressures, which all indicate that the G2 best fitting plane may have been (nearly)
The hydraulic energy supplied by fluid injection into the two stimulation wells is inconsistent with the observed seismic moments. We calculated the hydraulic energy of the injected fluid using $E_h = \int P(t)Q(t)\,dt$, where $P(t)$ and $Q(t)$ are the wellhead pressure and injection flow rate, respectively. The seismic energy ($E_s = \Delta \sigma M_s/2\mu$; Lay & Wallace, 1995) released by the $M_w$ 5.5 event corresponds to $3.7 \times 10^{13}$ J based on a stress drop ($\Delta \sigma$) of 5.6 MPa and shear modulus of 30 GPa ($\mu$), which is approximately 50 times larger than the cumulative hydraulic energy of $6.9 \times 10^{11}$ J. In contrast, the total seismic released energy from the mud loss period to the onset of the $M_w$ 3.2 event was $1.2 \times 10^{10}$ J, corresponding to 4% of the hydraulic energy of $4.8 \times 10^{11}$ J. De Barros et al. (2019) argued that the hydraulic energy released during the PX-2 stimulation was 2,370 times larger than the cumulative energy due to a large aseismic deformation related to the fluid injection. Therefore, the energy equilibrium for the $M_w$ 5.5 mainshock and possible large aseismic deformations (Guglielmi et al., 2015; Cornet et al., 2016) require other types of input besides the hydraulic energy, such that it is consistent with the large discrepancy between the observed and expected maximum moment magnitudes.

5. Conclusions

In this study, we were able to prove that there is a clear causal link between the origin of the 2017 $M_w$ 5.5 Pohang earthquake and injected fluids injected as part of an EGS project through spatiotemporal relations. For this, we constructed a local 1-D velocity structure and analyzed the characteristics of the earthquakes that occurred in the vicinity of the EGS site until the occurrence of the Pohang earthquake. We did not detect any earthquakes within 5 km of PX-2 based on our matched filter analysis until an $M_L$ −0.1 event that occurred on 1 November 2015, when a large volume of heavy mud was lost from PX-2. The relocated seismicity indicates that the earthquakes during the PX-1 stimulation periods represented an ellipsoidal distribution to the NW-SE, whereas seismicity during the PX-2 stimulations had the shape of an NE-SW fault plane dipping to the SE. Except for the aftershocks associated with the second largest $M_w$ 3.2 event, earthquakes on the fault plane appear to have migrated to the SW during an interstimulation period. A series of hierarchical clustering analyses suggested that they share similar fault parameters. The intersection depth between the fault plane and PX-2 was approximately 3,800 m, which corresponds to the intervals of heavy mud loss, as well as the breakage of the borehole by an image-logging device. However, unlike the earthquakes in the stimulation periods, the seismic moment of the Pohang earthquake far exceeded the possible ranges estimated by both the $M_{max}, V$ scaling laws and energy equilibrium. These results provide strong evidence that the $M_w$ 5.5 Pohang earthquake initiated on the fault zone that was reactivated by fluid injection, representing a self-sustained rupture process that released a large amount of energy via tectonic loading rather than being a directly induced earthquake via fluid injection. The Pohang earthquake case provides the lesson that large earthquakes on well critically stressed faults can be triggered even if critically stressed (Ellsworth et al., 2019; Geological Society of Korea, 2019). Second, the size of the $M_{max}$ steeply increased over time, far beyond the expected $M_{max}$ value (Figure 10). This is contrary to the temporal evolution of $M_{max}$ and $V$ observed in a geothermal stimulation in Finland (Kwiatek et al., 2019), which followed several scaling models (see Figure 5 in Kwiatek et al., 2019). Van der Elst et al. (2016) revisited the equation proposed by McGarr (2014) using statistical approaches, finding that the seismic moment scales up with $V^{3/2b}$ rather than $V$ alone. The derived equation from Van der Elst et al. (2016) can partially explain the dramatic change in $M_{max}$ that increased due to $V$ when our estimated $b$ value of 0.66 was used instead of the commonly assumed $b$ value of 1. However, the seismic moment of the Pohang earthquake was still larger than the expected $M_{max}$ by over 2 orders of magnitude.


