Spatial Pattern of the Seismicity Induced by Geothermal Operations at the Geysers (California) Inferred by Unsupervised Machine Learning

Mauro Palo[ID], Emanuele Ogliari[ID], *Member, IEEE*, and Maciej Sakwa[ID], *Graduate Student Member, IEEE*

Abstract— We analyzed the earthquake density of the Geysers geothermal field (California) as a function of time and space over a decade. We grouped parts of the volume of the geothermal area sharing similar earthquake rates over time; in this way, we found three concentric spatial domains centered on the principal exploitation area and labeled as A, B, and C, moving from the inner- to the outermost domain and characterized by peculiar time-history of the earthquake rates and different stress conditions. The earthquake density decreases moving from domains A to C, and different slopes of the earthquake frequency-magnitude distribution appear for the domains A–B and domain C. Stress field propagates via a diffusive mechanism up to about 3.5 km from the center of the geothermal area outwards, and a mean hydraulic diffusivity of about 0.05 m²/s is estimated; at larger distances, a poroelastic stress transfer dominates. Our approach can identify spatiotemporal patterns of physical mechanisms driving induced seismicity and can, in principle, be extended to other settings of man-induced earthquakes. Moreover, it potentially allows a differentiated assessment of the seismic risk within each domain, as well as the identification of domains with no or minimal induced seismicity.

Index Terms— Clustering, Coulomb stress, induced earthquakes, pore pressure, the Geysers geothermal field.

I. INTRODUCTION

GEOTHERMAL energy exploits the heat from inner substrates of the Earth's crust by pumping water from the surface. This facilitates the heat exchange between the hotter substrates and the pumped water to generate high-pressure steam which is conveyed through pipelines up to the power plant. Here, the steam is transformed into energy in the form of heat or electricity [1]. This simplified principle of operation covers both shallow (usually less than 300 m) and deep geothermal (3–5 km) technologies [1] producing electrical power, heat, cold water, and hot water [2].

Different from other traditional renewable energy sources [(RESs)—mainly wind or photovoltaic], geothermal energy is less affected by intermittent power generation issues mainly related to intrinsic unpredictability due to weather behavior [3], [4]. In addition, existing geothermal heat and power technologies could deliver substantial cost-efficient base-load energy [5], leading to long-term decarbonization [6]. For these reasons, geothermal energy could play a crucial role within the set of greenhouse gas mitigation policies in Europe [7] and it has grown worldwide, in terms of installed capacity, over the years [3], [8]. Geothermal energy in the past decade has become a significant source of electricity and heat with prosperous development in countries such as Italy [1], Iceland [9], and the USA [10]. Its future potential exploitation has also been evaluated with different perspectives in other countries such as China [11], [12], [13], Bangladesh [14], [15], and Indonesia [16] given their location in highly seismically active areas (such as the Ring of Fire in the Pacific Ocean and the Tibetan Plateau) and additional funding from the UN for clean growth of developing countries [17].

However, some limitations to this technology should be considered: appropriate locations require a deep geological assessment to better evaluate the economic potential of sites for their extensive exploitation with new installations, as reported in [18], which serves as a useful working and decision-making basis for stakeholders. Besides, nowadays, it is difficult to fully understand the legal system regarding geothermal energy in a given European country [19] without some acquaintance with the overarching EU framework [20]. Furthermore, geothermal energy plants are classified as "renewable" only if the heat extraction rate does not exceed the reservoir replenishment rate [21], [22]. Exploitation through wells, sometimes using downhole pumps in the case of non-electrical uses, leads to the extraction of very large quantities of fluid and consequently to a reduction or depletion of the geothermal resource in place [1]. Finally, seismicity produced by human activities (i.e., induced seismicity) [23], [24] has been widely reported over the past years, not limited to geothermal field exploitation [25], causing human discomfort and temporary stops to the operation of power plants.

Geoscience data acquisition has played a key role in recent academic research. Therefore, more sophisticated Earth observation systems, such as satellite imaging, have become more and more popular in the past decades [26] resulting in higher availability and spread of open datasets [27], [28]. In this regard, earthquake detection and recording in geothermal areas represent a standard tool for risk monitoring [29], [30], [31]. In particular, a comprehensive database of earthquakes has

been collected from data recorded by the seismic geophone network at the enhanced geothermal system (EGS) of The Geysers, USA.

The Geysers geothermal power plant is currently the largest single geothermal power installation in the world, and California, where high-temperature resources at the Geysers, the Salton Sea, and Coso are located, has an installed nameplate capacity of 2,627 MW of geothermal power, which is 72% of the total U.S. geothermal power production [32]. The power plant generates electricity from 22 geothermal power plants, which are spread across 45 square miles of the geothermal field [33]. Recent studies have utilized a range of remote-sensing techniques to better understand the subsurface structure and dynamics of the EGS [34], while other studies include satellite-based synthetic aperture radar (SAR) data [35], satellite-based infrared thermal images [36], and seismic tomography [37].

In particular, studies on the relation between microearthquakes (MEQs) and steam production have been conducted since the mid-1970s [38]. In the past, The Geysers was characterized by a low seismic activity. Since the beginning of the utilization of the hot water field, a significant increase in the frequency of micro- to medium-magnitude seismic events [39] has been detected. Through the years, The Geysers suffered a long-term reduction of pressure and a resulting decrease in steam production, which was counteracted by the installation of water injection wells that caused an increase in seismic activity [40]. The correlation between the injected amount of water over the years and MEQs has been largely studied [41], [42], but the causal link between the different exogenous sources and earthquake nucleation is not completely explained yet. The high seismicity rate at The Geysers is being constantly monitored by a robust local high-quality (and high-resolution) seismic monitoring system [43].

In this article, we propose, by exploiting the available historical earthquake dataset at The Geysers, an unsupervised machine-learning (ML)-based approach that employs temporal data structures as indicators of spatial behavior and the response of the geothermal field. This approach should be considered innovative as it is purely data-driven and relies on a reduced number of parameters. The proposed method is capable of detecting patterns of the seismicity induced by the different underlying stress fields, providing a framework to model the relation between fluid injection and earthquake nucleation and to spatially constrain the effects of energy production in terms of seismic risk.

The article is structured as follows. The case study and the dataset are presented in Section II, the adopted methodology is proposed in Sections III and IV, while the results and further analysis are described in Sections V–VII together with an extended discussion in Section VIII. Finally, Section IX presents the main conclusions and the future perspectives of this work.

II. CASE STUDY

The Geysers EGS is characterized by a complex network of faults and fractures that facilitates the circulation of hot fluids through the subsurface and accommodates via fault slipping the stress increase induced by the industrial processes con-
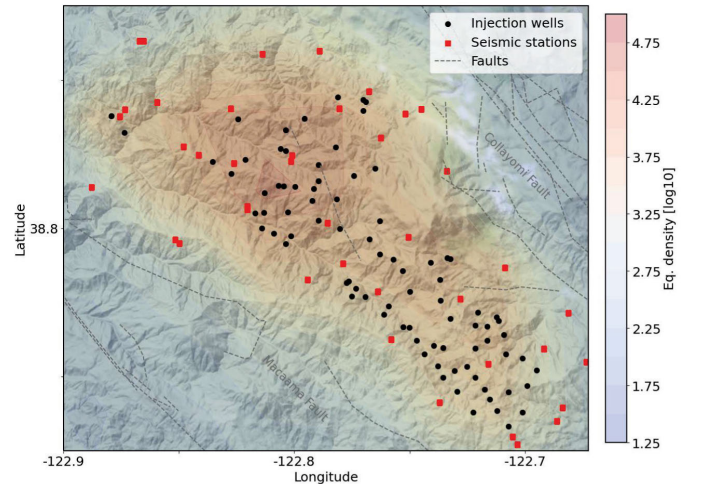

Fig. 1. Map of The Geysers EGS, California, USA. Earthquake density over the analyzed decade is displayed together with the position of the injection wells (stars), seismic stations (squares), and faults (dashed lines).

nected to the energy production. Geologically, the geothermal field is bordered by two major regional faults. One of those, the Maacama Fault, is considered to be active based on an average slip rate of about 10–15 mm/year [44], whereas the other, the Collayomi Fault, is currently considered to be almost inactive [45]. The Geysers reservoir is mostly composed of highly fractured greywacke with low porosity (1%–2%) that is capped by a geological structure formed by a mixture of low-permeability rocks and silica depositions [46]. In Fig. 1, the spatial density of the earthquakes in the area is plotted together with the location of water injection wells and seismic stations.

Our work focuses on the analysis of an earthquake catalog recorded at The Geysers EGS. The dataset has been provided by the Northern California Earthquake Data Center (NCEDC) which is an archive and distribution center for earthquake data recorded in Northern and Central California [43]. This archive is a joint project of the Berkley Seismology Laboratory (BSL) and the US Geological Survey (USGS) and it allows for specific queries on the entire dataset related to the geographical location, depth, event time, and magnitude of the recorded earthquakes.

For this work, we selected earthquakes from the NCEDC occurring between January 2006 and June 2016 (126 months) in the region between 38.7° and 38.9° North, and between −122.95° and −122.65° West, that encloses all injection wells. In total, our dataset is composed of 421 344 earthquakes and includes seismicity up to about 10 km from the wells. The analyzed period includes about two cycles of water injection, during which the largest earthquake rate ever was detected [42], [47]. Moreover, in the selected time interval, the overall water injection level oscillates around a nearly constant value [42], [47] allowing for critical analysis of the detected seismicity in relationship with the main triggering factor during a "stationary" phase.

III. EARTHQUAKE DENSITY TIME HISTORY

To analyze the earthquake data contained in the studied dataset, we perform a few aggregatory steps. First, we divide

The Catalogue

↓

Density Time History

↓

Normalized Distance Matrix

↓

Hierarchical Agglomerative Clustering
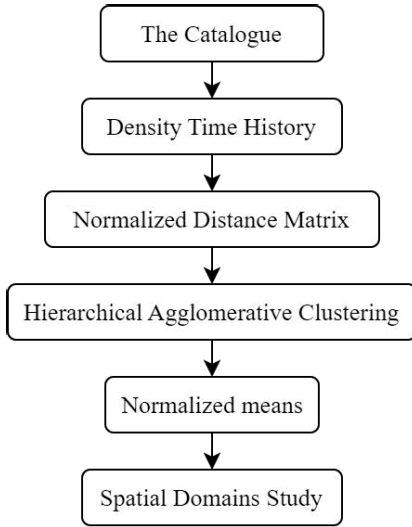
↓

Normalized means

↓

Spatial Domains Study

Fig. 2. Block scheme representing the data manipulation pipeline starting from the original catalog and finishing with the definition of spatial domains.

the control volume into unit volumes with the size of $0.25° \times 0.25° \times 0.25$ km (long-lat-depth) each, creating a grid of individual cells that contain a number of seismic events each. For each month of the observations (that is between January 2006 and June 2016), we sum the number of events occurring in each cell in that time frame, and as a result, a spatial density dataset with a one-month time-step is produced. Then, by extracting the data of a single cell and arranging them in chronological order, we create an earthquake density time history (DTH) of each unit volume. DTH of a cell is thus the monthly rate of the earthquake that nucleates in that cell. The division units of the cell volume were fixed after a sensitivity analysis, in which the effects of the cell size on the dynamic ranges of the DTHs and the spatial sampling of the EGS volume were evaluated (large cells imply better-sampled DTHs but worse spatial resolution and vice versa). The adopted cell size was eventually fixed as the best tradeoff between the two effects. Based on the same principle, the temporal aggregation unit has been set to one month, which is also consistent with the sampling of the available water injection data. This extraction and ordering procedure has been performed for all the cells in the analyzed dataset, resulting in a total of 5254 DTH series. The entire pipeline is summarized by a block scheme in Fig. 2.

In the next step of the analysis, we group by a clustering approach the cells based on the similarity of the DTH. For this aim, we treat each DTH as a 126-D vector, where each dimension is the earthquake rate in a month. Each vector is then normalized by dividing it by its largest component, and the Euclidean distance between each pair of these normalized vectors is calculated as follows:

$$D_2(p, q) = \sqrt{\sum_i (p_i - q_i)^2} \qquad (1)$$

where p and q define a pair of cells and D_2 is the Euclidean distance. We define in this way a "normalized distance matrix" that describes the similarity between all pairs of cells.

IV. CLUSTERING

Clustering algorithms are at the core of ML techniques and data analysis [48]. They have a conceptually simple task of organizing a set of objects into a number of groups, in a way that members of the same group (or cluster) are more similar to each other than to the members of other groups. The main problem regarding this type of analysis is the creation and implementation of a proper similarity metric that can be used to divide the dataset [49]. This metric is often referred to as the "clustering criterion" [50]. Currently, there are countless implementations of clustering algorithms and the definition of this metric can be approached from numerous different perspectives. In this research activity, we choose to proceed with hierarchical agglomerative clustering (HAC), which can provide information on the similarities between objects in the data space at all levels of aggregation, giving a clear structure of the data space [51] and has been successfully applied in different research areas to identify groups of signals emitted from a common source [52], [53], [54], [55], [56], [57].

HAC builds a binary merge tree (dendrogram) starting from the data point located at the ends (leaves) until the root node is reached. The relation between any given subsets of the merged dataset is called "linkage distance" [51]. Based on this criterion, the members are grouped together, with a limit on the maximum allowed distance between the clusters. Depending on the order of operation, an agglomerative (bottom-up) or a divisive (top-down) procedure can be defined. Here, we use the bottom-up technique, so we group at first very similar objects.

HAC is then performed on the normalized distance matrix calculated above. With the metric defined, the linkage distance criterion also has to be established. We use Ward's criterion, which has the advantage of being less sensitive to the outliers [58]. This method merges two clusters S_i and S_j that have the smallest cost to merge measured by the function [59]

$$\Delta(S_i, S_j) = \frac{N_{S_i} N_{S_j}}{N_{S_i} + N_{S_j}} d(c_{S_i}, c_{S_j}) \qquad (2)$$

where $N_{S_{i,j}}$ and $c_{S_{i,j}}$ represent the cardinality and centroid of the clusters $S_{i,j}$, and $d()$ is a function returning the distances between the centroids of the two clusters.

Stopping the aggregation of objects in the HAC implies also the definition of the clusters. However, the threshold distance at which the merging should be stopped is not known a priori. For this aim, here, we first estimate the number of clusters to be extracted and then define the distance threshold accordingly: it will be the value of Ward's distance of (2) at which the merging must be stopped.

Finally, the number of clusters is estimated through a sensitivity analysis using the Silhouette coefficient, which for each sample is defined as follows:

$$s = \frac{b_s - a_s}{\max(a_s, b_s)} \qquad (3)$$

where a_s is the mean distance between a sample and all other points in the same cluster and b_s is the mean distance between a sample and all the other samples in the nearest cluster [60]. The calculation of s is repeated for several numbers of clusters, and the optimal is chosen corresponding to the maximum value of s.
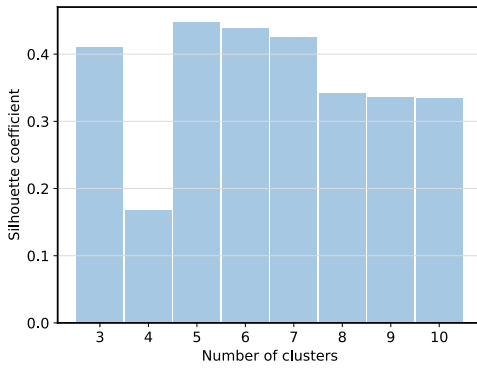
Fig. 3. Silhouette coefficient for different number of clusters, best distribution is related to the number of clusters equal to 5.
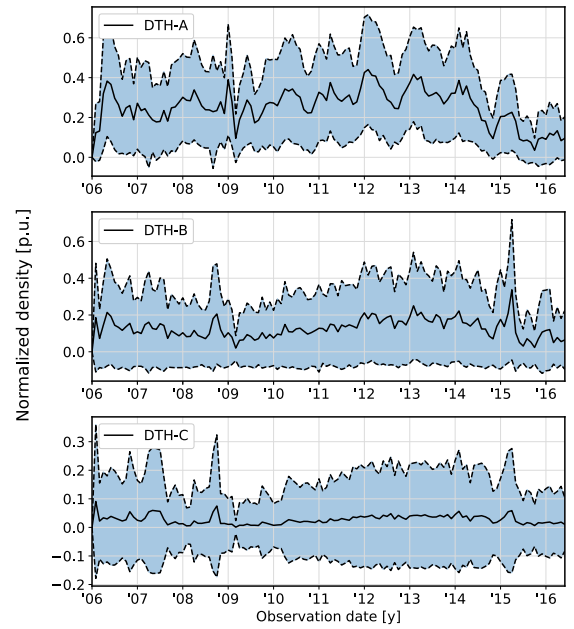

Fig. 4. Mean normalized DTH and SD range (in blue) of the chosen clusters; each plotted value is a mean (or SD) across all the samples present in a domain for a given month.

As it can be seen in Fig. 3, the best distribution is achieved for five clusters. We fix the distance threshold accordingly, and HAC consequently assigns a cluster label ranging between A and E to each cell. However, for the rest of the article, we will discard clusters D and E. The reason is that they involve a very limited number of earthquakes (204 and 2554, respectively), corresponding to the 0.1% and 0.6% of the whole catalog, despite a high number of cells that belong to these clusters (2304 and 203, respectively). In particular, cluster D is composed of cells with at most one earthquake per month (including cells with no earthquakes over the decade); the earthquakes belonging to this cluster are spatially spread over a wide area well outside the core region hosting the injection wells, while the cells with no earthquakes are mostly located at depths larger than 5 km. Cluster E shows DTHs with only a large peak in one month in the analyzed decade, and hypocenters are mostly concentrated at a few spatial spots within the injection region and at depths around 4 km or more. Given the lack of freely available detailed information on the local injection histories, we can only hypothesize this behavior is presumably connected to peculiar local injection operations. As we are interested in the long-term seismic response of the EGS as a whole and in the definition of large-scale spatial domains, we focus on clusters A–C.

To estimate the information lost by discarding these two clusters, we have applied the principal component analysis (PCA) to the mixture composed of the normalized DTHs of the five clusters. PCA projects data from the original feature space into the space of the principal components, providing also the data variance explained by each component [61]. It is usually applied for data reduction, which is performed to minimize the data variance lost in the projection. We found that discarding two dimensions corresponding to clusters D and E implies a 9% loss of the variance of the mixtures made by the normalized DTHs of the five clusters (most of the variance of this 2-D space is due to the large one-point maximum of DTH-E). This means that the majority of the information on the seismic response of the system EGS is described by clusters A–C. A similar or even larger loss of variance when reducing data dimension via PCA has been adopted also by other authors for the selection of the dimension reduction [55], [62], [63].

TABLE I
NUMBER OF CELLS AND EARTHQUAKES IN EACH DOMAIN

Cluster	Number of cells	Number of earthquakes
A	175	299,608
B	339	94,934
C	2,233	20,097

In Fig. 4, the mean normalized DTH of the three selected clusters are plotted and indicated as DTH-A, DTH-B, and DTH-C. To account for time-changing instrumental setup and station coverage, in the definition of the mean DTHs A–C, only events with magnitude above the magnitude of completeness have been selected, which was yearly estimated via the maximum curvature method [64] and varied in the decade in the range (0.3–0.9).

By clustering, we have thus defined three main spatial domains of the volume of the EGS that are composed of the cells of clusters A–C. The number of cells and earthquakes composing each domain is reported in Table I. We remark that the vast majority of earthquakes occur in domain A, which is the domain with the smallest number of nodes. In Fig. 5, the spatial distribution of these domains is displayed in its top view (upper side) and sectional view (lower side) in the underground. The three domains are roughly concentric, with A being the innermost domain and C the outermost domain. Domain A encloses most of the injection wells, in line with the fact that most earthquakes nucleate in that region. Domain A is elongated in the NW-SE direction and shows two main maxima, in a very similar manner as the spatial distribution of the overall earthquake density. Domain A appears to be concentrated in-depth, whereas domains B and C contain cells also at shallow depths. For all domains, the depth of the nodes ranges mostly between 0 and 6 km, although only domain C contains nodes with larger depths.
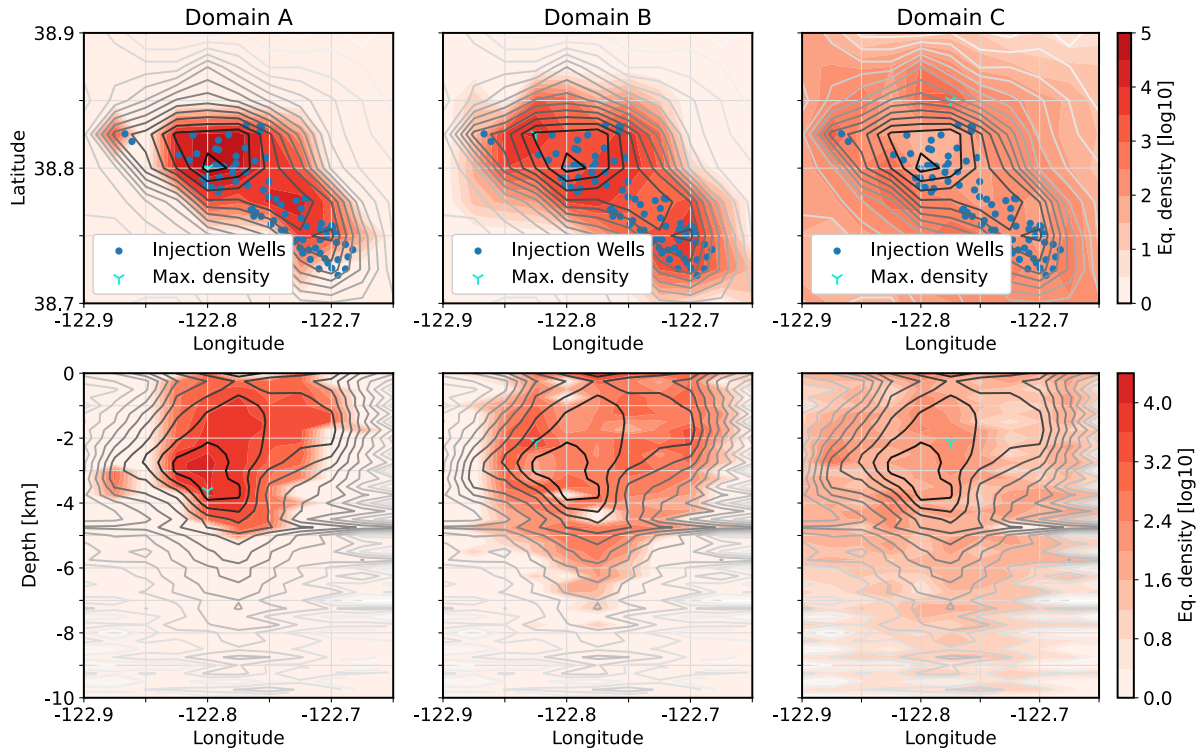
Fig. 5. Spatial earthquake density (in log scale) in defined domains. In the top row, the top-down view is presented, and in the bottom—the side view. For comparison purposes, the contour plot in the background (in gray) displays the overall density of earthquakes (in log scale) in the Geysers region in the selected decade. Blue dots mark the position of the injection wells. Green markers show the position of the cells hosting the largest number of earthquakes over the decade in each domain.

It must be underlined that clustering was performed while only employing DTHs of all the cells as a single input, meaning that any possible information about the cell's spatial location was not included in the clustering analysis. Despite that, the three selected clusters, which have not been created under supervision, are surprisingly composed of cells that are spatially aggregated. This allowed us to define spatial domains providing also an indirect indication that close points in the volume of EGS experience similar stress fields and that local intercell heterogeneities in the seismic response mostly contribute as a second-order effect to the DTHs.

V. DENSITY TIME HISTORIES AND WATER INJECTION HISTORY

Section IV shows that seismicity at EGS can be grouped into three main spatial domains based on the similarity of the unit cells DTHs. Here, we further analyze the relationship between the three main DTHs and the overall water injection history (WIH). In Figs. 6 and 7, the three chosen mean DTHs and WIH are overlapped both in the time and frequency domains. DTH-A and DTH-B show an increase of the earthquake rate broadly in the time interval 2009–2011, which matches the beginning of a new injection cycle and is visible as a step-like increase in the yearly WIH in this time interval, while a decrease in the earthquake rate appears at the end of the injection cycle (2014–2016). WIH is dominated by components with periods of 10–16 months, which originates from the seasonal modulation of the injected water, with a larger amount of water injected in the winter time [38], [65]. These low-frequency components dominate also the spectra of
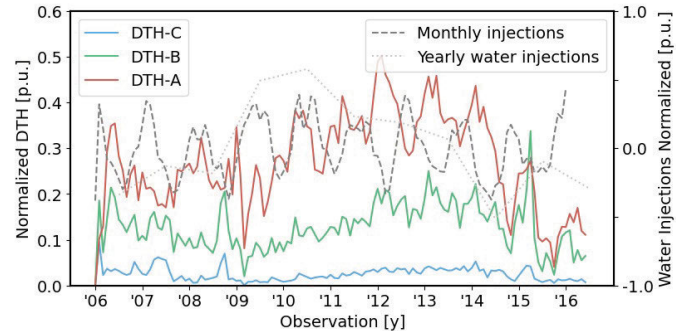

Fig. 6. Mean DTH of the domains A–C together with the monthly and yearly water injection history.

DTH-A–C and are especially visible in DTH-A, but in a less evident fashion, as expected considering the nonlinearity and self-similarity of the seismic response. It is straightforward to locate in domain A the effective source triggering the earthquakes, that is, the local pressure increase due to EGS water injections. Considering the mean injection well depth of 2.6 km [47], in Fig. 5, it can be seen that the location of the wells closely matches the volume defined by domain A. Specifically, we identify the cell with the largest maximum in the DTH as the one hosting (or the closest to) the source of stress change that induces the earthquakes, as it is clearly shown in Fig. 5. We define this cell as S-A and its position is marked with a green marker in the same figure. Here, the positions of the cells of domains B and C with the largest DTH maxima are displayed, as well; they will be referred to as S-B and S-C, respectively.
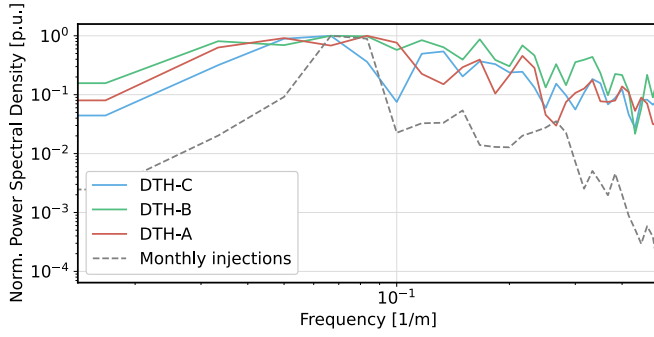
Fig. 7.   Spectra of mean DTH of domains A–C and of the monthly and yearly water injection history. Spectra have been calculated after high-pass filtering the corresponding DTH in the time domain with a cut-off period of three years to remove the step-like effect in the spectrum. Spectra have been normalized to the maxima to compare the relative amplitudes of the frequency peaks.

By looking at the monthly trends depicted in Fig. 6, in some particular years, it is a remarkable an apparent phase lag between the monthly WIH (dashed line) and mean DTHs (especially those of domains A and B). This is confirmed also by the Pearson correlation coefficient between WIH and DTHs that shows a weak inverse correlation (-0.10 for domain A, -0.12 for B, and -0.09 for C), which in turn could indicate a possible delayed triggering mechanism between the WIH and the DTHs.

VI. COULOMB STRESS ESTIMATES

In this section, we try to connect the spatial variability of the earthquake rates over the EGS volume with the amplitude of the triggering stress. Dieterich [66] developed a model that relates changes in Coulomb stress to changes in the seismicity rate. The model is based on the assumption that the sources of the earthquakes are ruled by a rate-and-state friction law [67]; a population of earthquakes ruled by this law evolves through instability in a way that for constant stressing rate, a background seismicity with constant rate is triggered [68]. Coulomb stress τ is defined as

$$\tau = \tau_S + f(\sigma + p) \tag{4}$$

where τ_S is the shear stress acting on a plane, σ is the normal stress, f is the friction coefficient, and p is the pore pressure [67].

Dieterich's model leads to a law describing the time evolution of the seismicity rate R as a function of the Coulomb stress that is a first-order nonlinear differential equation [68]

$$\frac{dR}{dt} = \frac{R}{t_a}\left(\frac{\dot{\tau}}{\dot{\tau}_0} - R\right) \tag{5}$$

where $\dot{\tau}$ is the Coulomb stressing rate, $\dot{\tau}_0$ is the background stressing rate, and t_a is a characteristic decay time defined as

$$t_a \equiv \frac{a\bar{\sigma}}{\dot{\tau}_0} \tag{6}$$

where a is a constitutive parameter quantifying the direct effect on fault slip rate in the rate-state friction law, and $\bar{\sigma}$ is the background effective normal stress [68].

If the stressing rate $\dot{\tau}$ is much larger than the background stressing rate $\dot{\tau}_0$, then $\dot{\tau}/\dot{\tau}_0 \gg R$, and (5) becomes [68]

$$R \approx e^{\Delta\tau/a\bar{\sigma}} \tag{7}$$

where $\Delta\tau$ is the change in Coulomb stress. Here, we assume that the changes of the product $a\bar{\sigma}$ over the volume of the EGS are negligible with respect to the spatial changes of $\Delta\tau$. Under this assumption, $a\bar{\sigma}$ can be considered constant over the whole volume of EGS, and the ratio C_R between the Coulomb stresses at two different positions of the volume can be directly connected to the different earthquake rates

$$C_R = \frac{\log R_j}{\log R_i} = \frac{\Delta\tau_j}{\Delta\tau_i} \tag{8}$$

where $R_{i,j}$ and $\Delta\tau_{i,j}$ are, respectively, the earthquake rates and the change in the Coulomb stress at two positions of the volume. In summary, we thus assume that as follows.

1) The stressing rate is far above the background [68].
2) The product between the constitutive parameter a and the background effective normal stress $\bar{\sigma}$ is fairly constant over the EGS volume.

Many authors estimated typical values of $a\bar{\sigma}$ in the order of $10^4 - 10^5$ Pa and postulated a relevant contribution of the pore pressure at seismogenic depths counteracting the lithostatic loading (see [69], [70] and references therein). For potential spatial variability of $a\bar{\sigma}$ of the same order of magnitude of the product itself, a $\Delta\tau$ well larger than $10^4 - 10^5$ Pa should guarantee that possible heterogeneity of $a\bar{\sigma}$ over the volume would be likely negligible. The existence of spatial domains with similar DTHs supports this assumption, as a strong local heterogeneity in the seismic response would sharply differentiate DTHs of nearby cells even for common stress histories. In general, both assumptions 1 and 2 are expected to be especially true for high injection rates, that is, at the early stages of the injection cycle [68].

We apply (8) using as R_i the maximum of the DTH of each cell and as R_j the maximum of the DTH of a reference cell that we choose as S-A. Therefore, (8) provides us an estimate of the Coulomb stress change between all cells and the reference cell S-A. Starting from the estimates of C_R for all cells, we calculate their distribution, distinguishing the cells belonging to the three domains. For the calculation of C_R, we properly compensated the earthquake rates of nodes of domain C to account for the higher magnitude of completeness of this domain (1.1) than the other two domains (0.9). The results are displayed in Fig. 8(a).

The figure shows that the three domains appear to be distinguished also by different Coulomb stress ratios. In other words, the boundaries of the domains are also roughly defined by the Coulomb stress that is experienced by the corresponding cells. In detail, we see that cells of domain A are mostly affected by a Coulomb stress ranging between the stress experienced by the reference cell and two orders of magnitude lower. Differently, Coulomb stresses affecting cells of domain B are between 2 and 5 orders of magnitude lower than the reference level. Cells with an even larger stress decrease from the reference cell mostly belong to domain C.
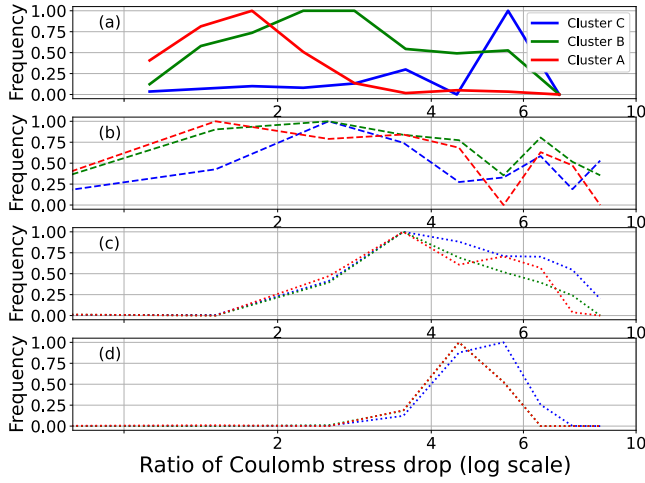
Fig. 8. (a) Distribution of the Coulomb stress ratio (in log scale) between all points of each domain of the EGS volume and a reference position, that is, the cell S-A. (b)–(d) Theoretical pressure drop distributions predicted after 12 days by a pore pressure diffusive model with constant water injection at S-A adopting a diffusivity of (b) $D = 0.05$ m^2/s, (c) $D = 0.5$ m^2/s, and (d) $D = 10$ m^2/s. All curves are normalized by the corresponding distribution maximum.

In the subplots Fig. 8(b)–(d) of the same figure, we have displayed the theoretical stress change experienced by all cells for a constant injection rate q (in Kg/s) at S-A propagating through the volume via pore pressure diffusion. The equation ruling the spatiotemporal pressure propagation is [68], [71], [72]

$$P(r, t) = \frac{q N}{\rho_f 4\pi D r} \mathrm{erfc}\left(\frac{r}{2\sqrt{Dt}}\right) \qquad (9)$$

where r is the distance from the injection point, t is the time from the injection start, D is the hydraulic diffusivity, N is a combination of poroelastic parameters with the dimension of Pa, ρ_f is the mass density of the injected fluid, and $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$, with erf the error function. In the approximation of low-porosity crystalline rocks, $N = [\phi/K_f + \alpha/K_g]^{-1}$, with ϕ porosity, α the Biot–Willis coefficient of the porous medium, $K_{f,g}$ the bulk moduli of the fluid and grain material, respectively [73]. Using $\phi = 0.02$ [40], $\alpha = 0.3$ [68], [73], $K_f = 2.3 \times 10^9$ Pa, and $K_g = 7 \times 10^{10}$ Pa [73], it results $N \simeq 8 \times 10^{11}$ Pa.

We defined $\overline{q} \equiv q/\rho_f$ (which has the dimension of m^3/s) and evaluated (9) using different values of $\overline{q}$ (in the range [0.01–10] m^3/s) and t (between $t = 5$ days and several months). Similarly, we let D range in [0.01–10] m^2/s, which covers most of the interval spanned by D in the crust [74]. Eventually, the combination of parameters that maximizes the overlap with the observed stress change distributions in the least-square sense was selected.

We found that the best fit between observed and theoretical stress change due to a diffusive pressure front propagation is observed for $\overline{q}_0 = 5$ m^3/s, $D_0 = 0.05$ m^2/s, and $t_0 = 12$ days [Fig. 8(b)]. For these values, the model predicts the main characteristics of the frequency distributions for the points of domains A and B (maximum at low-pressure changes for domain A and frequency decrease at larger pressure changes, shifted maximum for domain B). Instead, pressure changes at domain C appear to be mostly underestimated,
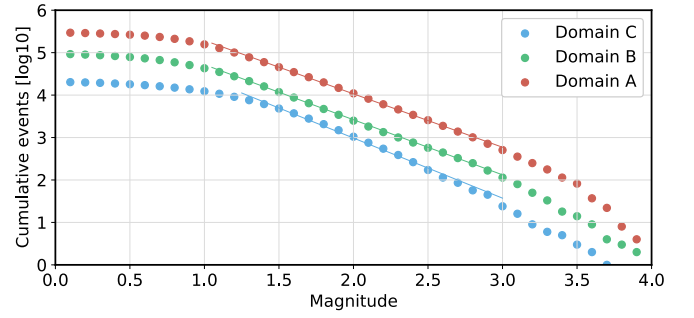

Fig. 9. GR distribution calculated using the earthquakes nucleated in the three domains. The linear fitting has been computed in the magnitude range $(M_C + 0.2)$–3.

suggesting that a pure pore-pressure diffusive model with the selected parameters cannot explain the observed seismicity. To match the observed earthquake ratio distribution of domain C, we need to use $D = 10$ m^2/s [Fig. 8(d)]; in this case, however, any difference in the seismicity rate of the three domains would basically disappear (the three earthquake ratio distributions are nearly overlapped), in contrast with the observations. As expected, an increase of D leads to a faster pressure front propagation and then to larger stress changes [see Fig. 8(b)–(d)]. We remark that t_0 must be intended as an indicative time interval during which the diffusive process acts, given the assumed idealization that the extended injection area is approximated with a point. Remarkably, dividing $\overline{q}_0$ by the number of injection wells (about 70), we found an injection rate per well of about 0.07 m^3/s, which corresponds to about 6000 m^3/day that is close to the typical injection peak rate at a well [40], [75], whereas the mean water injection rate over the considered decade is about 110×10^9 lbs per year [42] corresponding to about 1.4 m^3/s.

VII. b-VALUE

To test this distinction of the domains from the point of view of the magnitude of the stress field, we calculated the b-value of the Gutenberg–Richter (GR) distribution (a measure of the frequency of earthquakes as a function of magnitude) separately for the three domains. In the GR distribution, the number of earthquake N scales with the magnitude M with the form [76]

$$\log N(M) = a - bM \qquad (10)$$

where $N(M)$ is the number of earthquakes with a magnitude equal to or larger than M and the coefficient b rules the ratio between the frequencies of small and large earthquakes. Moreover, an increase in the b-value is normally interpreted as a stress decrease [77].

We calculate the GR distribution separately for the earthquakes nucleated in the cells of the three domains and estimate the b-value with a linear least-square best-fit approach (Fig. 9). To guarantee the fit in a magnitude range with well-populated bins, that is where the real power law scaling is best appreciable, for all domains, the fitting has been performed in the magnitude range $(M_C + 0.2)$–3, with M_C being the magnitude of completeness. M_C was estimated by the maximum curvature
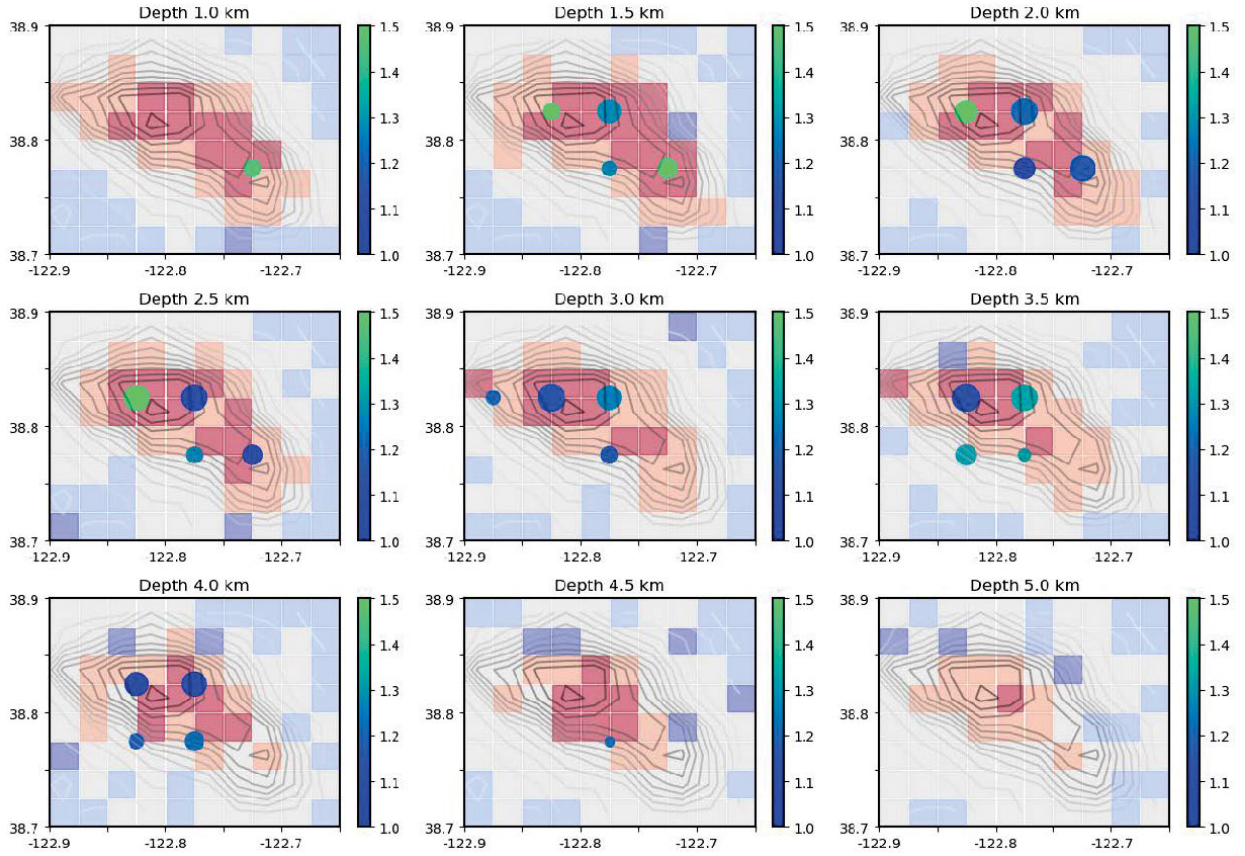
Fig. 10. Dots mark the estimated *b*-value within macrocells. The positions of the dots mark the center of the macrocell and their size inversely scales with the error (the larger the size, the lower the error). Background colors indicate the cluster the cell belongs to. Clusters A–E are shown, respectively, with the colors red, light orange, blight gray, light blue, and dark blue. In the title of each subplot, the upper bound of the depth range is shown (f.i., depth 1.0 km means that in the subplot the solutions in the depth range 1–1.5 km are displayed). Background overall earthquake density (in log scale) is shown by the gray isolines.

TABLE II

b-VALUES OF THE GR DISTRIBUTION OF THE THREE DOMAINS AND CORRESPONDING UNCERTAINTY

Domain	b-value	error
A	1.25	0.03
B	1.30	0.03
C	1.41	0.04

approach [64] and as 0.9 for domains A and B and 1.1 for domain C. We adopted a bootstrapping approach [78]: the fitting was repeated 100 times on distributions built by random sampling (with replacement) of the selected earthquake catalog. The resulting *b*-value and its uncertainty were eventually fixed as the mean value and the standard deviation (SD) of the ensemble, respectively.

The results are displayed in Table II and clearly indicate that domains A and B have similar *b*-values in the range 1.2–1.3, in line with the calculation of other authors [40], [79], who, however, processed datasets of hypocenters enclosed in a small region (about 1 × 1 km) falling in our domain A. In contrast, the *b*-value of domain C appears to be larger, even considering the errors, and is about 1.4. For the sake of completeness, the *b*-values of the discarded clusters D and E have also been calculated using the same approach and are, respectively, 1.74 ± 0.49 and 1.47 ± 0.13.

We have analyzed in more detail the spatial *b*-value variability splitting the EGS volume into cells of 0.5° × 0.5° × 0.5 km (latitude, longitude, and depth), that is, for this analysis, we have fused together eight neighboring cells of the original distribution. We refer to them as macrocells. To ensure the stability of the calculations, we have selected only macrocells with 1000 earthquakes or more and computed the calculation of the *b*-value as described above. In Fig. 10, the results are displayed in subplots corresponding to different depth intervals. Only solutions with errors lower than 0.2 are shown.

Given the spatial earthquake density required by this selection, we mostly sample the volumes of domains A and B. It is observed that *b*-values within the macrocells are generally consistent with the mean values of Table II and mostly range between 1.1 and 1.4. A deviation from the mean value appears in volumes located on the map close to the area with the largest earthquake density: for these volumes, the *b*-value falls above 1.4 at depths up to 3 km, while at larger depths it is lower than 1.3. This evidence can be considered as an indication that the injection significantly affects the natural variation of the *b*-value.

A similar *b*-value discontinuity appears also close to the secondary peak of the earthquake density (southeast of the primary), but the discontinuity occurs at depths around 2 km. Overall, a decrease of *b*-values can be observed while moving

from shallow to deep volumes, consistently with seismological observations [69].

VIII. Discussion

All results described above are compatible with a principal stress source located in the area of EGS with many injection wells and the largest density of earthquakes.

This induces a stress field that is mainly experienced by the cells of domain A and that triggers the earthquakes in this domain with a characteristic rate described by DTH-A. The density of injection wells in domain B is significantly lower than in domain A (Fig. 5). Similarly, the earthquake density is about three times lower than in domain A. On these bases, and considering that also DTH-B contains an imprinting of the WIH, DTH-B can be allegedly imputed to the wastewater injection and may represent a different overall seismic response to a common stress step broadly originated in S-A. This is also compatible with the evidence that the mean *b*-values of domains A and B are similar, suggesting similar mean local medium conditions.

On the other hand, the depth-dependent *b*-values in some parts of domains A and B suggest some small-scale medium heterogeneity; specifically, the *b*-value contrast at depths of 2–3 km can be related to volumes at depths of 2–4 km where tomographic studies locate low V_p and Q_p possibly caused by injected fluids [80], and/or to a transition from unaltered to thermally altered greywacke or granite in the reservoir [47], [81]. A domain could thus enclose different geological units, but a broadly common large-scale stress field acting on them triggers a site-independent earthquake production.

We impute the different shapes of DTH-A and DTH-B, and, in particular, the smoother dependence of earthquake density in DTH-B on the modulation of the injection cycle to a distance effect from the main stress source. The calculation of the Coulomb stress change indicates that the boundary between these two different regimes of seismic response to a unique stress source is given by a stress change of about two orders of magnitudes.

In Dietrich's model of triggered seismicity [66], the earthquake rate following a uniform stress change on a circular crack shows different regimes at different distances from the stress source, with a nearly distance-independent rate at short distances and a power law decay at larger distances (see [66, eq. (21) and Fig. 4]). In this framework, DTH-A and DTH-B would represent the seismic responses at different distances to the same stress source: with the distance effects in domain A being secondary in the earthquake production. Dietrich's model connects the distance from the source to the position at which the change of the seismic response occurs with the source (crack) size. Depending on the model parameters ($\Delta\tau$, $a\bar{\sigma}$), the distance separating the two domains is about 1.2–1.4 times the crack size. Considering here a distance separating the two regimes equal to the distance between the cells S-A and S-B (about 3–5 km), the crack size would be in the order of a few 2–3 km. It would represent the size of the EGS region where the water injection effectively acts as a macroscopic stress step. In this framework, an open question is: how does the stress propagate from the main
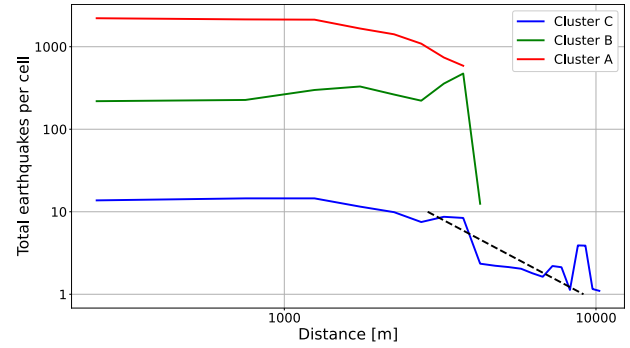

Fig. 11. Number of earthquakes as a function of distance from S-A distinguishing the three spatial domains. The dotted black line displays the theoretical spatial decay of the earthquake rate for a poroelastic mechanism.

stress source outwards? Goebel and Brodsky [82] analyzed two propagation mechanisms, that are pressure diffusion and poroelastic propagation, and showed that they are distinguished by the way the earthquake density decays by distance from the source. Specifically, the pressure diffusion dominates in the near field and shows a nearly constant earthquake density at distances up to (in order of magnitude) $\sim$1 km followed by an abrupt decay at larger distances, while by poroelastic propagation, the earthquake density decays with distance as a power law. We test this difference by plotting the number of earthquakes in the whole decade as a function of the distance from a reference cell that was chosen as S-A. In Fig. 11, we plot the logarithm of the mean number of earthquakes in each cell within a given distance interval from S-A distinguishing the cells of the three domains. Domain A and B trends show compatible patterns with those available in [82] which are imputed to the pressure diffusion effect. That provides a nearly constant earthquake density at short distances followed by a sharp drop, which is visible only for domain B. This suggests that the spatial volume of domain B encloses the whole range of seismic responses induced by a diffusion mechanism, while in domain A, the near-source effects prevail. On the contrary, the pattern shown by domain C matches very well the one typical of sites with high poroelastic coupling [82], that is a constant earthquake rate at short distances and a power-law decay at large distances. In our case, the separation between pore pressure-diffusion- and the poroelasticity-dominated regimes appears to occur at about $\bar{x} \simeq 3.5$ km from the source (Fig. 11), which should be considered as an upper bound given the idealized condition of a point source that we are assuming.

A pore pressure front propagates with the law $r = (4\pi Dt)^{1/2}$ [74]. Using D_0 and t_0, this equation predicts $r_0 = (4\pi D_0 t_0)^{1/2} \simeq 1$ km; the same equation instead for $r = \bar{x}$ and $D = D_0$ predicts $t' \simeq 225$ days, while keeping $t = t_0$, it provides a diffusivity of about 0.9 m²/s. Given the theoretical curves of Fig. 8, we consider unlikely a hydraulic diffusivity much larger than D_0 (under the validity of a pure diffusive pore-pressure triggering mechanism). As $\bar{x}$ is an upper limit of the distance separating the two regimes, t' must be considered an upper bound of the time interval during which the diffusive process may act (potentially promoting delayed triggering), whereas the diffusion must trigger most of the earthquakes (of domains A and B) in a time interval in the order of t_0.

A hydraulic diffusivity close to 0.05 m²/s was found also by [83] and [84] from the data of the hot dry rock experiment in Soultz, France (1993). On the contrary, our estimate of D is lower than the value estimated at The Geysers in [40], who provide $D = 10$ m²/s. However, this value has been found focusing on a much smaller area of The Geysers (in the order of 1×1 km) around two injection wells and falls close to the upper bound of the diffusivity range estimated for the Earth crust [74].

We depict thus a scheme where the earthquakes are induced by a pore-pressure-dominated diffusion mechanism for distances within about 3.5 km from S-A, which mostly encloses the domains A-B, while for larger distances a poroelastic effect dominates the earthquake triggering, and this effect is mostly visible for the earthquakes of domain C. This is compatible with the mismatch between the observed earthquake rate in domain C and the rate predicted by a pore pressure diffusion mechanism (Fig. 8). Based on the results of Section VI, we deduce that the transition between pressure diffusion and poroelastic effect occurs roughly when the stress decreases by five orders of magnitudes or more. A different triggering mechanism for domain C is also compatible with the larger b-value observed in this domain. Larger b-values are normally associated with lower stress conditions in the rock [77]. In line with that, [40] found a decrease in the b-value for an increase in the injected fluids. We interpret the different b-values of domains A–B and domain C—in combination with the different decays of earthquake rates and stress conditions—as the effect of different states and geometry of the activated fault networks. Moreover, domain C is the only one containing deep (> 6 km) earthquakes. All these elements let us hypothesize that the earthquakes of domain C nucleate on an existing prestressed fault network [82] and are mainly triggered by stress transferred from domain A via a poroelastic effect.

However, from Fig. 11, a peak in the earthquake density distribution of domain C is clearly visible for large distances (between 9 and 10 km from S-A) that mismatches the power decay of the poroelastic mechanism. We thus hypothesize a minor contribution from a secondary source not directly connected to the water injection detectable only at very large distances from the injection wells. Assuming that this source is identifiable with the tectonic loading, we try to estimate the yearly cumulative fault slip triggered by this supposed secondary source in the cell where it appears to be most effective, that is the cell located at distances larger than 9 km from S-A and showing the largest cumulative number of earthquakes over the decade. This cell has coordinates $(-122.75°, 38.7°, 12.75$ km) (lat, long, depth) and will be labeled as S-SEC. In S-SEC, 26 earthquakes nucleate over the decade.

From the local magnitudes (M_L) used in our catalog and assuming $M_W = (2/3)M_L + 1.14$ [85], an estimate of the seismic moment M_0 of each earthquake of S-SEC can be obtained from the relation [86]

$$M_0 = 10^{1.5 \ M_w + 9.1}. \tag{11}$$

For a circular fault experiencing a uniform and instantaneous stress drop over the whole slipping area, the seismic moment is related to the stress drop $\Delta\sigma$ and the radius a of the fault by the equation [87], [88]

$$M_0 = \frac{16}{7}\Delta\sigma a^3. \tag{12}$$

We assume a constant stress drop for all earthquakes ranging in $\Delta\sigma = 0.1$–5 MPa, which is based on previous estimates of different authors [75], [89], [90]. For a constant stress drop, the fault slip s is a function of the seismic moment M_0, the fault size a, and the shear modulus μ following the law [88]:

$$M_0 = \frac{7\pi}{16}\mu\frac{s}{A}. \tag{13}$$

Fixing the shear modulus to $\mu = 30$ GPa [91], we associate a fault slip with each earthquake. We estimate the fault slip of the earthquakes nucleated in cell S-SEC, then calculate the cumulative fault slip over the whole analyzed decade, and finally normalize the value to one year. The resulting yearly cumulative slip $\bar{s}_y$ ranges between 1 and 14 mm. Even though this is intended as a rough indication of the slip induced by the secondary source, it is nonetheless remarkable that $\bar{s}_y$ is in line with the mean values of the yearly slip on the two main faults enclosing the Geysers, that are the Collayomi and the Maacama fault, which show, respectively, a yearly fault slip of about 1 and 13 mm [92]. We remark moreover that S-SEC is located at larger depths (about 12 km) than most of the seismicity at EGS (mostly shallower than 6 km). Given all elements, we think that it is reasonable to hypothesize that the secondary source is at least partially an effect of the regional tectonic loading. Moreover, as S-SEC is located relatively close to the Maacama fault, which shows the largest mean cumulative slip, we can argue that the correct values of $\Delta\sigma$ for the events of this cell are probably in the range of 1–10 MPa. Assuming that the DTH of the cell S-SEC is totally imputable to this secondary injection-unrelated stress source, an estimate of the relevance of this source in inducing earthquakes with respect to the wastewater injection can be given by the equation

$$F_T = \frac{\overline{N}_{\text{S-SEC}}}{N_{\text{S-A}}} \tag{14}$$

where $\overline{N}_{\text{S-SEC}}$ is the total number of earthquakes over the decade on the cell S-SEC (compensated for the higher M_C in domain C) and $N_{\text{S-A}}$ is the total number of earthquakes over the decade on the cells S-A. In this way, we estimate $F_T \simeq 0.5\%$.

IX. CONCLUSION

We have described the seismic response to the exploitation of the EGS of The Geysers, in California, with a new ML-based approach that considers the entire volume as a whole mechanical system seismically reacting to the stress produced principally by the water injection. In the proposed approach, the time series of a limited number of parameters such as location, time, and magnitude of the earthquakes occurring over one decade were analyzed. They are considered as the medium response to local stress. In detail, we divided the volume of the EGS into cells of equal volumes and calculated the earthquake rate over time in each cell that

we treat as a proxy of the local stress history assuming that spatial small-scale (cell size) heterogeneities of the seismic response are negligible for the earthquake production with respect to the main stress field acting on larger scales. We do not take into account aseismic accommodation of the stress. Under these assumptions, by employing the proposed approach based on clustering, the following main results have been found.

1) Based on the similarity of the DTHs, three spatial domains have been defined that are concentric to the region where most of the earthquakes occurred.
2) The domains are principally a result of different seismic responses of the medium to a common pressure increase from the wastewater injection. The boundaries of these domains can be roughly identified by points of the volume where the Coulomb stress decreases by about two and five orders of magnitude.
3) The stress field propagates from the principal exploitation area via a diffusion process up to distances of about 3.5 km, whereas for larger distances, a poroelastic mechanism dominates potentially combined with the background tectonic loading. We infer an overall mean hydraulic diffusivity in the order of 0.05 m^2/s.
4) The earthquake rates of a few cells of the outermost domain located far away from the injection wells misfit the spatial decay predicted by a poroelastic mechanism. For these cells, we hypothesize the contribution of the regional tectonic loading, which can trigger up to about 0.5% of the injection-induced earthquakes corresponding to a cumulative yearly slip in the order of 1–10 mm, in line with the mean slipping rates of the active faults enclosing the EGS.
5) In terms of b-values, the volume can be divided into a low b-value (about 1.2–1.3) volume that is nearly composed of the innermost and intermediate spatial domains, and a high b-value (about 1.4) volume that broadly matches the outermost spatial domain. Insights on a contrast in the b-value at depths of 2–3 km in domains A and B are observed, as well.

The approach presented here can potentially allow for differentiation of the seismic risk into the domains based on the different earthquake rates, source distance, Coulomb stress change, and b-values. Moreover, this approach can identify domains of cells with very weak seismicity or totally aseismic, potentially providing a tool to highlight regions of EGS with depleted stress, with important implications for the seismic risk assessment (e.g., [93]).

Acknowledgment

References

[1] E. Barbier, "Geothermal energy technology and current status: An overview," *Renew. Sustain. Energy Rev.*, vol. 6, nos. 1–2, pp. 3–65, Jan. 2002.

[2] M. E. H. Assad, M. N. AlMallahi, M. AlShabi, and H. Delnava, "Heating and cooling by geothermal energy," in *Proc. Adv. Sci. Eng. Technol. Int. Conf. (ASET)*, Feb. 2022, pp. 1–7.

[3] H. Kulasekara and V. Seynulabdeen, "A review of geothermal energy for future power generation," in *Proc. 5th Int. Conf. Adv. Electr. Eng. (ICAEE)*, Sep. 2019, pp. 223–228.

[4] P. J. Ball, "A review of geothermal technologies and their role in reducing greenhouse gas emissions in the USA," *J. Energy Resour. Technol.*, vol. 143, no. 1, pp. 1–12, Jan. 2021.

[5] B. Vonsée, W. Crijns-Graus, and W. Liu, "Energy technology dependence—A value chain analysis of geothermal power in the EU," *Energy*, vol. 178, pp. 419–435, Jul. 2019.

[6] C. A. Vargas, L. Caracciolo, and P. J. Ball, "Geothermal energy as a means to decarbonize the energy mix of megacities," *Commun. Earth Environ.*, vol. 3, no. 1, p. 66, Mar. 2022.

[7] M.-A. Evans, C. Bono, and Y. Wang, "Toward net-zero electricity in Europe: What are the challenges for the power system?" *IEEE Power Energy Mag.*, vol. 20, no. 4, pp. 44–54, Jul. 2022.

[8] A. Anderson and B. Rezaie, "Geothermal technology: Trends and potential role in a sustainable future," *Appl. Energy*, vol. 248, pp. 18–34, Aug. 2019.

[9] T. J. Hammons and A. Gunnarsson, "Geothermal sustainability in Europe and worldwide," in *Proc. 43rd Int. Universities Power Eng. Conf.*, Sep. 2008, pp. 1–10.

[10] R. Bertani, "Geothermal power generation in the world update report," *Geothermics*, vol. 60, pp. 31–43, Mar. 2016.

[11] X. Yang, Y. Song, G. Wang, and W. Wang, "A comprehensive review on the development of sustainable energy strategy and implementation in China," *IEEE Trans. Sustain. Energy*, vol. 1, no. 2, pp. 57–65, Jul. 2010.

[12] N. Wu and X. Liu, "Geothermal energy current state and utilization in oilfield, China: An overview," in *Proc. Int. Conf. Electr. Control Eng.*, Sep. 2011, pp. 5278–5281.

[13] B. Li, F. Li, P. Zhou, T. Hong, and F. Liu, "The geothermal resources assessment of jinghong basin in Yunnan province," in *Proc. Int. Conf. Electr. Control Eng.*, Sep. 2011, pp. 2695–2698.

[14] K. Khan, M. Ahmed, Md. S. Parvez, and Md. M. Hossain, "Scope of geothermal potential of Bangladesh: A review," in *Proc. 3rd Int. Conf. Green Energy Technol. (ICGET)*, Sep. 2015, pp. 1–4.

[15] M. T. Islam et al., "Trends and prospects of geothermal energy as an alternative source of power: A comprehensive review," *Heliyon*, vol. 8, no. 12, Dec. 2022, Art. no. e11836.

[16] M. Rumbayan and K. Nagasaka, "Prioritization decision for renewable energy development using analytic hierarchy process and geographic information system," in *Proc. Int. Conf. Adv. Mech. Syst.*, Sep. 2012, pp. 36–41.

[17] Nasruddin et al., "Potential of geothermal energy for electricity generation in indonesia: A review," *Renew. Sustain. Energy Rev.*, vol. 53, pp. 733–740, Jan. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032115010023

[18] M. Frey, K. Bär, I. Stober, J. Reinecker, J. van der Vaart, and I. Sass, "Assessment of deep geothermal research and development in the upper Rhine graben," *Geothermal Energy*, vol. 10, no. 1, p. 18, Sep. 2022.

[19] K. P. Tsagarakis et al., "A review of the legal framework in shallow geothermal energy in selected European countries: Need for guidelines," *Renew. Energy*, vol. 147, pp. 2556–2571, Mar. 2020.

[20] P. Dumas, "Policy and regulatory aspects of geothermal energy: A European perspective," in *Geothermal Energy Soc.*, Cham, Switzerland: Springer Int. Publishing, 2019, pp. 19–37.

[21] L. Rybach and M. Mongillo, "Geothermal sustainability-a review with identified research needs," *GRC Trans.*, vol. 30, pp. 1083–1090, Sep. 2006.

[22] V. Stefansson, "The renewability of geothermal energy," in *Proc. World Geothermal Congress*, May 2000, pp. 883–888.

[23] G. R. Foulger, M. P. Wilson, J. G. Gluyas, B. R. Julian, and R. J. Davies, "Global review of human-induced Earthquakes," *Earth-Sci. Rev.*, vol. 178, pp. 438–514, Mar. 2018.

[24] C. Doglioni, "A classification of induced seismicity," *Geosci. Frontiers*, vol. 9, no. 6, pp. 1903–1909, Nov. 2018.

[25] A. Vengosh, R. B. Jackson, N. Warner, T. H. Darrah, and A. Kondash, "A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States," *Environ. Sci. Technol.*, vol. 48, no. 15, pp. 8334–8348, Aug. 2014.

[26] L. Di, P. Yue, H. K. Ramapriyan, and R. L. King, "Geoscience data provenance: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 11, pp. 5065–5072, Nov. 2013.

[27] X. Li et al., "Multi-GNSS meteorology: Real-time retrieving of atmospheric water vapor from BeiDou, galileo, GLONASS, and GPS observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6385–6393, Dec. 2015.

[28] S. Liu et al., "Detection of geothermal anomaly areas with spatiotemporal analysis using multitemporal remote sensing data," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 14, pp. 4866–4878, 2021.

[29] N. Deichmann and J. Ernst, "Earthquake focal mechanisms of the induced seismicity in 2006 and 2007 below Basel (Switzerland)," *Swiss J. Geosci.*, vol. 102, no. 3, pp. 457–466, Dec. 2009.

[30] M. Cardiff et al., "Geothermal production and reduced seismicity: Correlation and proposed mechanism," *Earth Planet. Sci. Lett.*, vol. 482, pp. 470–477, Jan. 2018, doi: 10.1016/j.epsl.2017.11.037.

[31] J. D. Smith, E. R. Heimisson, S. J. Bourne, and J.-P. Avouac, "Stress-based forecasting of induced seismicity with instantaneous earthquake failure functions: Applications to the Groningen gas reservoir," *Earth Planet. Sci. Lett.*, vol. 594, Sep. 2022, Art. no. 117697.

[32] J. C. Robins et al., "2021 U.S. geothermal power production and district heating market report," USA, Jul. 2021. [Online]. Available: https://www.osti.gov/biblio/1808679, doi: 10.2172/1808679.

[33] O. S. Boyd, "Analysis of seismic moment tensors, in-situ stress, and finite-source scaling of Earthquakes at the geysers geothermal field, California," Ph.D. dissertation, Dept. Earth Planet. Sci., UC Berkeley, Berkeley, CA, USA, 2018.

[34] J. Combs, S. K. Sanyal, and C. Haukwa, "Use of remote sensing for monitoring the geysers geothermal field, California," *GSA Special Papers*, vol. 1, no. 29, pp. 43–57, 2000.

[35] J. Rutqvist et al., "The Northwest geysers EGS demonstration project, California—Part 2: Modeling and interpretation," *Geothermics*, vol. 63, pp. 120–138, Sep. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0375650515001030

[36] K. Watson, "Geologic applications of thermal infrared images," *Proc. IEEE*, vol. 63, no. 1, pp. 128–137, 1975.

[37] R. C. Gunasekera, G. R. Foulger, and B. R. Julian, "Reservoir depletion at the geysers geothermal area, California, shown by four-dimensional seismic tomography," *J. Geophys. Res. Solid Earth*, vol. 108, no. B3, pp. 1–11, Mar. 2003.

[38] E. L. Majer and J. E. Peterson, "The impact of injection on seismicity at the geysers, California geothermal field," *Int. J. Rock Mech. Mining Sci.*, vol. 44, no. 8, pp. 1079–1090, Dec. 2007.

[39] D. Eberhart-Phillips and D. H. Oppenheimer, "Induced seismicity in the geysers geothermal area, California," *J. Geophys. Res. Solid Earth*, vol. 89, no. B2, pp. 1191–1207, Feb. 1984.

[40] P. Martínez-Garzón, G. Kwiatek, H. Sone, M. Bohnhoff, G. Dresen, and C. Hartline, "Spatiotemporal changes, faulting regimes, and source parameters of induced seismicity: A case study from the geysers geothermal field," *J. Geophys. Res. Solid Earth*, vol. 119, no. 11, pp. 8378–8396, Nov. 2014.

[41] P. Martínez-Garzón, G. Kwiatek, M. Bohnhoff, and G. Dresen, "Impact of fluid injection on fracture reactivation at the geysers geothermal field," *J. Geophys. Res. Solid Earth*, vol. 121, no. 10, pp. 7432–7449, Oct. 2016.

[42] R. Gritto and S. P. Jarpe, "Temporal variations of Vp/Vs-ratio at the geysers geothermal field, USA," *Geothermics*, vol. 52, pp. 112–119, Oct. 2014.

[43] 2014 *Northern California Earthquake Data Center, Uc Berkeley Seismological Laboratory*. Guindy, chennai, India: NCEDC.

[44] J. R. Murray, S. E. Minson, and J. L. Svarc, "Slip rates and spatially variable creep on faults of the northern san Andreas system inferred through Bayesian inversion of global positioning system data," *J. Geophys. Res. Solid Earth*, vol. 119, no. 7, pp. 6023–6047, Jul. 2014.

[45] S. G. Wesnousky, "Earthquakes, quaternary faults, and seismic hazard in California," *J. Geophys. Res. Solid Earth*, vol. 91, no. B12, pp. 12587–12631, Nov. 1986.

[46] S. Enedy, K. Enedy, and J. Maney. (1991). *Reservoir Response To Injection in the Southeast Geysers*. [Online]. Available: https://www.osti.gov/biblio/887486

[47] J. Farison, "Geysers lessons learned and egs results," in *Proc. Geothermal Res. Soc. Jpn. EGS Workshop*, 2009, pp. 1–61.

[48] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts Techniques*. Burlington, MA, USA: Morgan Kaufmann, 2022.

[49] V. Estivill-Castro, "Why so many clustering algorithms: A position paper," *ACM SIGKDD Explor. Newslett.*, vol. 4, no. 1, pp. 65–75, Jun. 2002.

[50] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *J. Intell. Inf. Syst.*, vol. 17, no. 2, pp. 107–145, Dec. 2001.

[51] F. Nielsen, "Hierarchical clustering," in *Introduction to HPC With MPI for Data Science*. Springer, 2016, pp. 195–211.

[52] R. Xu and D. C. Wunsch, "Clustering algorithms in biomedical research: A review," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 120–154, 2010.

[53] M. Palo, B. Schubert, J. Wei, and W. Liu, "Clustering-based discrimination of multiple partial discharge sources: A case study," in *Proc. IEEE Milan PowerTech*, 2019, pp. 1–6, doi: 10.1109/PTC.2019.8810743.

[54] M. Palo, B. Schubert, E. Ogliari, J. Wei, K. Gu, and W. Liu, "Detection, features extraction and classification of radio-frequency pulses in a high-voltage power substation: Results from a measurement campaign," in *Proc. IEEE 3rd Int. Conf. Dielectr.*, Jul. 2020, pp. 677–680.

[55] P. Shi, L. Seydoux, and P. Poli, "Unsupervised learning of seismic wavefield features: Clustering continuous array seismic data during the 2009 L'Aquila Earthquake," *J. Geophys. Res. Solid Earth*, vol. 126, no. 1, Jan. 2021, Art. no. e2020JB020506.

[56] M. Palo, M. Picozzi, G. De Landro, and A. Zollo, "Microseismicity clustering and mechanic properties reveal fault segmentation in southern Italy," *Tectonophysics*, vol. 856, Jun. 2023, Art. no. 229849, doi: 10.1016/j.tecto.2023.229849.

[57] E. Ogliari, M. Sakwa, J. Wei, W. Liu, B. Schubert, and M. Palo, "General machine learning-based approach to pulse classification for separation of partial discharges and interference," *IEEE Sensors J.*, vol. 23, no. 21, pp. 26839–26849, Nov. 2023.

[58] R. C. de Amorim, "Feature relevance in ward's hierarchical clustering using the Lp norm," *J. Classification*, vol. 32, no. 1, pp. 46–62, Apr. 2015.

[59] J. Ward, "Hierarchical grouping to optimize an objective function," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.

[60] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.

[61] G. H. Dunteman, "Principal component analysis," *Chemometrics Intell. Laboratory Syst.*, vol. 69, nos. 1–3, pp. 37–52, Aug. 1989.

[62] S. De Martino, M. Palo, and G. B. Cimini, "A statistical study of the stromboli volcano explosion quakes before and during eruptive crisis," *J. Geophys. Res.*, vol. 116, no. B4, pp. 1–15, Apr. 2002.

[63] B. Schubert, M. Palo, and T. Schlechter, "Review of uhf-based signal processing approaches for partial discharge detection," in *Proc. Comput. Aided Syst. Theory—EUROCAST, 16th Int. Conf.*, vol. 10672, pp. 219–226, 2018. [Online]. Available: http://link.springer.com/10.1007/978-3-319-74727-9_26

[64] S. Wiemer and K. Katsumata, "Spatial variability of seismicity parameters in aftershock zones," *J. Geophys. Res. Solid Earth*, vol. 104, no. B6, pp. 13135–13151, Jun. 1999.

[65] P. Martínez-Garzón, M. Bohnhoff, G. Kwiatek, and G. Dresen, "Stress tensor changes related to fluid injection at the geysers geothermal field, California," *Geophys. Res. Lett.*, vol. 40, no. 11, pp. 2596–2601, Jun. 2013.

[66] J. Dieterich, "A constitutive law for rate of earthquake production and its application to earthquake clustering," *J. Geophys. Res. Solid Earth*, vol. 99, no. B2, pp. 2601–2618, Feb. 1994.

[67] C. H. Scholz, "Earthquakes and friction laws," *Nature*, vol. 391, no. 6662, pp. 37–42, Jan. 1998.

[68] P. Segall and S. Lu, "Injection-induced seismicity: Poroelastic and earthquake nucleation effects," *J. Geophys. Res. Solid Earth*, vol. 120, no. 7, pp. 5082–5103, Jul. 2015.

[69] F. Catalli, M. Meier, and S. Wiemer, "The role of Coulomb stress changes for injection-induced seismicity: The Basel enhanced geothermal system," *Geophys. Res. Lett.*, vol. 40, no. 1, pp. 72–77, Jan. 2013.

[70] T. J. Ader, N. Lapusta, J.-P. Avouac, and J.-P. Ampuero, "Response of rate-and-state seismogenic faults to harmonic shear-stress perturbations," *Geophys. J. Int.*, vol. 198, no. 1, pp. 385–413, May 2014.

[71] J. W. Rudnicki, "Fluid mass sources and point forces in linear elastic diffusive solids," *Mech. Mater.*, vol. 5, no. 4, pp. 383–393, Dec. 1986.

[72] T. M. Müller, "Spatiotemporal pore pressure evolution due to fluid-mass point sources in dynamic poroelasticity," *Geophys. J. Int.*, vol. 165, no. 3, pp. 906–912, Jun. 2006.

[73] S. A. Shapiro, E. Huenges, and G. Borm, "Estimating the crust permeability from fluid-injection-induced seismic emission at the KTB site," *Geophys. J. Int.*, vol. 131, no. 2, pp. F15–F18, Nov. 1997.

[74] M. Parotidis, S. A. Shapiro, and E. Rothert, "Evidence for triggering of the vogtland swarms 2000 by pore pressure diffusion," *J. Geophys. Research: Solid Earth*, vol. 110, no. B5, pp. 1–12, May 2005.

[75] G. Kwiatek, P. Martínez-Garzón, G. Dresen, M. Bohnhoff, H. Sone, and C. Hartline, "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*, vol. 120, no. 10, pp. 7085–7101, Oct. 2015, doi: 10.1002/2015jb012362.

[76] B. Gutenberg and C. F. Richter, "Earthquake magnitude, intensity, energy, and acceleration," *Bull. Seismol. Soc. Amer.*, vol. 46, no. 2, pp. 105–145, Apr. 1956.

[77] C. H. Scholz, "On the stress dependence of the earthquake b value," *Geophys. Res. Lett.*, vol. 42, no. 5, pp. 1399–1402, Mar. 2015.

[78] B. Efron, "The bootstrap and modern statistics," *J. Amer. Stat. Assoc.*, vol. 95, no. 452, p. 1293, Dec. 2000.

[79] K. Leptokaropoulos, M. Staszek, S. Lasocki, P. Martínez-Garzón, and G. Kwiatek, "Evolution of seismicity in relation to fluid injection in the north-western part of the geysers geothermal field," *Geophys. J. Int.*, vol. 212, no. 2, pp. 1157–1166, Feb. 2018.

[80] H. Guo and C. Thurber, "Double-difference seismic attenuation tomography method and its application to the geysers geothermal field, California," *Geophys. J. Int.*, vol. 225, no. 1, pp. 926–949, Dec. 2020.

[81] R. Gritto, S.-H. Yoo, and S. P. Jarpe, "Three-dimensional seismic tomography at the geysers geothermal field," in *Proc. 38th Workshop Geothermal Reservoir Eng.*, 2013, pp. 1–12.

[82] T. H. W. Goebel and E. E. Brodsky, "The spatial footprint of injection wells in a global compilation of induced Earthquake sequences," *Science*, vol. 361, no. 6405, pp. 899–904, Aug. 2018.

[83] S. A. Shapiro, P. Audigane, and J.-J. Royer, "Large-scale in situ permeability tensor of rocks from induced microseismicity," *Geophys. J. Int.*, vol. 137, no. 1, pp. 207–213, Apr. 1999.

[84] M. Parotidis and S. A. Shapiro, "A statistical model for the seismicity rate of fluid-injection-induced Earthquakes," *Geophys. Res. Lett.*, vol. 31, no. 17, pp. 1–5, Sep. 2004.

[85] L. Malagnini and I. Munafò, "On the relationship between ML and mw in a broad range: An example from the apennines, Italy," *Bull. Seismological Soc. Amer.*, vol. 108, no. 2, pp. 1018–1024, Jan. 2018.

[86] P. Bormann and D. Di Giacomo, "The moment magnitude m w and the energy magnitude m e: Common roots and differences," *J. Seismol.*, vol. 15, no. 2, pp. 411–427, Apr. 2011.

[87] J. D. Eshelby, "The determination of the elastic field of an ellipsoidal inclusion, and related problems," *Proc. Roy. Soc. London. Ser. A. Math. Phys. Sci.*, vol. 241, no. 1226, pp. 376–396, 1957.

[88] T. Lay and T. C. Wallace, *Modern Global Seismol.* Amsterdam, The Netherlands: Elsevier, 1995.

[89] M. Picozzi, A. Oth, S. Parolai, D. Bindi, G. De Landro, and O. Amoroso, "Accurate estimation of seismic source parameters of induced seismicity by a combined approach of generalized inversion and genetic algorithm: Application to the geysers geothermal area, California," *J. Geophys. Res. Solid Earth*, vol. 122, no. 5, pp. 3916–3933, May 2017.

[90] V. Longobardi, S. Nazeri, S. Colombelli, R. Rea, G. De Landro, and A. Zollo, "Time domain source parameter estimation of natural and man-induced microEarthquakes at the geysers geothermal field," *Energies*, vol. 16, no. 3, p. 1121, Jan. 2023.

[91] C. Aiken, X. Meng, and J. Hardebeck, "Testing for the-predictability' of dynamically triggered Earthquakes in the geysers geothermal field," *Earth Planet. Sci. Lett.*, vol. 486, pp. 129–140, Mar. 2018.

[92] S. Bentz, P. Martínez-Garzón, G. Kwiatek, G. Dresen, and M. Bohnhoff, "Analysis of microseismicity framing ML > 2.5 Earthquakes at the geysers geothermal field, California," *J. Geophys. Res. Solid Earth*, vol. 124, no. 8, pp. 8823–8843, Aug. 2019.

[93] K. Im, J.-P. Avouac, E. R. Heimisson, and D. Elsworth, "Ridgecrest aftershocks at coso suppressed by thermal destressing," *Nature*, vol. 595, no. 7865, pp. 70–74, Jul. 2021.

Mauro Palo received the Ph.D. degree in physics from the University of Salerno, Salerno, Italy, in 2008.

He is a Senior Assistant Professor with the Department of Physics "Ettore Pancini," University of Naples "Federico II," Naples, Italy, where he teaches electromagnetism at the Faculty of Management Engineering. Since 2004, he has been working on unsupervised machine-learning methods combined with geophysical techniques to study the spatiotemporal properties of seismicity in tectonic and volcanic regions.

Emanuele Ogliari (Member, IEEE) was born in Milan, Italy, in 1980. He received the M.Sc. and Ph.D. degrees in electrical engineering from the Politecnico di Milano, Milan, in 2016.

He is an Associate Professor with the Department of Energy, Politecnico di Milano, where he teaches electrical engineering and photovoltaic-based systems. He is a Staff Member of the Solar Tech Laboratory (SolarTech^LAB) and the Laboratory of MicroGrids (MG²LAB) with Politecnico di Milano.

Maciej Sakwa (Graduate Student Member, IEEE) received the M.Sc. degree in energy engineering from Politecnico di Milano, Milan, Italy, in 2022.

He is a Research Fellow with the Department of Energy, Politecnico di Milano. His research interests include renewable energy sources and the implementation of modern computational intelligence algorithms.