Article

Deep learning forecasts the spatiotemporal evolution of fluid-induced microearthquakes

Check for updates

Jaehong Chung [1], Michael Manga[2], Timothy Kneafsey[3], Tapan Mukerji[1,4] & Mengsu Hu [3] ✉

Microearthquakes generated by subsurface fluid injection record the evolving stress state and permeability of reservoirs. Forecasting their spatiotemporal evolution is therefore critical for applications such as enhanced geothermal systems, carbon dioxide sequestration and other geoengineering applications. Here we propose a transformer neural network model that ingests hydraulic stimulation history and prior microearthquake observations to forecast four key quantities: cumulative microearthquake count, cumulative logarithmic seismic moment, and the 50th- and 95th-percentile extents of the microearthquake cloud. Applied to the EGS Collab Experiment 1 dataset, the model achieves $R^2 > 0.98$ for the 1-s forecast horizon and $R^2 > 0.88$ for the 15-s forecast horizon across all targets, and supplies uncertainty estimates through a learned standard deviation term. These accurate, uncertainty-quantified forecasts enable real-time inference of fracture propagation and permeability evolution, demonstrating the strong potential of deep-learning approaches to improve seismic-risk assessment and guide mitigation strategies in future fluid-injection operations.

Subsurface applications for climate mitigation and sustainability are essential to achieving the net-zero emissions target set by the Intergovernmental Panel on Climate Change for 2050[1]. Key geo-engineering strategies include the development of enhanced geothermal systems (EGS) for renewable energy generation and the geological storage of carbon dioxide (CO_2) to reduce atmospheric greenhouse gas concentrations. The U.S. Geological Survey (USGS) estimates that EGS could provide over 500 GWe of electricity in the western United States alone[2]. In addition, carbon dioxide sequestration has the potential to store at least 1000 $GtCO_2$ in saline aquifers, with further storage capacity available in depleted oil and gas reservoirs and coal formations[3,4]. Despite the immense potential to reduce greenhouse gases through these subsurface applications, a key challenge is the induced seismicity that can result from fluid injection operations[5–7]. Fluid injection perturbs in-situ stress fields in the subsurface, potentially leading to the reactivation of preexisting faults or the creation of new fractures, both potentially compromising the integrity of reservoirs[8]. Notable examples include the magnitude 5.7 2015 Prague and magnitude 5.8 2016 Pawnee earthquakes in Oklahoma after wastewater injection[9–12]—and a magnitude 3.9 earthquake following circulation tests for the EGS project in Vendenheim, France[13]. These events underscore the critical need for accurate forecasting of induced seismicity to ensure the safe implementation of subsurface technologies.

Accurately forecasting fluid-induced seismicity remains a challenge due to the complex interactions between geological, hydrological, and mechanical factors[5,14]. Traditional approaches rely on physics-based models to estimate induced seismicity by coupling fluid flow, mechanical deformation, and seismicity rates[15–18]. Although these models can capture intricate subsurface interactions, they face limitations in real-world applications. Challenges include uncertainties in fracture geometries, material heterogeneity, and in-situ stress conditions. Moreover, assumptions such as isotropic material properties or idealized fracture networks are often required, reducing predictive accuracy. High computational costs associated with three-dimensional modeling with complex fracture geometries further restrict their use in practical forecasting and operational decision-making[15,17]. As a result, discrepancies between modeled and observed seismicity frequently occur.

From a statistical perspective, the Epidemic-Type Aftershock Sequence (ETAS) model provides a forecasting approach for both natural and fluid-induced seismicity, based on the assumption that an earthquake can trigger clusters of aftershocks[19,20]. In particular, nonstationary ETAS models have effectively demonstrated their capability in detecting the impacts of fluid-induced seismicity by employing a nonstationary background rate[19,21–23]. This capability positions ETAS as a valuable tool for generating probabilistic earthquake forecasts. However, determining key parameters, including the

[1]Department of Geophysics, Stanford University, Stanford, CA, USA. [2]Department of Earth and Planetary Science, University of California Berkeley, Berkeley, CA, USA. [3]Energy Geoscience Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [4]Department of Energy Science and Engineering, Stanford University, Stanford, CA, USA. ✉e-mail: mengsuhu@lbl.gov
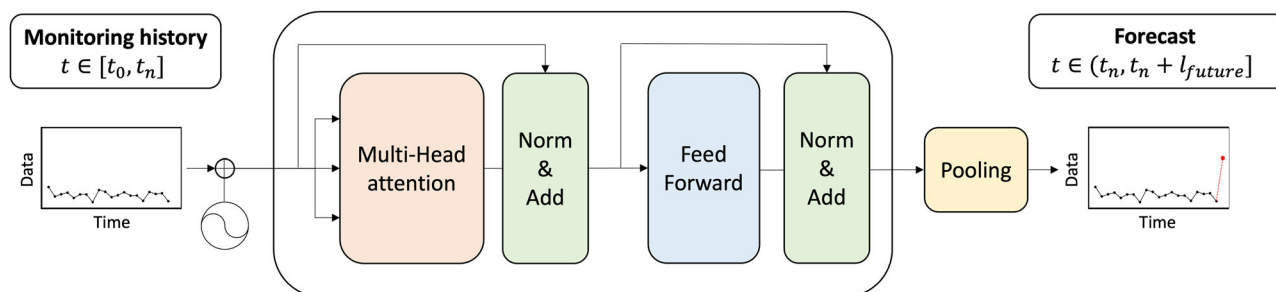
Fig. 1 | Architecture of the transformer-based MEQ forecasting model. Given input history from time steps t_0 through t_n, the model predicts MEQ features at future time steps t_{n+1} through $t_n + l_{\text{future}}$, where l_{future} is the forecast range (see section "Method: Transformer neural network architecture").

timing of peak activity, solely based on statistical analysis has been challenging[24]. Thus, successful applications of ETAS models to spatio-temporal forecasting of microearthquakes (MEQs) due to fluid injection may be limited.

Data-driven approaches—particularly machine learning—have emerged as powerful complements or alternatives to traditional frameworks including both physics-based and statistical approaches, in a range of geoscientific applications[25–37]. These methods do not require detailed prior knowledge of uncertain subsurface properties but instead leverage large datasets from monitoring systems to identify patterns and correlations that can be used for forecasting. For instance, deep learning—with and without physical constraints—was used to forecast the seismicity rate, which was then used to estimate the maximum magnitude of fluid-induced microseismicity[38]. A bidirectional long short-term memory neural network predicted fluid-induced permeability evolution based on MEQ features, including seismic rate and cumulative logarithmic seismic moment[39]. In addition, an LSTM model was employed to predict average permeability changes inferred from the seismicity data. Another LSTM model was used to predict pore pressure and associated fault displacements given the fluid injection cycles[40]. These studies demonstrate that deep learning approaches can effectively capture the temporal evolution of permeability or microseismicity based on operational parameters. However, they often focus solely on temporal predictions without considering the spatial evolution of MEQs, which is critical for assessing the extent of affected areas and potential impacts. Furthermore, these models rely on simplified assumptions for permeability changes, such as the migration of the triggering front of the MEQ cloud assuming proportionality to the square root of time since the initiation of injection, which is inconsistent with observed MEQ data[41]. These idealizations limit the applicability and accuracy of the models in complex scenarios.

Our study advances the forecasting of the spatiotemporal evolution of MEQs induced by hydraulic stimulation using a deep learning approach that tackles these challenges. Specifically, we employ transformer networks, a type of neural network architecture that uses self-attention mechanisms to capture complex dependencies within data sequences[42,43]. Compared with recurrent neural networks such as LSTMs, transformer networks can model long-range temporal dependencies more efficiently and are less susceptible to issues like vanishing gradients[44]. Their ability to focus on different parts of the input data through attention mechanisms makes them particularly well-suited for capturing both spatial and temporal patterns in MEQ data. Based on hydraulic stimulation history, our model predicts key MEQ features, including the cumulative number of MEQs, cumulative seismic moment, and the spatial extent of induced micro-seismicity. By incorporating both spatial and temporal information, the model provides more comprehensive forecasts that can inform real-time monitoring and risk mitigation strategies in subsurface activities.

Results

We use hydraulic stimulation data and MEQ history from the EGS Collab[45,46]. Figure 1 shows the architecture of our transformer model for forecasting the spatiotemporal evolution of MEQs based on hydraulic stimulation and MEQ histories (see Section " Method: Transformer neural network architecture ").

EGS Collab hydraulic stimulation datasets

We utilize hydraulic stimulation and MEQ data from the EGS Collab project, intermediate-scale (10–20 m) field tests at the Sanford Underground Research Facility in Lead, South Dakota. This study focuses on Experiment 1 data, aimed at producing a fracture network connecting an injection well to a production well via hydraulic fracturing[47]. A series of stimulations and flow tests were conducted at a depth of 1.5 km to re-open and generate hydraulic fractures in crystalline rock under reservoir-like stress conditions, with passive seismic data cataloged[48] and Continuous Active-Source Seismic Monitoring[45,49,50].

Figure 2 shows the stimulation-induced MEQs for each stimulation event along with the injection and production wells. Two 60 m boreholes were used for injection (E1-I) and production (E1-P), respectively. A total of five stimulation episodes were carried out in May 2018. During the first two stimulations, injections at flow rates less than 1L/min produced few MEQs. In addition, water leakage was observed between the production well and one monitoring well. Thus, the injection point was moved to a notch at a depth of 50 m in the injection hole (red triangle in Fig. 2) starting from Stimulation 3 and used through Stimulation 5. From Stimulations 3–5, three continuous hydraulic stimulations were performed using controlled step-rate injections to re-open or create fractures around the injection well, with the maximum injection rate reaching up to 5 L/min, resulting in rich MEQ signals[46,51]. Thus, this study uses data from Stimulations 3 to 5, generated from the same injection point with a rich MEQ history, to train neural networks. The data were recorded at 1-s intervals. Stimulations 3 and 4 each lasted approximately 1 h (3600 time steps), and the first 1 h and 10 min of Stimulation 5 were used (4100 time steps). These continuous records were segmented into overlapping input-output windows for supervised training, validation, and testing, as described in section "Data pre-processing: crop and normalization".

Figure 3 presents the series of stimulations along with the spatio-temporal MEQ data and corresponding magnitudes. Detailed information about the MEQs—including location, time, and magnitudes—was continuously monitored during the hydraulic stimulations[45,46]. In addition, to quantify the spatial extent of MEQs in response to fluid injection, we extracted the 95th and 50th percentiles (median) distances of the MEQ clouds from the injection points as a function of time. Although the monitoring array is extensive, the catalog still carries intrinsic uncertainties: hypocenter locations are accurate to about 1 m and there is no reported uncertainty range for magnitude[45]. These uncertainties limit the fidelity of the training data and establish a floor on achievable forecast accuracy. Additionally, including all raw events—without excluding those below the magnitude of completeness—could constrain the neural network's capability to learn underlying MEQ patterns (Supplementary Fig. 1).

Fig. 2 | Spatial distribution of fluid-induced microearthquakes during hydraulic stimulations. The figure shows the three-dimensional spatial distribution of microearthquakes (MEQs) generated during three hydraulic stimulation episodes. The solid black line represents the injection well, and the dashed black line represents the production well. A red triangle marks the injection point located at the 50 m notch in the injection well. Colored scatter points indicate MEQ locations: yellow for stimulation 3, green for stimulation 4, and purple for stimulation 5.
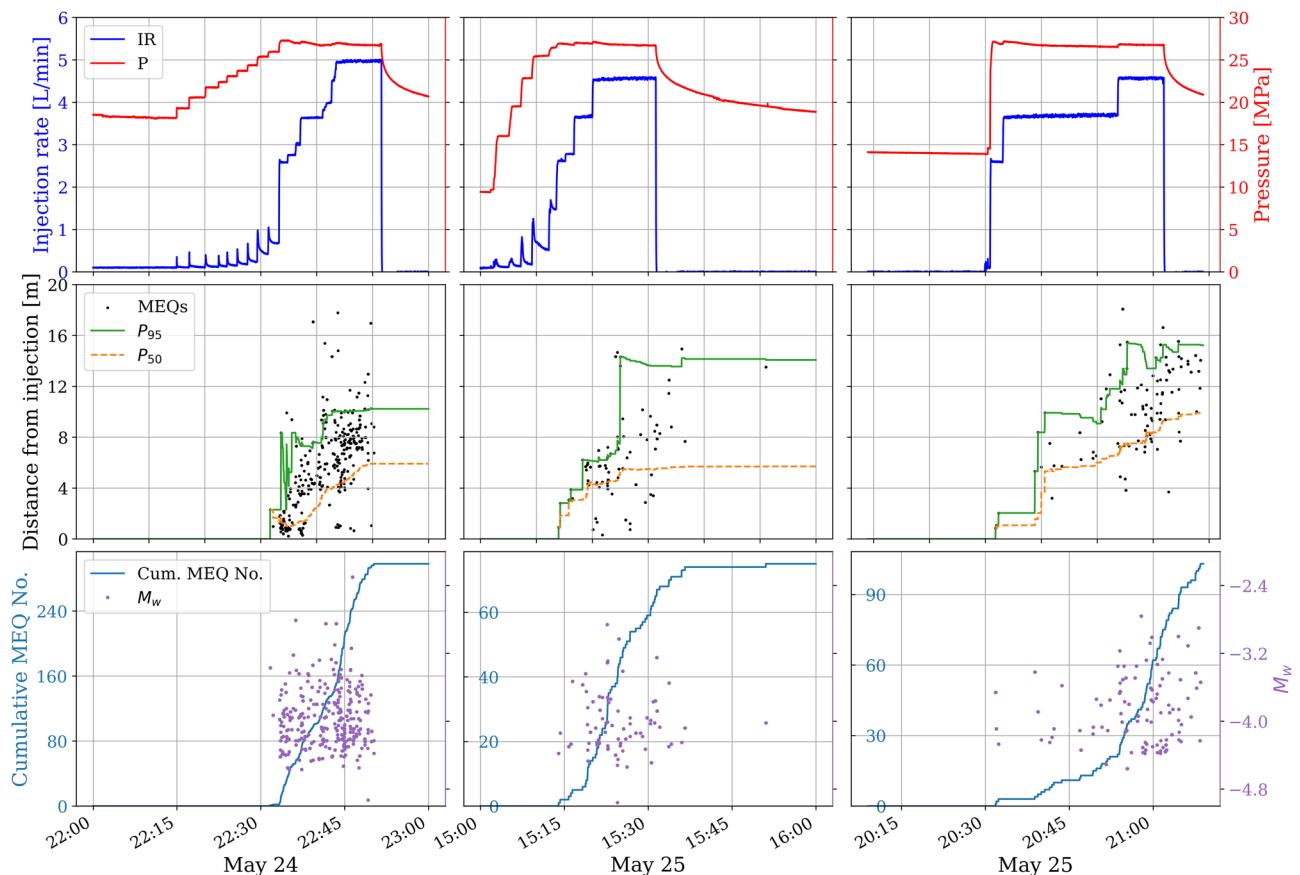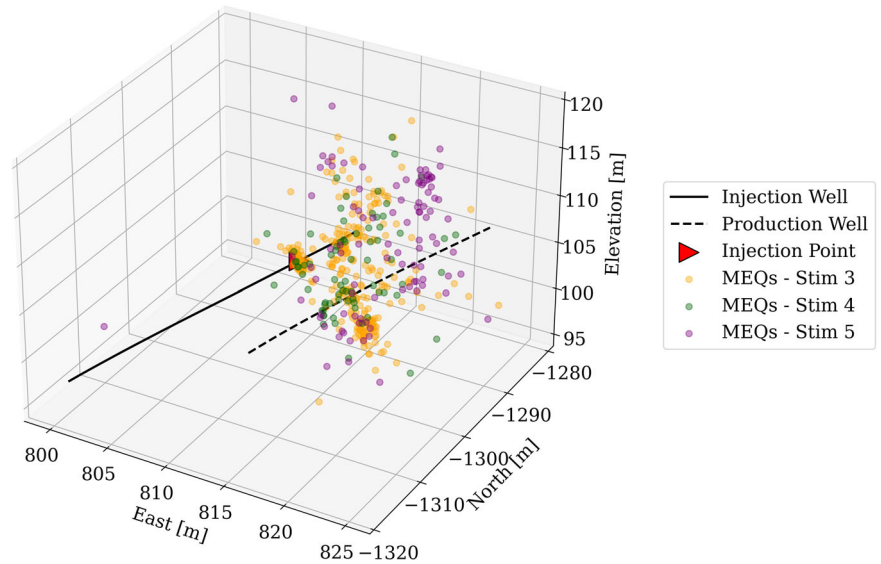

Fig. 3 | Microearthquake and injection history for EGS Collab Stimulation 3-5. Columns correspond to: Stimulation 3 (training data), Stimulation 4 (validation data), and Stimulation 5 (test data). The first row presents the hydraulic stimulation history, showing injection rate (blue) and injection pressure (red). The second row displays the locations of microearthquakes (MEQs) relative to the injection point, with distances calculated as the Euclidean distance between the injection point and observed microseismic events. P_{95} and P_{50} represent the 95th and 50th percentile distances over time. The third row shows the cumulative number of MEQ events and the magnitude of each discrete event.

Forecasting performance

We evaluate three forecast intervals—1 s, 15 s, and 30 s—using a sliding-window strategy. At each forecasting instant t_n, the model ingests the entire monitoring history $[t_0, t_n]$ and predicts subsequent interval $[t_{n+1}, t_n + l_{future}]$, where l_{future} is the forecast range (e.g., 1 s, 15 s, or 30 s). For instance, when using a 15 s range, the model forecasts the next 15 s (e.g., $t_{101} - t_{115}$) based on the history data $t_1 - t_{100}$. Once actual monitoring for these 15 s is recorded, these new data ($t_{101} - t_{115}$) are appended to the monitoring history. The model then uses the extended history $t_1 - t_{115}$ to forecast the following segment $t_{116} - t_{130}$, and this procedure repeats until the
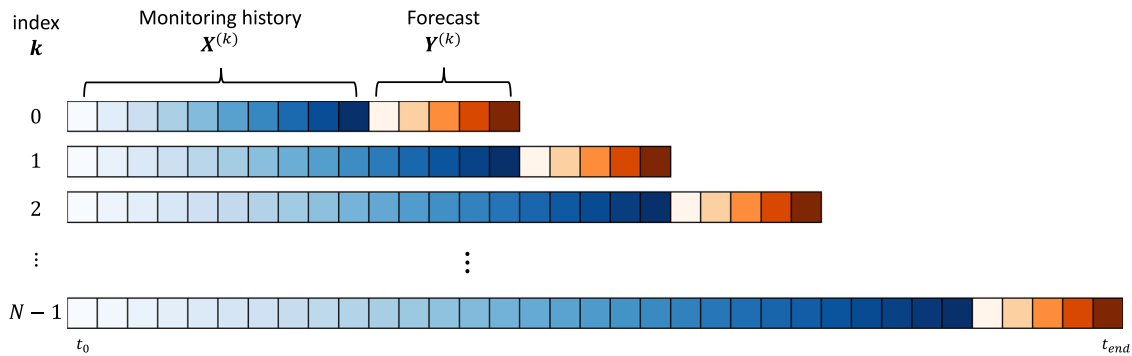
Fig. 4 | Schematic of the forecasting procedure. $X^{(k)}$ denotes the cumulative monitoring input and $Y^{(k)}$ the corresponding forecast window; k is the segment index. The forecasting range is l_{future} (see section "Method: Transformer neural network architecture").
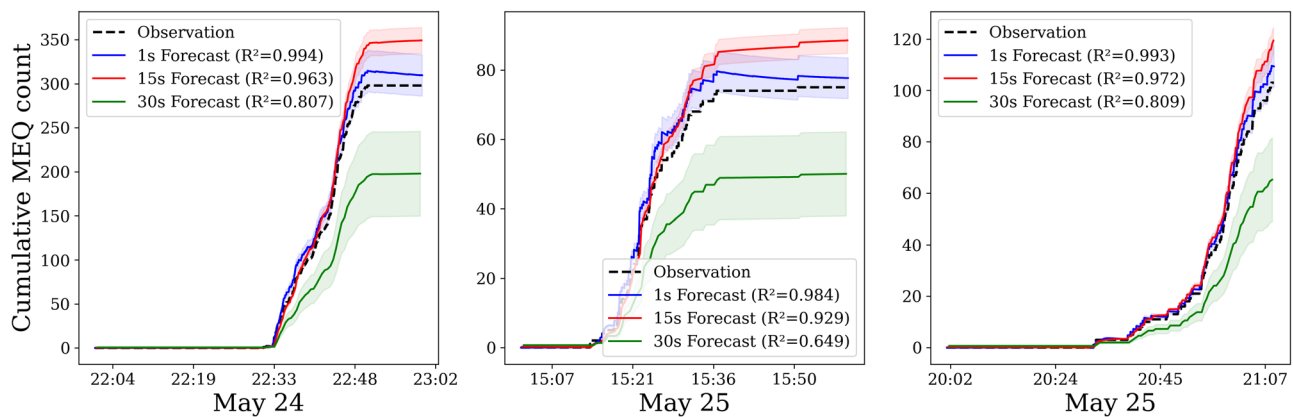

Fig. 5 | Cumulative MEQ counts: observed data (black dotted) versus forecasts for the 1-s (blue), 15-s (red), and 30-s (green) models. Each forecast curve is constructed by predicting successive, non-overlapping segments whose length equals the forecast interval and concatenating them to cover the full record. Panels show the training (left), validation (middle), and test (right) sets. Shaded bands denote $\pm\sigma$ (one predicted standard deviation, corresponding to $\approx 68\%$ coverage under a Gaussian assumption).

monitoring concludes. Since the model consistently utilizes actual measurements without recycling previously predicted outputs, forecasting errors do not accumulate over successive forecasts (Fig. 4).

Figure 5 compares the forecasted and observed cumulative MEQ counts. For the 1-second forecast model the predicted curves are virtually indistinguishable from the ground truth, even on unseen data (validation $R^2 = 0.999$, test $R^2 = 0.980$). The 15-second forecast model maintains high fidelity (validation $R^2 = 0.929$, test $R^2 = 0.972$), with a slight tendency to overestimate MEQ growth during the most intense injection phases. The 30-s forecast model still captures the overall trend but systematically underpredicts the MEQ count late in each episode (validation $R^2 = 0.649$, test $R^2 = 0.809$). These results show that the transformer delivers excellent short-term forecasts, with accuracy declining gradually as the forecast window lengthens.

Second, we forecast the cumulative logarithmic seismic moment, a proxy for the activated reservoir volume and thus a key metric for planning new production wells[52,53]. The cumulative moment $\mathcal{M}$ is defined as[39]

$$\mathcal{M}(t_i) = \int_{t_0}^{t_i} \log M_0 \, dt, \tag{1}$$

with

$$\log M_0 = 1.5\, M_w + 13.5, \tag{2}$$

where M_0 is the seismic moment, M_w the moment magnitude, t_0 the start of injection, and t_i the current injection time.

Figure 6 compares the predicted and observed cumulative moments for the 1-, 15-, and 30-s forecast models across the three data splits. The 1-s forecast model reproduces the observations almost exactly (validation $R^2 = 0.999$, test $R^2 = 0.978$). Performance remains high at 15-s forecast model (validation $R^2 = 0.878$, test $R^2 = 0.935$), although the predictive bands widen compared with the 1-s case. At 30-s forecast the model still captures the overall trend but underestimates the released seismic energy (validation $R^2 = 0.546$, test $R^2 = 0.765$). These results confirm that our neural network effectively links hydraulic-energy input to seismic-energy release, providing reliable short-term estimates of cumulative moment while showing a gradual and interpretable loss of accuracy as the forecast range increases.

Accurately forecasting the spatial evolution of MEQ clouds is critical for delineating the affected area, guiding mitigation, and optimizing future well placement[15]. Figure 7 compares the spatial extent of the MEQ clouds across the training, validation, and test sets, quantified by the 50th and 95th percentiles of the Euclidean distance from the injection point. The 1-s and 15-s forecast models reproduce the ground truth trajectories of both the median distance (P_{50}) and the far distance (P_{95}), achieving $R^2 > 0.97$ for the 1-s forecast model and $R^2 > 0.94$ for the 15-s forecast model.

Figure 8 illustrates the final stabilized extents predicted by these models: absolute errors are below 0.4 m for the 1-s model and below 2 m for the 15-s model (Table 1). For the 1-second case, the observed-predicted differences lie within the model's $\pm\sigma$ band, indicating that the discrepancies are consistent with the reported uncertainty. In contrast, the 15-s differences exceed σ, revealing the limitations of the mid-range model. The 30-s model underestimates both P_{50} and P_{95} in all data splits, highlighting its reduced reliability for long-range spatial forecasts.
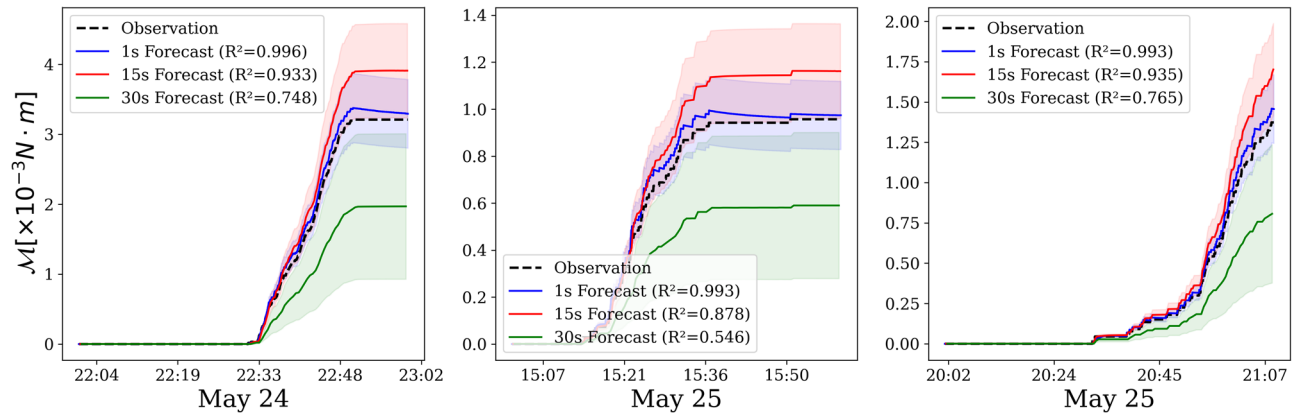
Fig. 6 | Cumulative logarithmic seismic moment: observed data (black dotted) versus forecasts from the 1-, 15-, and 30-s models (blue, red, green) for the training (left), validation (middle), and test (right) sets. Shaded bands denote $\pm\sigma$.
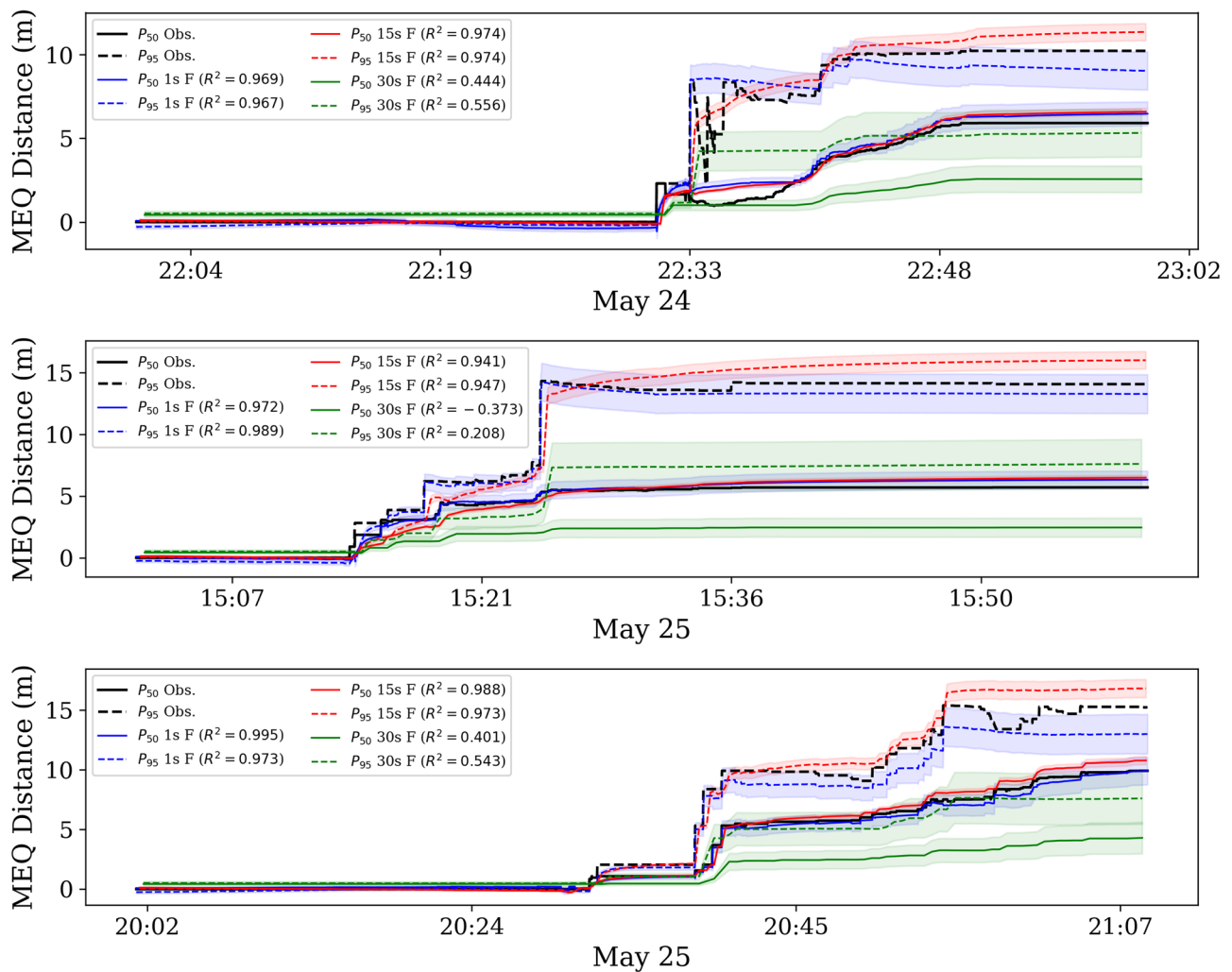

Fig. 7 | Temporal evolution of the MEQ cloud's spatial extent. The three rows correspond to the training (top), validation (middle), and test (bottom) datasets. In each row, the solid curve shows the observed 50th-percentile distance (P_{50}) and the dashed curve the observed 95th-percentile distance (P_{95}). Forecasts from the 1-, 15-, and 30-s models are plotted in blue, red, and green, respectively. Shaded regions denote $\pm\sigma$ (standard deviation).

Discussion and conclusion

Our transformer model accurately forecasts fluid-induced MEQs, capturing both their temporal evolution and spatial growth (Table 2). This dual capability is, to the best of our knowledge, novel; earlier studies focused mainly on temporal predictions[39,54]. Reliable spatiotemporal forecasts are essential for estimating permeability changes and mitigating the risks associated with induced seismicity. In the following, we discuss how permeability enhancement can be inferred from monitoring data and model outputs, how fracture characteristics can be estimated, and the potentials and limitations of deep-learning-based forecasting for field-scale, fluid-induced earthquakes.
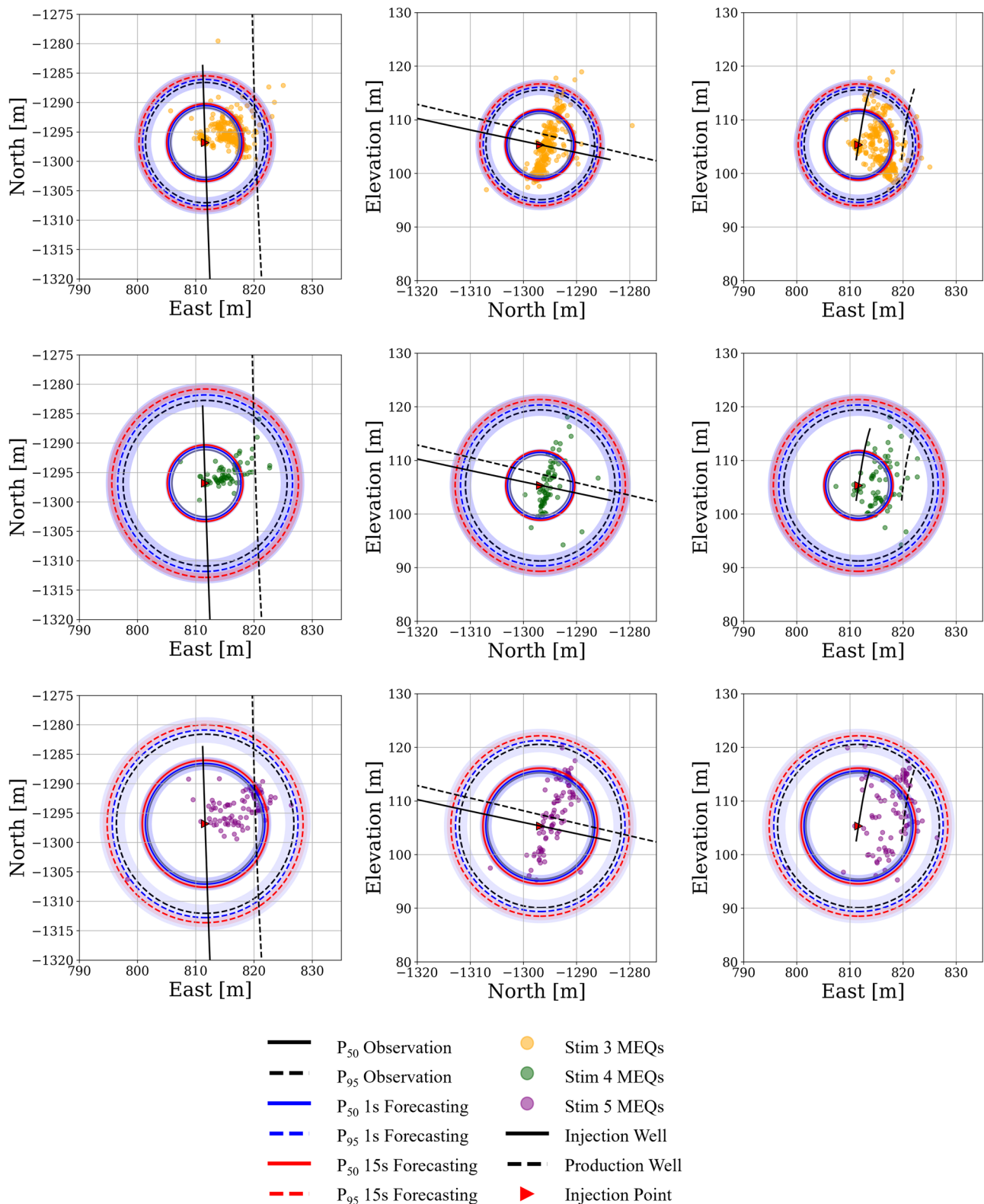
Fig. 8 | Spatial evolution of microearthquake (MEQ) clouds and forecast performance. The figure shows the spatial distribution of MEQs and forecast results for different datasets and time horizons. Each row represents a dataset: training (top), validation (middle), and test (bottom). Each column shows projections on the XY, YZ, and ZX planes. Solid circles indicate the observed 50th-percentile radius (P_{50}), while dashed circles represent the 95th-percentile radius (P_{95}). Blue and red lines show forecasts from the 1-s and 15-s models, respectively, with shaded regions denoting $\pm\sigma$ (one predicted standard deviation). Colored dots represent MEQs from different stimulation phases: yellow for Stim 3, green for Stim 4, and purple for Stim 5. Solid black lines indicate the injection well, dashed black lines the production well, and red triangles mark the injection point. All spatial dimensions are in meters.

Table 1 | Final MEQ spatial extent

Dataset	Pct	Obs [m]	1 s model			15 s model			30 s model		
			Pred [m]	Diff [m]	σ [m]	Pred [m]	Diff [m]	σ [m]	Pred [m]	Diff [m]	σ [m]
Train	P_{50}	5.92	6.29	0.38	0.67	6.59	0.68	0.20	2.56	3.35	0.79
	P_{95}	10.23	10.74	0.51	1.46	11.36	1.13	0.51	5.32	4.91	1.43
Val	P_{50}	5.71	6.16	0.46	0.65	6.50	0.79	0.20	2.45	3.25	0.77
	P_{95}	14.08	15.01	0.93	2.04	16.02	1.94	0.71	7.61	6.47	1.99
Test	P_{50}	9.92	10.25	0.34	1.11	10.79	0.87	0.34	4.29	5.62	1.32
	P_{95}	15.23	15.95	0.72	2.18	16.82	1.58	0.76	7.60	7.63	2.16

For each split and percentile, the observed radius is followed by the neural-network (NN) prediction (Pred), absolute difference |Obs − Pred| (Diff), and model uncertainty (σ) for the 1-, 15-, and 30-s forecasting ranges.

Table 2 | R^2 scores for all metrics and forecast models

Dataset	Cum. MEQ count			Cum. moment ($\mathcal{M}$)			50th-percentile distance (P_{50})			95th-percentile distance (P_{95})		
	1 s	15 s	30 s	1 s	15 s	30 s	1 s	15 s	30 s	1 s	15 s	30 s
Train	0.994	0.963	0.807	0.996	0.933	0.748	0.969	0.974	0.444	0.967	0.974	0.556
Val	0.984	0.929	0.649	0.993	0.878	0.546	0.972	0.941	−0.373	0.989	0.947	0.208
Test	0.993	0.972	0.809	0.993	0.935	0.765	0.995	0.988	0.401	0.973	0.973	0.543

The 1 s, 15 s, and 30 s columns correspond to models with forecast ranges of 1, 15, and 30 s, respectively.

Estimation of permeability enhancement

Estimating permeability enhancement is a critical task in EGS, yet direct measurements are challenging in the subsurface. This limitation also applies to our study—we aim to understand how permeability evolves during hydraulic stimulation, but no direct measurements were available from the field experiment. Although the correlation between MEQs and permeability remains elusive[55], we derive a physically grounded rationale to indirectly estimate permeability using model outputs. Specifically, we apply the cubic law for permeability, which relates changes in fracture aperture to permeability change[56,57]:

$$\Delta k = \frac{(b_0 + \Delta b)^3}{12s} - \frac{b_0^3}{12s} \tag{3}$$

where Δk is the permeability change, b_0 is the initial fracture aperture, Δb is the aperture change, and s is the spacing between parallel fractures. Assuming that the initial aperture b_0 is negligible compared to the aperture change (i.e., $b_0 \ll \Delta b$), we approximate the permeability evolution as $\Delta k \approx \frac{\Delta b^3}{12s}$. Given that the EGS Collab Experiment 1 aimed to establish fracture networks via hydraulic fracturing (i.e., tensile fractures)[55,58], we assume the seismic moment is linked to normal displacement by tensile opening. The equivalent moment M_0 for a tensile opening can be expressed as[39]:

$$M_0 = 2GA\Delta u_n \tag{4}$$

where G is the shear modulus, A is the area of the fracture, and Δu_n is the normal displacement across the fracture. Assuming the area A of the fracture is proportional to the aperture ($A \propto \Delta b$)[59], we establish a direct proportionality between seismic moment M_0 and permeability change as[60]:

$$\log M_0 \propto \frac{2}{3} \log \Delta k \tag{5}$$

With these scaling relationships, we infer that the overall logarithmic permeability increment is linearly proportional to the logarithmic seismic moment, though this assumption primarily holds during early stimulation, where the initial aperture is substantially smaller than the aperture increment (i.e., $b_0 \ll \Delta b$).

During the first stimulation, the observed cumulative logarithmic seismic moment reaches ≈3 (Fig. 6 left), implying a permeability increase of roughly two orders of magnitude. The 1-s forecast reproduces this estimate, whereas the 15-s forecast model overpredicts the moment by about one order of magnitude, and the 30-s forecast model underpredicts it by a similar amount. Because the cumulative seismic moment predicted by our network can be mapped directly to permeability changes, the model provides a practical, indirect means of tracking permeability evolution during hydraulic stimulation—though this mapping is valid only for the initial seismic—moment range where the derivation's assumptions hold.

Inference of the fracture characteristics

In fluid injection operations, we need to control the spatial extent of fracturing. As an example, in EGS fields, it is crucial to prevent MEQs from extending beyond the region between injection and production wells while enhancing permeability within this region through fracturing. Our model provides estimates of two spatial extents of MEQs: the 95th percentile distance (P_{95}) and the 50th percentile distance (P_{50}). P_{95} represents the far extent of MEQs, while P_{50} indicates the most active MEQ regions, which likely correspond to areas of greatest permeability increase due to fracture generation and re-opening.

The importance of tracking P_{95} and P_{50} becomes clear when the spatial extents from each stimulation are compared (Table 1). From stimulation 3 (training) to stimulation 4 (validation), the observed P_{95} grows by 3.85 m (from 10.23 to 14.08 m), while P_{50} retreats by 0.21 m (from 5.92 to 5.71 m), indicating a slight shrinkage of the seismically active zone. Our 1-s forecast model reproduces these shifts almost exactly, predicting a 4.27 m increase in P_{95} (from 10.74 to 15.01 m) and 0.13 m retreat in P_{50} (from 6.29 to 6.16 m); all absolute errors fall within the 1-s forecast model's $\pm\sigma$ band. Between stimulation 4 (validation) and stimulation 5 (test), the observed P_{95} increases by 1.15 m (from 14.08 to 15.23 m), whereas P_{50} advanced by 4.21 m (from 5.71 to 9.92 m). The 1-s forecast model again captures these trends, predicting a 0.94 m rise in P_{95} (from 15.01 to 15.95 m) and 4.09 m increase in P_{50} (from 6.16 to 10.25 m). By accurately forecasting P_{50} and P_{95} in real time, the network enables practitioners to infer fracture propagation and activation, making it a practical tool for managing stimulation where direct measurements are not feasible.

Potential and challenges of deep learning forecasting

Among the various deep learning approaches, we chose the transformer model as our core architecture. The success of the transformer model is driven by several key factors. First, the self-attention mechanism allows the model to capture long-term dependencies[42,61,62], which are crucial in fluid-induced seismicity, where MEQs are influenced by cumulative fluid injection, pore pressure changes, and perturbed in-situ stress conditions[2]. In particular, fluid-induced seismicity often exhibits long time intervals between injection and seismicity. For instance, the largest earthquake (local magnitude 3.9) at the deep geothermal site GEOVEN in Vendenheim occurred more than six months after shut-in[63]. The self-attention mechanism enables the model to weigh the importance of different input features over time, making it highly suited for sequential data[44].

Second, transformers excel at processing spatiotemporal data[64], which is vital for accurately predicting the spatial distribution of MEQs. This ability provides critical insights into fracture propagation[65] and fluid migration[66], both of which are key factors in assessing the effectiveness of hydraulic stimulation. The model's performance in predicting the spatial extent of seismic events reflects its capacity to capture both the temporal and spatial dynamics of fluid injection-induced microseismicity. Third, the transformer's non-recurrent architecture allows it to handle irregular time series data[67], a common occurrence in microseismic monitoring due to variable injection schedules and operational pauses. This flexibility enhances the model's robustness across different stimulation phases and geological settings, making it adaptable to varying conditions and data availability—a common challenge in real-world geophysical applications.

While the model shows promising results, extending it to large-scale field operations introduces additional uncertainties due to unknown geological heterogeneity and the extended temporal dependencies inherent to fluid-induced seismicity. The data used in this study were collected from an intermediate-scale (10–20 m) experiment with comprehensive monitoring tools from the EGS Collab project[47,50]. Such dense instrumentation may not be feasible in reservoir-scale engineering applications, raising questions about the model's generalizability to less controlled, large-scale environments. One promising strategy for adapting deep learning forecasting techniques to larger-scale fluid-induced seismicity applications involves transfer learning with fine-tuning. For example, successful transferability between datasets from Utah FORGE and EGS Collab was recently demonstrated using appropriate fine-tuning methods[39]. Although further fine-tuning will likely be required to adjust the model to larger operational scales, the fundamental assumption remains that the neural network model learns generalizable signal patterns associated with fluid-induced MEQs. Additionally, integrating uncertainty quantification into predictions becomes increasingly important given the higher uncertainty inherent in real-field-scale operations. By incorporating these strategies, along with judicious monitoring, transformer networks could be systematically validated and effectively implemented at larger scales. Future work could involve training and validating the model's performance with field-scale fluid-induced seismic data and hydraulic stimulation histories, thus ensuring robustness in more complex geological settings.

In summary, despite limitations related to monitoring systems and scale, this study presents a deep learning based approach for forecasting MEQs in response to fluid injection. The transformer model's ability to predict both temporal and spatial evolution highlights its potential as a valuable tool in subsurface operations, offering substantial improvements in safety and efficiency.

Method: transformer neural network architecture

We employ a transformer neural network to forecast the spatiotemporal evolution of fluid-induced microearthquakes (MEQs). The attention mechanism captures dependencies in the monitoring time series, allowing the model to learn patterns across multiple temporal scales. Figure 1 illustrates the overall architecture. Given a sequence of past monitoring data, the model predicts the future MEQ features. The following subsections describe data processing, network architecture, loss function, and hyperparameter tuning.

Data preprocessing: crop and normalization

We first construct training segments by sliding a growing stimulation history across the cumulative time series and advancing the forecast horizon in non-overlapping blocks. The monitoring data at discrete time index t are defined as:

$$\boldsymbol{x}(t) = \left[x_1(t), x_2(t), ..., x_6(t) \right]^T \in \mathbb{R}^M, \tag{6}$$

where the monitoring dimension $M = 6$ includes hydraulic stimulation features —(1) flow rate (x_1) and (2) well head pressure (x_2)— and spatiotemporal MEQ features —(3) cumulative MEQ numbers (x_3), (4) $\log M_0$ (x_4), (5) 95th percentile distance (x_5), (6) 50th percentile distance)(x_6).

The cropping procedure is controlled by two hyperparameters. The minimum history length $l_{\min}$ specifies the number of monitoring samples always available, and the forecast horizon l_{future} specifies how many future steps are predicted at once. For a monitoring ending at t_{end}, the number of segments is

$$N = \frac{t_{end} - l_{\min}}{l_{\text{future}}} \tag{7}$$

For each segment index $k \in \{0, \ldots, N-1\}$ the split time is set as

$$t_k^{\text{split}} = l_{\min} + k l_{\text{future}} \tag{8}$$

Thus, the cumulative monitoring input ($\boldsymbol{X}^{(k)}$) and the subsequent forecast window ($\boldsymbol{Y}^{(k)}$) are defined as:

$$\boldsymbol{X}^{(k)} = \{\boldsymbol{x}(t) \mid 1 \le t \le t_k^{\text{split}}\} \in \mathbb{R}^{t_k^{\text{split}} \times M} \tag{9}$$

$$\boldsymbol{Y}^{(k)} = \{\boldsymbol{x}(t) \mid t_k^{\text{split}} < t \le t_k^{\text{split}} + l_{\text{future}}\} \in \mathbb{R}^{t_k^{\text{future}} \times F} \tag{10}$$

where $F = 4$ corresponds to the forecasting MEQ features: (1) cumulative MEQ count, (2) $\log M_0$, (3) P_{95}, and (4) P_{50}. Each successive segment index k advances the split by l_{future}, ensuring that the predicted time blocks $\boldsymbol{Y}^{(k)}$ are non-overlap and contiguous, while the input window grows monotonically. This approach yields continuous, leakage-free forecasting segments that can be applied in real time once at least $l_{\min}$ monitoring have been acquired (Fig. 4).

To fairly normalize the data without information leakage from future steps, normalization is applied individually to each input window $\boldsymbol{X}^{(k)}$. For each monitoring dimension $m \in \{1, \ldots, M\}$ and each segment k, we define the normalization using only the known input window as follows:

$$\tilde{\boldsymbol{x}}_m^{(k)} = \frac{\boldsymbol{x}_m^{(k)} - \min\limits_{1 \le t \le t_k^{\text{split}}} x_m(t)}{\max\limits_{1 \le t \le t_k^{\text{split}}} x_m(t) - \min\limits_{1 \le t \le t_k^{\text{split}}} x_m(t)}, \quad 1 \le t \le t_k^{\text{split}} \tag{11}$$

The normalization parameters obtained from each input window $\boldsymbol{X}^{(k)}$ are then consistently applied to scale the corresponding forecast window $\boldsymbol{Y}^{(k)}$. This ensures that normalization relies exclusively on information available at the prediction time, thus avoiding any data leakage from future observations.

Neural network architecture

Our transformer neural network architecture employs a multi-head attention mechanism designed to effectively capture temporal dependencies from variable-length sequences. Given an input monitoring sequence $\boldsymbol{X}^{(k)}$, the multi-head attention layer processes the input as follows[42]:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}, \tag{12}$$

where $\mathbf{Q} = \mathbf{X}^{(k)}\mathbf{W}_Q$, $\mathbf{K} = \mathbf{X}^{(k)}\mathbf{W}_K$, and $\mathbf{V} = \mathbf{X}^{(k)}\mathbf{W}_V$ are the query, key, and value matrices, respectively; $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ are learnable weight matrices; d_k is the dimension of key vectors.

Following the attention layer, a feed-forward network (FFN)[68] is applied independently to each time step. The FFN consists of two linear transformations with a Rectified Linear Unit (ReLU) activation function:

$$\text{FFN}(\mathbf{z}) = \text{ReLU}(\mathbf{z}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2, \quad (13)$$

where $\mathbf{z}$ denotes the input from the attention output, and $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{b}_1$, and $\mathbf{b}_2$ are learnable parameters.

To enhance training stability, layer normalization and residual connections are applied after both attention and feed-forward layers. These ensure effective gradient propagation and prevent training instabilities.

After attention and feed-forward layers, global average pooling and dense layers reduce the sequence to a single vector, producing predictions for the forecasting window $\mathbf{Y}^{(k)}$. In particular, the model predicts both the mean (μ) and log-variance ($\log \sigma^2$) of these forecasting MEQ features to quantify prediction uncertainty:

$$\mathbf{y}_{\text{pred}} \in \mathbb{R}^{l_{\text{future}} \times 2F}, \quad \mathbf{y}_{\text{pred}}(t) = [\mu_1(t), ..., \mu_F(t), \log \sigma_1^2(t), ..., \log \sigma_F^2(t)] \quad (14)$$

The model is trained using the Adam optimizer[69] with a heteroscedastic Gaussian negative log-likelihood (NLL) loss function[70,71], augmented by a monotonicity penalty weighted by the hyperparameter (λ):

$$\mathcal{L} = \text{NLL}(\mathbf{y}_{\text{true}}, \mathbf{y}_{\text{pred}}) + \lambda \, \text{Penalty}_{\text{mono}} \quad (15)$$

The NLL explicitly measures the discrepancy between predictions and true values, accounting for predictive uncertainty. Given the predicted mean (μ_{pred} and log-variance ($\log \sigma^2_{\text{pred}}$), the NLL is defined as:

$$\text{NLL}(\mathbf{y}_{\text{true}}, \mathbf{y}_{\text{pred}}) = \frac{1}{2NF} \sum_{i=1}^{N} \sum_{f=1}^{F} \left[\frac{(y_{i,f}^{\text{true}} - \mu_{i,f}^{\text{pred}})^2}{\sigma_{i,f}^2} + \alpha \log\left(\sigma_{i,f}^2\right) \right], \quad (16)$$

where N is the number of time steps in the forecast window, F is the number of MEQ target features, and α is the hyperparameter to discourage the model from inflating variance. This formulation captures both prediction accuracy and model confidence, penalizing over- or under-confident forecasts.

To enforce non-decrease for the cumulative term forecastings, a monotonicity penalty is applied to cumulative MEQ count and cumulative logarithmic seismic moment. The penalty is defined as:

$$\text{Penalty}_{\text{mono}} = \sum_{t=2}^{T} \left| \min\left(0, \, \text{pred} \, t\mathbf{y} - \mathbf{y}_{\text{pred}}^{t-1}\right) \right|, \quad (17)$$

where only the selected cumulative features are included in the penalty term.

Finally, all predictions are rescaled using the inverse of the normalization applied during preprocessing. The model performance is evaluated using the coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (18)$$

where Y includes the four spatiotemporal MEQ features.

Neural-network hyper-parameter tuning
The transformer model is trained to forecast spatiotemporal MEQs from hydraulic-stimulation history and past MEQ responses. While network weights are learned automatically, several settings—loss-function coefficients, architectural widths, batch size, dropout rate, and penalty weights—

must be chosen by the user[72,73]. Supplementary Table 1 lists the values that remain fixed in every experiment.

Two coefficients are tuned by grid search: β (the variance-regularization weight inside the heteroscedastic Gaussian NLL term) and λ (the weight on the monotonic-increase penalty applied to cumulative MEQ count and cumulative seismic moment). For each forecast horizon $l_{\text{future}} \in \{1, 15, 30\}$ models are trained with $\beta, \lambda \in \{0.1, 1.0, 10.0\}$. Validation R^2 scores identify the optimal pair ($\beta^\star, \lambda^\star$); the corresponding results appear in Supplementary Table 2.

Short-horizon models—forecast windows of up to fifteen seconds—achieve excellent accuracy; for example, the $l_{\text{future}} = 15$ model reaches $R^2_{\text{val}} = 0.924$. As the horizon lengthens, performance degrades: at $n_{\text{future}} = 30$ the best model attains $R^2_{\text{val}} = 0.046$. The horizon-specific models reported in Supplementary Table 2 are used for all subsequent experiments.

Data availability
The EGS Collab experiment's stimulation data and seismic catalog are available at https://doi.org/10.15121/1651116 and https://doi.org/10.15121/1557417.

Code availability
The code used in this study is available on GitHub at https://github.com/jh-chung1/Transformer_MEQ_Forecasting.

References
1. Metz, B., Davidson, O., De Coninck, H., Loos, M. & Meyer, L. *IPCC Special Report on Carbon Dioxide Capture and Storage* (Cambridge University Press, 2005).
2. Williams, C. F., Reed, M. J., Mariner, R. H., DeAngelo, J. & Galanis, S. P. *Assessment of Moderate-and High-Temperature Geothermal Resources of the United States*. Technical Report (Geological Survey (US), 2008).
3. Bachu, S. & Adams, J. J. Sequestration of CO_2 in geological media in response to climate change: capacity of deep saline aquifers to sequester CO_2 in solution. *Energy Convers. Manag.* **44**, 3151–3175 (2003).
4. Damen, K., Faaij, A. & Turkenburg, W. Health, safety and environmental risks of underground CO_2 storage–overview of mechanisms and current knowledge. *Clim. Change* **74**, 289–318 (2006).
5. Rutqvist, J. The geomechanics of CO_2 storage in deep sedimentary formations. *Geotech. Geol. Eng.* **30**, 525–551 (2012).
6. Yeo, I., Brown, M. R., Ge, S. & Lee, K. Causal mechanism of injection-induced earthquakes through the M_w 5.5 Pohang earthquake case study. *Nat. Commun.* **11**, 2614 (2020).
7. Ellsworth, W. L., Giardini, D., Townend, J., Ge, S. & Shimamoto, T. Triggering of the Pohang, Korea, earthquake M_w 5.5 by enhanced geothermal system stimulation. *Seismol. Res. Lett.* **90**, 1844–1858 (2019).
8. Wang, C.-Y., Manga, M., Shirzaei, M., Weingarten, M. & Wang, L.-P. Induced seismicity in Oklahoma affects shallow groundwater. *Seismol. Res. Lett.* **88**, 956–962 (2017).
9. Rajesh, R. & Gupta, H. K. Characterization of injection-induced seismicity at north central Oklahoma, USA. *J. Seismol.* **25**, 327–337 (2021).
10. Johann, L., Shapiro, S. A. & Dinske, C. The surge of earthquakes in Central Oklahoma has features of reservoir-induced seismicity. *Sci. Rep.* **8**, 11505 (2018).
11. Hincks, T., Aspinall, W., Cooke, R. & Gernon, T. Oklahoma's induced seismicity strongly linked to wastewater injection depth. *Science* **359**, 1251–1255 (2018).

12. Manga, M., Wang, C.-Y. & Shirzaei, M. Increased stream discharge after the 3 September 2016 M_w 5.8 Pawnee, Oklahoma earthquake. *Geophys. Res. Lett.* **43**, 11–588 (2016).

13. Fiori, R., Vergne, J., Schmittbuhl, J. & Zigone, D. Monitoring induced microseismicity in an urban context using very small seismic arrays: the case study of the Vendenheim EGS project. *Geophysics* **88**, WB71–WB87 (2023).

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

15. Boyet, A., Vilarrasa, V., Rutqvist, J. & De Simone, S. Forecasting fluid-injection induced seismicity to choose the best injection strategy for safety and efficiency. *Philos. Trans. A* **382**, 20230179 (2024).

16. Lu, J. & Ghassemi, A. Coupled thermo–hydro–mechanical–seismic modeling of EGS collab experiment 1. *Energies* **14**, 446 (2021).

17. McClure, M. W. & Horne, R. N. Investigation of injection-induced seismicity using a coupled fluid flow and rate/state friction model. *Geophysics* **76**, WC181–WC198 (2011).

18. Zhai, G., Shirzaei, M., Manga, M. & Chen, X. Pore-pressure diffusion, enhanced by poroelastic stresses, controls induced seismicity in Oklahoma. *Proc. Natl Acad. Sci.* **116**, 16228–16233 (2019).

19. Kumazawa, T. & Ogata, Y. Nonstationary ETAS models for nonstandard earthquakes. *Ann. Appl. Stat.* **8**, 1825–1852 (2014).

20. Ritz, V. A. et al. Pseudo-prospective forecasting of induced and natural seismicity in the Hengill geothermal field. *J. Geophys. Res. Solid Earth* **129**, e2023JB028402 (2024).

21. Hainzl, S. & Ogata, Y. Detecting fluid signals in seismicity data through statistical earthquake modeling. *J. Geophys. Res. Solid Earth* **110**, B05S07 (2005).

22. Kumazawa, T. & Ogata, Y. Quantitative description of induced seismic activity before and after the 2011 Tohoku-Oki earthquake by nonstationary ETAS models. *J. Geophys. Res.: Solid Earth* **118**, 6165–6182 (2013).

23. Petrillo, G., Kumazawa, T., Napolitano, F., Capuano, P. & Zhuang, J. Fluids-triggered swarm sequence supported by a nonstationary epidemic-like description of seismicity. *Seismol. Res. Lett.* **95**, 3207–3220 (2024).

24. Aochi, H., Maury, J. & Le Guenan, T. How do statistical parameters of induced seismicity correlate with fluid injection? case of oklahoma. *Seismol. Soc. Am.* **92**, 2573–2590 (2021).

25. Qin, Y., Chen, T., Ma, X. & Chen, X. Forecasting induced seismicity in Oklahoma using machine learning methods. *Sci. Rep.* **12**, 9319 (2022).

26. Zhang, W. et al. Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: comprehensive review and future challenge. *Gondwana Res.* **109**, 1–17 (2022).

27. Chung, J., Ahmad, R., Sun, W., Cai, W. & Mukerji, T. Prediction of effective elastic moduli of rocks using graph neural networks. *Comput. Methods Appl. Mech. Eng.* **421**, 116780 (2024).

28. Camps-Valls, G., Tuia, D., Zhu, X. X. & Reichstein, M. *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences* (John Wiley & Sons, 2021).

29. Maniar, H., Ryali, S., Kulkarni, M. S. & Abubakar, A. Machine-learning methods in geoscience. In *Proc. SEG International Exposition and Annual Meeting*, SEG–2018 (SEG, 2018).

30. Bergen, K. J., Johnson, P. A., de Hoop, M. V. & Beroza, G. C. Machine learning for data-driven discovery in solid earth geoscience. *Science* **363**, eaau0323 (2019).

31. Yu, S. & Ma, J. Deep learning for geophysics: current and future trends. *Rev. Geophys.* **59**, e2021RG000742 (2021).

32. Mousavi, S. M. & Beroza, G. C. Deep-learning seismology. *Science* **377**, eabm4470 (2022).

33. Zhu, W. & Beroza, G. C. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophys. J. Int.* **216**, 261–273 (2019).

34. Reichstein, M. et al. Deep learning and process understanding for data-driven earth system science. *Nature* **566**, 195–204 (2019).

35. Anikiev, D. et al. Machine learning in microseismic monitoring. *Earth-Sci. Rev.* **239**, 104371 (2023).

36. Jinqiang, W., Basnet, P. & Mahtab, S. Review of machine learning and deep learning application in mine microseismic event classification. *Min. Miner. Depos.* **15**, 19–26 (2021).

37. Mousavi, S. M., Horton, S. P., Langston, C. A. & Samei, B. Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression. *Geophys. J. Int.* **207**, 29–46 (2016).

38. Li, Z., Eaton, D. W. & Davidsen, J. Physics-informed deep learning to forecast m^ max during hydraulic fracturing. *Sci. Rep.* **13**, 13133 (2023).

39. Yu, P. et al. Crustal permeability generated through microearthquakes is constrained by seismic moment. *Nat. Commun.* **15**, 2057 (2024).

40. Mital, U., Hu, M., Guglielmi, Y., Brown, J. & Rutqvist, J. Modeling injection-induced fault slip using long short-term memory networks. *J. Rock Mech. Geotech. Eng.* **16**, 4354–4368 (2024).

41. Hummel, N. & Shapiro, S. A. Nonlinear diffusion-based interpretation of induced microseismicity: A Barnett shale hydraulic fracturing case study. *Geophysics* **78**, B211–B226 (2013).

42. Vaswani, A. Attention is all you need. *Advances in Neural Information Processing Systems 30* (2017).

43. Zeng, A., Chen, M., Zhang, L. & Xu, Q. Are transformers effective for time series forecasting? *Proc. AAAI Conf. Artif. Intell.* **37**, 11121–11128 (2023).

44. Zhou, H. et al. Informer: beyond efficient transformer for long sequence time-series forecasting. In *Proc. AAAI Conference on Artificial Intelligence*, vol. 35, 11106–11115 (2021).

45. Schoenball, M. et al. Creation of a mixed-mode fracture network at mesoscale through hydraulic fracturing and shear stimulation. *J. Geophys. Res. Solid Earth* **125**, e2020JB019807 (2020).

46. Fu, P. et al. Close observation of hydraulic fracturing at EGS Collab Experiment 1: fracture trajectory, microseismic interpretations, and the role of natural fractures. *J. Geophys. Res. Solid Earth* **126**, e2020JB020840 (2021).

47. Kneafsey, T. J. et al. An overview of the EGS Collab project: field validation of coupled process modeling of fracturing and fluid flow at the Sanford Underground Research Facility, Lead, SD. In *Proc. 43rd Workshop on Geothermal Reservoir Engineering*, vol. 2018 (Curran Associates, Inc., 2018).

48. Qin, Y. et al. Source mechanism of khz microseismic events recorded in multiple boreholes at the first EGS Collab Testbed. *Geothermics* **120**, 102994 (2024).

49. Feng, Z. et al. Monitoring spatiotemporal evolution of fractures during hydraulic stimulations at the first EGS collab testbed using anisotropic elastic-waveform inversion. *Geothermics* **122**, 103076 (2024).

50. Kneafsey, T. J. et al. EGS Collab project: status and progress. In *Proc. 44th Workshop on Geothermal Reservoir Engineering* (Stanford University, 2019).

51. Kneafsey, T. J. et al. The EGS Collab project: learnings from experiment 1. In *Proc. 45th Workshop on Geothermal Reservoir Engineering*, 10–12 (Stanford University 2020).

52. Rothert, E. & Baisch, S. Passive seismic monitoring: mapping enhanced fracture permeability. In *Proc. World Geothermal Congress* 25–29 (European Association of Geoscientists & Engineers, 2010).

53. Baisch, S., Vörös, R., Weidler, R. & Wyborn, D. Investigation of fault mechanisms during geothermal reservoir stimulation experiments in the Cooper Basin, Australia. *Bull. Seismol. Soc. Am.* **99**, 148–158 (2009).

54. Li, Z., Elsworth, D., Wang, C. & EGS-Collab. Induced microearthquakes predict permeability creation in the brittle crust. *Front. Earth Sci.* **10**, 1020294 (2022).

55. Kneafsey, T. et al. The EGS Collab project: Outcomes and lessons learned from hydraulic fracture stimulations in crystalline rock at 1.25 and 1.5 km depth. *Geothermics* **126**, 103178 (2025).

56. Witherspoon, P. A., Wang, J. S., Iwai, K. & Gale, J. E. Validity of cubic law for fluid flow in a deformable rock fracture. *Water Resour. Res.* **16**, 1016–1024 (1980).

57. Ouyang, Z. & Elsworth, D. Evaluation of groundwater flow into mined panels. *Int. J. Rock Mech. Min. Sci. Geomech. Abstracts*, **30**, 71–79 (1993).

58. Morris, J. P. et al. Experimental design for hydrofracturing and fluid flow at the DOE EGS collab testbed. In *Proc. ARMA US Rock Mechanics/Geomechanics Symposium*, ARMA–2018 (ARMA, 2018).

59. Olson, J. E. Sublinear scaling of fracture aperture versus length: an exception or the rule? *J Geophys Res. Solid Earth* **108**, 2413 (2003).

60. Ishibashi, T., Watanabe, N., Asanuma, H. & Tsuchiya, N. Linking microearthquakes to fracture permeability change: the role of surface roughness. *Geophys. Res. Lett.* **43**, 7486–7493 (2016).

61. Chen, Z., Ma, M., Li, T., Wang, H. & Li, C. Long sequence time-series forecasting with deep learning: a survey. *Inf. Fusion* **97**, 101819 (2023).

62. Nie, Y., Nguyen, N. H., Sinthong, P. & Kalagnanam, J. A time series is worth 64 words: long-term forecasting with transformers. arXiv preprint. (2022).

63. Lengliné, O. et al. The largest induced earthquakes during the geoven deep geothermal project, Strasbourg, 2018–2022: from source parameters to intensity maps. *Geophys. J. Int.* **234**, 2445–2457 (2023).

64. Giuliari, F., Hasan, I., Cristani, M. & Galasso, F. Transformer networks for trajectory forecasting. In *Proc. 25th International Conference on Pattern Recognition (ICPR)*, 10335–10342 (IEEE, 2021).

65. Gischig, V. S. Rupture propagation behavior and the largest possible earthquake induced by fluid injection into deep reservoirs. *Geophys. Res. Lett.* **42**, 7420–7428 (2015).

66. Bhattacharya, P. & Viesca, R. C. Fluid-induced aseismic fault slip outpaces pore-fluid migration. *Science* **364**, 464–468 (2019).

67. Chen, Y. et al. Contiformer: continuous-time transformer for irregular time series modeling. *Adv. Neural Inf. Process. Syst.* **36** (2024).

68. Svozil, D., Kvasnicka, V. & Pospichal, J. Introduction to multi-layer feed-forward neural networks. *Chemometr. Intell. Lab. Syst.* **39**, 43–62 (1997).

69. Kingma, D. P. Adam: a method for stochastic optimization. arXiv preprint. https://doi.org/10.48550/arXiv.1412.6980 (2014).

70. Nix, D. A. & Weigend, A. S. Estimating the mean and variance of the target probability distribution. In *Proc. IEEE International Conference on Neural Networks (ICNN'94)*, vol. 1, 55–60 (IEEE, 1994).

71. Lakshminarayanan, B., Pritzel, A. & Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Proc. Advances in Neural Information Processing Systems.* Vol **30** (Curran Associates, Inc., 2017).

72. Feurer, M. & Hutter, F. Hyperparameter optimization. *Automated Machine Learning: Methods, Systems, Challenges* 3–33 (Springer, 2019).

73. Yang, L. & Shami, A. On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing* **415**, 295–316 (2020).

Author contributions
J. C. Conceptualization, Methodology, Investigation, Visualization, Writing —original draft, Review and editing. M. M. Conceptualization, Investigation, Supervision, Review and editing. T. K. Conceptualization, Investigation, Supervision, Review and editing. T. M. Investigation, Supervision, Review and editing. M. H. Conceptualization, Investigation, Funding acquisition, Project administration, Supervision, Review and editing.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-025-02644-z.

Correspondence and requests for materials should be addressed to Mengsu Hu.

Peer review information *Communications Earth and Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Marisol Monterrubio-Velasco and Joe Aslin. [A peer review file is available].

Reprints and permissions information is available at http://www.nature.com/reprints

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