Check for updates

RESEARCH ARTICLE

Real-time forecasting of induced seismicity in geological carbon storage with a machine learning workflow

[version 1; peer review: 2 approved with reservations]

Linus Walter[1,2], Kayla A. Kroll[3], Qingkai Kong[3], Sara Hanson-Hedgecock[1], Víctor Vilarrasa[1]

[1]Global Change Research Group (GCRG), Mediterranean Institute of Advanced Studies, Esporles, Illes Balears, Spain
[2]Department of Civil and Environmental Engineering (DECA), Universitat Politecnica de Catalunya, Barcelona, Catalonia, Spain
[3]Lawrence Livermore National Laboratory Physical and Life Sciences Directorate, Livermore, California, USA

v1 **First published:** 13 May 2026, **6**:164
https://doi.org/10.12688/openreseurope.23413.1
Latest published: 13 May 2026, **6**:164
https://doi.org/10.12688/openreseurope.23413.1

Abstract

Geological carbon storage (GCS) can significantly reduce emissions from hard-to-abate industries. However, large-scale deployment is hindered by the risk of induced seismicity, which has led to multiple project shutdowns. Mitigating such events would require unprecedented reliable seismicity forecasting. We present an adaptive near-real-time workflow that forecasts seismic variables. Using a random forest trained on sequentially partitioned data, we forecast seismicity rate and maximum magnitude. Applied to the Illinois Basin Decatur Project, the model receives operational parameters and the seismic catalog as input. Our results identify well pressure, past event counts, and prior maximum magnitudes as key predictors. However, accuracy declines over time as fluid pressure diffuses into the far-field, highlighting the need to incorporate spatiotemporal pressure diffusion for improved long-term forecasts. Our framework is the first to offer near-real-time forecasting from the operation start with uncertainty quantification, providing a foundation for next-generation seismic hazard mitigation in geo-energy operations.

Plain Language summary

Geological carbon storage (GCS) is essential for reducing CO2 emissions in hard-to-abate industries, such as cement or steel production. However, its widespread implementation is limited by the risk of inducing earthquakes during CO2 injection. Various fluid-injection projects were shut down after causing perceivable seismicity, like the Castor underground gas storage site, which caused three M4 events. Current strategies to mitigate undesired seismicity are limited due to the absence of reliable forecasting models. Our research addresses this gap by proposing a real-time model to forecast seismic trends during GCS operations. The proposed machine learning

Open Peer Review

Approval Status ? ?

	1	2
version 1 13 May 2026	? view	? view

1. **Jize Zhang**, The Hong Kong University of Science and Technology, Hong Kong, China
2. **Wendong Wang**, Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan, China

Any reports and responses or comments on the article can be found at the end of the article.

workflow continuously forecasts the seismic event count and maximum magnitude from the injection start, without requiring large training data. For the example of the Illinois Basin Decatur Project, the model quantifies uncertainties and reveals the evolving importance of physical parameters such as well pressure and past seismicity. By enabling earlier and more informed decisions of GCS operators, this adaptive workflow serves as a building block in the next generation of frameworks for seismic hazard control. It supports safer operation of GCS projects, which paves the way for the establishment of this technology as a viable and publicly accepted tool to reduce the environmental impact of industrial activities.

Keywords
induced seismicity, real-time seismicity forecasting, Illinois Basin Decatur Project, machine learning, Random Forest

This article is included in the European Research Council (ERC) gateway.

This article is included in the Horizon 2020 gateway.

Corresponding authors: Linus Walter (linuswalter@mailbox.org), Víctor Vilarrasa (victor.vilarrasa@csic.es)

Author roles: Walter L: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **A. Kroll K**: Conceptualization, Data Curation, Methodology, Software, Supervision, Writing – Review & Editing; **Kong Q**: Conceptualization, Methodology, Software, Supervision, Writing – Review & Editing; **Hanson-Hedgecock S**: Supervision, Writing – Review & Editing; **Vilarrasa V**: Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 801809 (GEoREST), from the Maria de Maeztu Excellence Unit programme under grant agreement No. CEX2021-001198, and from the Laboratory Directed Research and Development (LDRD) programme under grant agreement No. 24-ERD-012.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Walter L *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Walter L, A. Kroll K, Kong Q *et al.* **Real-time forecasting of induced seismicity in geological carbon storage with a machine learning workflow [version 1; peer review: 2 approved with reservations]** Open Research Europe 2026, **6**:164 https://doi.org/10.12688/openreseurope.23413.1

First published: 13 May 2026, **6**:164 https://doi.org/10.12688/openreseurope.23413.1

1. Introduction

Mitigating anthropogenic climate change requires not only decarbonizing the energy sector, but also the reduction of greenhouse gases from the industry sector.[1] Whereas the energy sector has great potential to meet the demand for heat and electricity from renewable sources, the decarbonization of the hard-to-abate industry, that is, heavy industries such as iron and steel, cement, and chemicals, is much more difficult and costly.[2] The main challenge is the requirement of high process temperatures as well as greenhouse gas emissions from the inherent chemical process reactions.[3] One viable solution for decarbonization is Geological Carbon Storage (GCS), as reported with high confidence in the Intergovernmental Panel on Climate Change (IPCC) report of 2023.[4] For exploiting the full potential of GCS for these industries, the current global injection capacity of $\approx 50 Mt/y$ needs to be upscaled by one order of magnitude towards a potential maximum capacity of $\approx 267 Mt/y$ until 2030 and by another order of magnitude to $\approx 3000 Mt/y$ until 2050.[1–5] Apart from these global storage objectives, the required storage capacity of a single GCS project is based on to the CO_2 output of nearby high-emitting industrial plants. The common production rates of CO_2 range from $q \approx 12 kg/s$ for biofuel production[6] to $q \approx 30 kg/s$ from a medium-sized cement plant[7] to the extreme case of $q \approx 1000 kg/s$ for one of the largest steel production plants in China.[8] GCS is based on injection of CO_2 from these industrial sources into a permeable reservoir rock at a depth below $800 m$ which is confined by a low-permeability caprock.[9,10] The injection of CO_2 causes fluid pressure buildup, which may extend tens to hundreds of kilometers during a multi-decade injection. This migrating fluid pressure reduces the effective normal stress along preexisting fractures and/or faults, which lowers contact friction along preexisting fault planes.[11] This phenomenon might result in seismic slip, depending on the prevalent shear stress in the system.[12]

This reactivation of preexisting faults is a major challenge to the large-scale implementation of GCS. Fortunately, induced seismicity is very unlikely to cause a leakage of CO_2 by damaging the caprock due to the favorable geomechanical characteristics of the soft sedimentary GCS reservoirs.[13,14] However, perceivable moment magnitudes $M_{w,max}$ might occur due to fault reactivation within the crystalline basement below a reservoir. In an unfavorable geological setting as the one of the Castor underground gas storage site (Spain), these conditions led to three magnitude 4 seismic events. The project was terminated following this incident, resulting in a total cost to the Spanish taxpayer exceeding 4 billion euros.[15,16] This example illustrates that successful management of induced seismicity is a prerequisite for large-scale implementation of GCS.

Operators can implement various strategies to mitigate the occurrence of perceivable seismic events with a maximum moment magnitude $M_{w,max} > 3$.[17] The predominant risk management framework to address induced seismicity is the traffic light system (TLS), which is voluntarily implemented by operators.[18–20] The TLS is a decision scheme that defines distinct operational response categories, namely green, yellow, and red, corresponding to threshold values of decision variables, such as seismic rate R, $M_{w,max}$, and ground motion. Measurements of these decision variables within the green TLS category imply that the operation continues as planned. Within the yellow category, mitigation measures should be assigned, for example, a reduction of the pumping rate q_{well}. The red category indicates a major intervention that could involve the termination of the operation.[18] For each color of the TLS, the threshold values of R, $M_{w,max}$, and ground motion are commonly based on site-specific risk and hazard assessment maps.[21,22]

A current limitation of the TLS is that its categories are associated with the monitored seismicity rather than the forecasted one.[19,23–25] The rigid coupling of the operational response to the measured values of R and $M_{w,max}$ is based on general empirical values that are not necessarily transferable to the geological conditions at the injection site. If the threshold values of R or $M_{w,max}$ used in the TLS are underestimated, the operation could fall into the red category although there is no large-magnitude event about to happen. Thus, the injection may be unnecessarily stopped, resulting in an economic loss. An example is the incomplete reservoir stimulation at the enhanced geothermal system (EGS) site at Otaniemi (Finland) due to fixed injection limits based on the total injected volume.[23–26] Alternatively, overestimation of the threshold values of R or $M_{w,max}$ could result in operators missing the opportunity to adjust the flow rate in time, and therefore being unable to prevent an unexpected high-magnitude event. TLS also fails to prevent large-magnitude earthquakes occurring after the stop of injection, a common phenomenon in EGS, like at the EGS in Basel (Switzerland) with $M_{w,max} = 3.4$[27] and Pohang (South Korea) with $M_{w,max} = 5.5$.[28–30] Although such negative examples have not yet occurred in the GCS industry, we strongly recommend treating them as warning signs. Instead of using fixed threshold values for R or $M_{w,max}$ within TLS, we firmly advocate investing in the development of accurate near-real-time induced seismicity forecasts to ensure the economic and social success of GCS projects.

Despite its significant potential, near-real-time forecasting of induced seismicity still requires major developments. This is because existing forecasting models either lack sufficient complexity to generate reliable forecasts (e.g., primarily statistical models) or are too computationally demanding for near-real-time calibration and execution (e.g., physics-based models).[31] Statistical models have been applied and developed in the seismological community mainly based on the

Gutenberg-Richter-Law and other decay principles similar to natural seismicity.[32–35] The advantage of such probabilistic approaches is that they require relatively low computational resources and provide an immediate output of a target variable, such as R or $M_{w,max}$. At the same time, they do not consider triggering mechanisms based on fluid injection itself. The coupling with simplified physical relationships can be provided through an analytical solution in hybrid physical-statistical models, as Shapiro et al. (2002)[36] show with their model for axisymmetric pressure diffusion. The simplified pressure front can be linked to the systems seismic response, allowing an estimate of the possible $M_{w,max}$.[37] This assumption of pore pressure as main driver of induced seismicity allowed Langenbruch et al. (2018)[38] to estimate R for $M_{w,max} > 3$ applied to Oklahoma's oil fields (USA). At the same time, such purely hydraulic models are not applicable in fracture-dominated reservoirs where hydro-mechanical stress coupling dominates the seismicity.[12–15,39] This shortcoming can be addressed by using multi-physical numerical models as input for probabilistic models instead. Traditionally, such complex physics-based models are mainly used to identify triggering mechanisms[15–40] of induced seismicity or modeling injection scenarios to prevent such unfavorable events.[41,42] In particular, numerical codes provide a high-resolution spatiotemporal distribution of state variables such as pressure, displacement and temperature based on physical laws.[43,44] As deterministic numerical models do not adequately address the uncertainty of occurring seismicity, the relationship between state variables and observed seismicity is typically established through probabilistic or empirical relationships.[45,46] Such hybrid workflows have been developed[47,48] and applied to well-documented injection sites, as demonstrated by Dempsey and Suckale (2017)[49] in Groningen (Netherlands), and Hill et al. (2024)[50] in the Raton Basin (USA). However, empirical and analytical equations that define the relationships between state variables and seismicity, such as rate-and-state friction, face limitations in their general applicability. This is due to uncertainties in the distribution of the required model parameters, and the potential influence of neglected physical effects, such as fault lubrication.[51]

A promising option to complement these shortcomings can be provided by machine learning (ML) algorithms. Advancements have been made with regard to correlating measured data with target variables such as R and $M_{w,max}$. For example, ML has been successfully tested at scales where large data are available, namely the lab scale[52–54] and the long time-series history of natural earthquakes. For the latter ones, De Vries et al. (2018)[55] applied a deep neural network to forecast aftershock sequences over time and space. Although their method outperforms a compared physics-based model applied for natural earthquakes, the applicability has not been demonstrated for human-induced earthquakes where the main shocks themselves would be of major interest. Similarly, Saad et al. (2023)[56] proposed a framework based on a Random Forest (RF) model for the spatiotemporal forecast of R and $M_{w,max}$. rf has also been applied to induced seismicity by Qin et al. (2022)[57] to a large dataset from Oklahoma (USA), who analyzed the relationship between various operational features and the seismic rate R over time and space. Their hindcast proofs its applicability by successfully associating hydraulic data such as q_{well}, well pressure p_{well}, and data from the seismic catalog for the same time window.

While such retrospective analysis offers valuable insight into the correlations of operations and observed seismicity, a primary challenge lies in the forecast of R and $M_{w,max}$. Near-real-time prospective forecasting for seismicity is still a young domain, with great potential for development and knowledge transfer into the domain of subsurface fluid injection. Until now, some pioneering works have been published, using different features of time-series data, while usually splitting the data into a fixed training period and a forecasting period. For example, Li et al. (2023)[58] used an artificial neural network to forecast the maximum magnitudes $M_{w,max}$ based on the previously measured data from $M_{w,max}$ at the example of the Tony Creek Dual Microseismic Experiment (Canada). Yu et al. (2023)[59] used a lagged linear regression to forecast R based on injection rates. Similarly, Mignan et al. (2024)[60] linked R to the injected volume based on the approach of Shapiro et al. (2011)[37] by applying a 'least absolute shrinkage and selection operator' model. Using a more suitable architecture for time-series forecasting, Picozzi and Iaccarino (2021)[61] passed features from a seismic catalog in a recurrent neural network to predict the preparatory phase of an $M_{w,max} = 4$ event.

Much of the previous work provides a framework that can be extended to near-real-time forecasting. However, most studies do not explore the potential of assessing their models' continuous performance from the beginning of the injection operation.[58–60] Therefore, we propose in this study a near-real-time forecasting workflow based on a random forest model that continuously assimilates new measurements and re-trains the model as new data are obtained. We establish our model utilizing data from the demonstration GCS injection site of the Illinois Basin Decatur Project (IBDP) (Section 5.2). These data are preprocessed with a shifted moving window approach as data-partitioning technique (Section 5.1). In Section 2, we describe the results of our forecast for the target variables, namely seismic event count N and $M_{w,max}$. An improvement in forecasting accuracy, highlighted by changes in feature importance and a better R^2 score, is observed after the initial IBDP seismic event cluster. Subsequently, the forecast performance diminishes after the second cluster, probably due to the growing importance of fluid pressure diffusion into the far field of the reservoir. Following these results, we evaluate in Section 3 the model performance when using solely hydraulic data compared to using both hydraulic and seismological data. We also provide guidance on the optimal window length as a key parameter within our modeling framework. Finally, we assess the current model's limitations and propose future development paths towards spatial forecasting. In

this way, our work contributes a significant component for adaptive near-real-time forecasting of induced seismicity in GCS.

2. Results

We present our modeling results for forecasting the number of seismic events N and the maximum magnitudes $M_{w,max}$. Each of these two target variables is forecasted with a separate model, since the importance of the respective characteristic features varies significantly among them. We preprocess the training data of both models by binning the ground truth to moving windows, where we use a window length $T = 8d$. In addition, we set the window shift to $\Delta t = 1d$ for every timestep (see Figure 7). This seems a pragmatic choice, since the assimilation and evaluation of operational and seismological data at a project site are likely to follow a daily cycle. The initial model is trained with all data that are obtained within the first 20 days. The continuous forecast starts after this initial traing, which means that the model incorporates new data and is re-trained after each timestep (see Figure 8).

2.1 Forecasting the seismic event count

We present the results of our continuous training and forecasting workflow for N in Figure 1. In the cumulative graph of N, labeled as N_c, we compare the evolution of the binned ground truth as $N_{c,true}$ (black line) and the model output as $N_{c,pred}$ (yellow dashed line). In addition, we use the Symmetric Absolute Percentage Error (SAPE) metric from Equation 2 to assess the forecasting performance, which is the most suitable metric for the given data range (see Extended Data S4 at Walter (2025)[62]). Regarding the evolution of our continuous forecast, we observe the first deviation from the ground truth during the first earthquake cluster C1 at $t = 75d$. As clearly visible in the zoomed inset figure of C1, the model output $N_{c,pred}$ underestimates $N_{c,true}$ until C1 ends, which is also indicated by the dominantly red color of the SAPE. It reaches values up to $SAPE_{max,o} = 96\%$ for C1 while the upper bound is 100%. After C1 ends, $N_{c,pred}$ overestimates $N_{c,true}$, shown by the mostly blue SAPE. For the subsequent cluster C2, the model also misses the onset of a sudden increase in the event rate. However, it adapts comparably faster to the transition from a high rate to a low rate of $N_{c,true}$ at the end of the cluster, as shown in the zoomed plot for C2 in Figure 1a and as the lower maximum error of $SAPE_{max,o} = 90\%$ indicates. In general, the forecasting performance of our model improves after 400 days, which is indicated by the reduction in the moving average for 100 windows of SAPE (dark blue). In total, our model overestimates $N_{c,true}$ in 64.4% of the windows while it underestimates 35.6%. Overestimation is generally more favorable by preventing negative effects on the project operation.

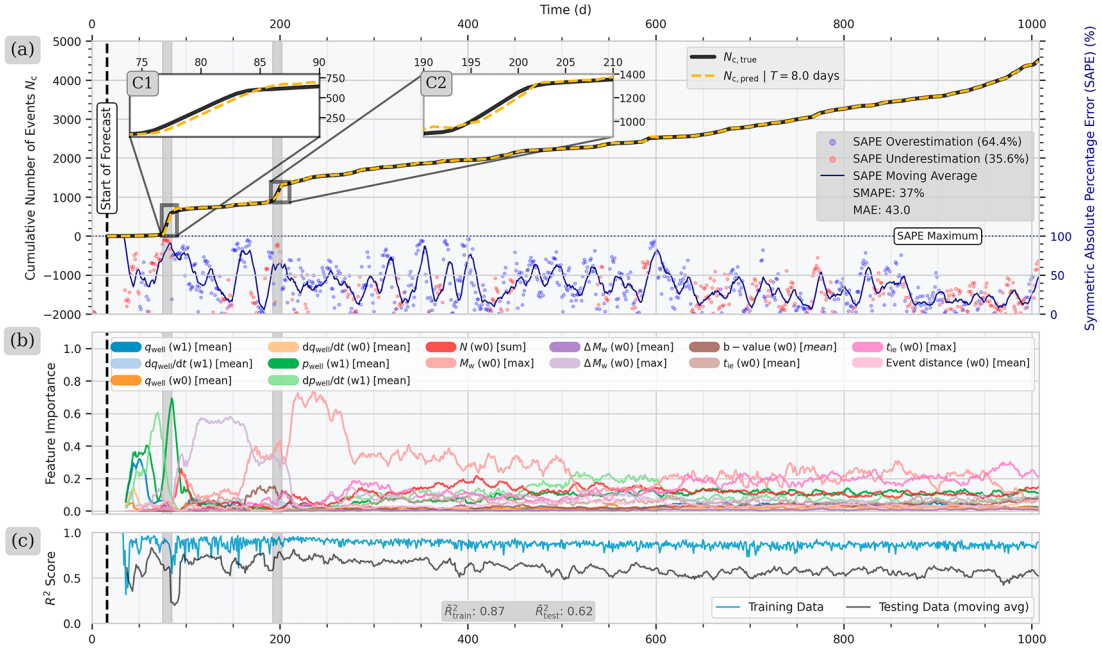

Figure 1. (a) Iterative forecast of the cumulative number of seismic events N_c for a window length of 8 days starting from day 4 with the ground truth $N_{c,true}$ (black) and the forecast values $N_{c,pred}$ (dark yellow). The Symmetric Absolute Percentage Error (SAPE) shows higher values after the first seismic cluster C1 in comparison to the period after of the second cluster C2. (b) The initially important training features are the hydraulic variables $p_{well}(w1)$ and $q_{well}(w0)$, which are replaced after C1 by the seismic features $\Delta M_{w,max}(w0)$ and $M_{w,max}(w0)$. (c) The R^2 score for training (blue) and testing (black) decrease slightly in the period between C2 and the end of the injection.

Apart from the apparent forecasting results, we provide insight into model training by displaying the evolution of feature importance and the R^2 score over time in Figure 1b and Figure 1c, respectively. While using both hydraulic and seismological features for training, our forecasting workflow identifies the hydraulic features $p_{well}(w1)$ and $q_{well}(w0)$ as the two most important ones until the onset of C1. This observation implies that the seismicity in that stage is mostly related to pore pressure changes. However, the elevated SAPE values observed during this period indicate that this feature importance can only be partially relied upon. After C1, the maximum magnitude difference $\Delta M_{w,max}(w0)$ is the dominant feature until day 160. Then, the maximum magnitude $M_{w,max}(w0)$ takes over shortly after the end of C2 with maximum values up to 0.75. Both, the importance of $\Delta M_{w,max}(w0)$ and subsequently $M_{w,max}(w0)$, could be correlated with static stress transfer caused during C1 and C2. The importance of $M_{w,max}(w0)$ remains in the range of $0.2 < M_{w,max}(w0) < 0.4$ during the period $270d < t < 500d$, while the error range increases again with values up to $SAPE = 98\%$. Similarly, there is no clearly dominant input feature observable after day 500, indicating that the model has difficulty relating the input variables to $N_{c,true}$. This observation is also supported by the evolution of the R^2 score in Figure 1c. The training score R^2_{train} reaches its highest values between C1 and C2 with $max\,R^2_{train} = 0.97$, while steadily decreasing afterwards. The R^2_{test} score behaves similarly, reaching $max\,R^2_{test} = 0.78$ at C1 and shortly after C2, showing a similar decline just afterward. We assume that the initial increase of both values is related to the increasing amount of available data and to a more constant range of the distribution of N-values, while the following decrease might be due to pressure diffusion in the far field, which we discuss in more detail in Section 3.3.

2.2 Forecasting maximum magnitudes

For the forecast of the maximum moment magnitude $M_{w,max}$ use principally the same input features, consisting of a combination of operational and seismological variables. A main challenge for modeling $M_{w,max}$ is that underestimations of the future observed $M_{w,max}$ could have more severe implications for the operation, in the worst case a cessation of the project, whereas an overestimation could only lead to a minor economic loss. We address this issue by scaling the target variable $M_{w,max}$ exponentially via $y_{true} = 10^{M_{w,max}}$, penalizing larger magnitudes more severely than smaller ones (see Equation 5). In this way, we achieve an underestimation rate of 15% and an overestimation of 42% (see pie chart in

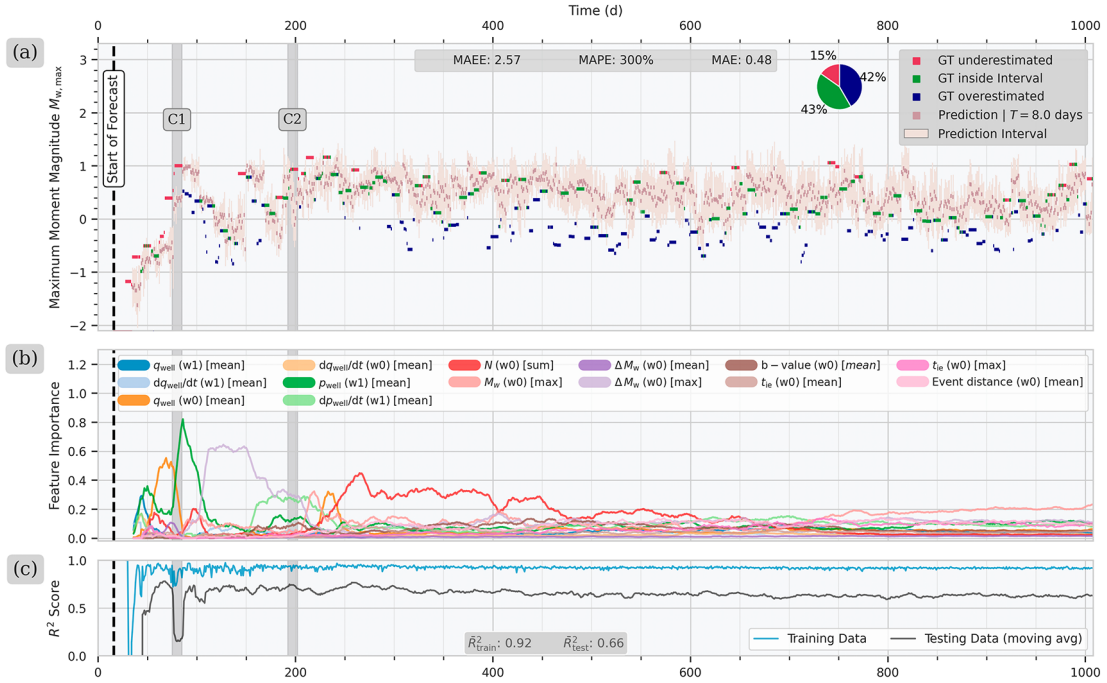

Figure 2. (a) Continuous forecast for the exponentially scaled maximum magnitude $M_{w,max}$ for a window length of 8 days starting after 4 days of injection. The forecasted values of $M_{w,max}$ are in dark pink and the associated standard deviation in light pink. Values of $M_{w,max}$ within the standard deviation are colored in green, while underestimates windows are colored in pink, and overestimated ones in blue. The pie chart in the upper-right corner summarizes the count of each category. (b) Feature importance values through time. The fluid pressure p_{well} is the most important feature after the onset of cluster C1, however it steadily decreases in importance as operation continues. (c) R^2_{test} score over time for the training data (blue) and testing data (black). The trend shows higher R^2_{test} scores after C1 and a slight decrease after day 350.

Figure 2a). In comparison, an alternative realization of the model that uses $y_{true} = M_{w,max}$ shows a fairly equal distribution of under- and overestimates with 26% and 27%, respectively (see Figure 1 in Extended Data at Walter (2025)[62]).

When analyzing the temporal trends of $M_{w,max}$ in Figure 2a, the measured magnitudes exhibit a range of $-1.17 < M_{w,max} < 1.1$ with significant fluctuations within the injection period. The forecasted magnitudes show a similar range between $-1.6 < M_{w,max} < 1.08$. From the start of injection until the end of the first cluster C1, our model underestimates most of the observed $M_{w,max}$ values. Shortly after C1 occurred, it rather overestimates, although it produces a few accurate predictions around day 145. During this time span, hydraulic features such as q_{well} and p_{well} have the highest importance with values of up to 0.8. In between C1 and C2, the importance of the hydraulic features decreases while the maximum magnitude difference $\Delta M_{w,max}$ becomes the most important. In the same time frame, the R^2_{test} score constantly increases to $R^2_{test} = 0.77$, while showing a drop during C1. In comparison, the observed maximum magnitudes increase within 50 days after C2 up to $M_{w,max} = 1.1$, when the forecast again shows a certain delay to adapt, as indicated by the red markers in Figure 2a. The following period after day 250 shows a majority of correct forecasting or overestimation, when the event count N is the most conclusive input feature, while R^2_{test} again reaches values greater than 0.75. Between day 600 and day 800, the model fails to forecast an event with $M_{w,max} = 1.05$, despite the overall satisfactory performance. During this period, no feature stands out as particularly important and the R^2_{test} score shows a slight decrease. Overall, the chosen forecasting model effectively reduces the underestimation of $M_{w,max}$ by using exponential scaling, although the forecast accuracy still shows potential for improvement.

3. Discussion

3.1 Forecast comparison: Hydraulic-only versus hydraulic and seismological features

The performance of our near-real-time forecasting workflow is mainly influenced by the quality of the input features provided to the model. The forecast results for N (Figure 1) and M_w (Figure 2), presented in Section 2, were obtained by training the RF model on both hydraulic and seismological input features, following referred to as RF_{HS}. To understand the impact of the seismic catalog on the model performance, we compare these results with a model called RF_H, which only uses the hydraulic features $p_{well}(w1)$ and $q_{well}(w0)$ for training. This distinction may be of interest for operators of GCS projects since these hydraulic parameters can be more conveniently obtained in near-real-time compared to the preparation of a seismic catalog. Table 1 lists the training features for each respective model, as well as several evaluation metrics for the forecast of N and $M_{w,max}$. For the forecast of N, the comparison of RF_H and RF_{HS} shows a slight, but not

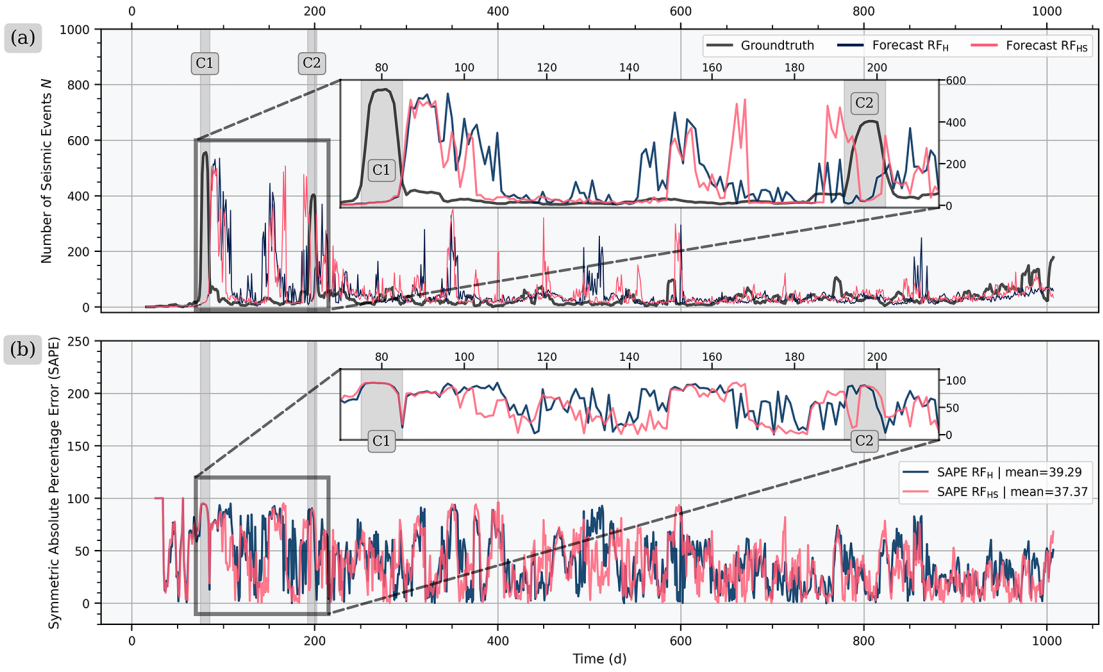

Figure 3. Comparison of forecasting performance for the number of seismic events N of model RF_H with only operational input features versus model RF_{HS} with operational and seismological input features. (a) While both models forecast a similar number of seismic events (N) during the early times, RF_{HS} shows less high-magnitude outliers over the total injection period. (b) For the Symmetric Absolute Percentage Error (SAPE), both models have the highest values during C1, while RF_{HS} has again a better performance afterwards, especially in between C1 and C2.

significant improvement of RF_{HS} compared to RF_H (see forecasting results of RF_H for N in Figure 2 and for $M_{w,max}$ in Figure 3 in Extended Data at Walter (2025)[62]).

In particular, the evolution of the forecast of both RF_H and RF_{HS} exhibits a nearly identical trend until the end of C1 (Figure 3a), due to the high importance of hydrological features in both models (recall Figure 1). Subsequently, both models miss the onset of C1, as it is clearly visible in the zoomed inset of Figure 3a. In between C1 and C2, both models overestimate the observed N values. In particular, RF_H and RF_{HS} forecast a high N right before C2, although both are unable to predict the actual increase of $N = 400$ during C2. Between day 200 and day 600, the observed values of N remain below 100, while the model RF_H forecasts $N > 200$ in seven occasions, and RF_{HS} in six occasions. These overestimations translate into several peaks of the SAPE, as shown in Figure 3b, exceeding 75% at those points. After day 600, the frequency and intensity of overestimates decreases, resulting in a generally decreasing trend of the SAPE (Figure 3b). In summary, the comparison of RF_H and RF_{HS} for forecast of N reveals only a marginal improvement from incorporating seismological features into the model.

For the forecast of $M_{w,max}$, Table 1 shows that the results for the most suitable performance metrics Mean Absolute Error (MAE) and AEE are nearly identical (see Extended Data S5 at Walter (2025)[62] for a discussion of the metrics). The overlapping trends of RF_H and RF_{HS} illustrate the dominance of hydraulic features until the end of C1, after which both models begin to deviate from each other (Figure 4a). In particular the feature importances of q_{well} and p_{well} exhibit a very similar pattern, indicating a strong correlation between hydraulic parameters and $M_{w,max}$ during this injection period (compare Figure 1 for RF_{HS} versus Figure 3 for RF_H). After C1, both models overestimate the observed $M_{w,max}$, which is shown by the negative exponential error in Figure 4b. Afterwards, both models miss the $M_{w,max} = 0.86$ event on day 142. Interestingly, both models forecast the onset of the $M_{w,max} = 0.94$ event at the beginning of C2, although they underestimate it by 0.2. The following event $M_{w,max} = 1.16$ is missed by both models, while the onset of the subsequent $M_{w,max} = 1.17$ is forecasted by both (see the inset axis in Figure 4a). In the remaining forecasts, both models rather overestimate $M_{w,max}$, while underestimation usually occurs after the onset of a larger $M_{w,max}$ event, for example, on days 570, 642, 739 and 984.

From the comparison of hydraulic input features in RF_H and hydraulic and seismic input features in RF_{HS}, both models show a very similar performance, both for the forecasts of N and $M_{w,max}$. For the given dataset of the IBDP, the performance gain from incorporating the seismic catalog in RF_{HS} is minor. Karimpouli et al. (2025)[63] observed a similar effect when forecasting the cumulative seismic moment for the dataset of the Otaniemi EGS (Finland), where the

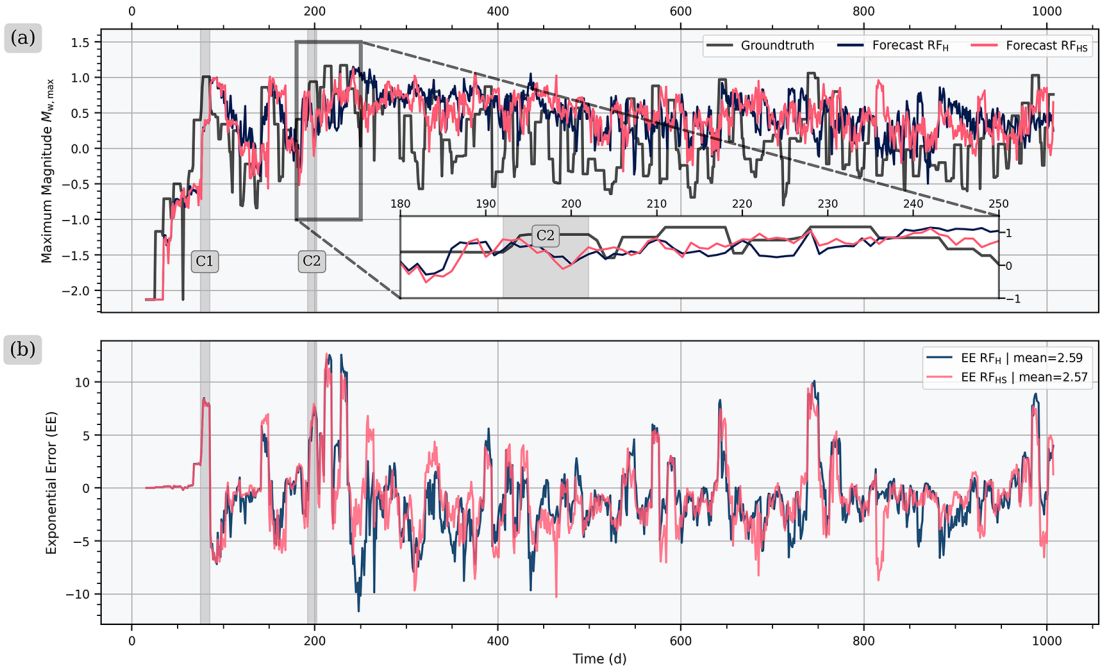

Figure 4. Comparison of forecasting performance for the maximum magnitude $M_{w,max}$ of model RF_H with only hydraulic input features versus model RF_{HS} with hydraulic and seismological input features for (a) $M_{w,max}$ and (b) the exponential error.

seismicity remained rather contained with limited event clustering. In contrast, the seismic catalog played an important role in forecasting cascading events of induced seismicity in the Cooper Basin EGS, where interaction of seismicity and reactivation of faults was more prevalent.

Therefore, the ability of our modeling workflow to incorporate both hydraulic and seismological data remains important, as the occurrence of such cascading earthquake sequences can not be known *a priori* to an operation.

3.2 Evaluating the model performance based on the window length

In addition to the input features, a main characteristic of our modeling workflow is that the forecast period equals exactly one window length T. Therefore, the choice of T determines how far in the future the target variable can be forecasted. We evaluate the influence of T on the forecast of both target variables by comparing in Figure 5 the evaluation metrics Symmetric Mean Absolute Percentage Error (SMAPE) from Equation 3 and R^2_{test} for the forecast of N, and MAE from Equation 5 and R^2_{test} for the forecast of $M_{w,max}$. For this analysis, we keep the step length fixed to $\Delta t = 1 day$.

The choice of the optimal window length in Figure 5a is based on the trends of both the SMAPE (lower is better) and the R^2_{test} score with $R^2_{test} \in (0,1)$ (higher is better). Taking both metrics into account, a window length of $T = 8$ days is optimal for forecasting N. Although R^2_{test} reaches even higher values at $T = 20$ days, SMAPE has its minimum value at 8 days. A smaller T allows for a higher resolution of the actual timely changes of N, while a larger T leads to a smoothing effect (see Supplementary Material S4). We visualize this trend in the inset of Figure 5, where the distribution of N shifts toward larger values for increasing T. For the same reason, we choose the SMAPE as a relative metric over an absolute metric like the MAE. Specifically, the increasing count of N for a larger T results in a larger absolute error range between y_{true} and y_{pred}, which leads to a nearly proportional increase of the MAE. The use of SMAPE avoids this effect by dividing the absolute error by y_{true}, which therefore ensures a better comparability of the results (see a detailed comparison for the performance metrics MAE, Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE) and SMAPE in Extended Data S4 at Walter (2025)[62]). While SMAPE evaluates the forecasting error as an extrapolation over time, we use the R^2_{test} score to evaluate the model regression performance on the input dataset. Here, the trend in Figure 5a shows a significantly reducing R^2_{test} score for 8 days, which might correspond to overfitting. Therefore, we choose 8 days as the optimal window length.

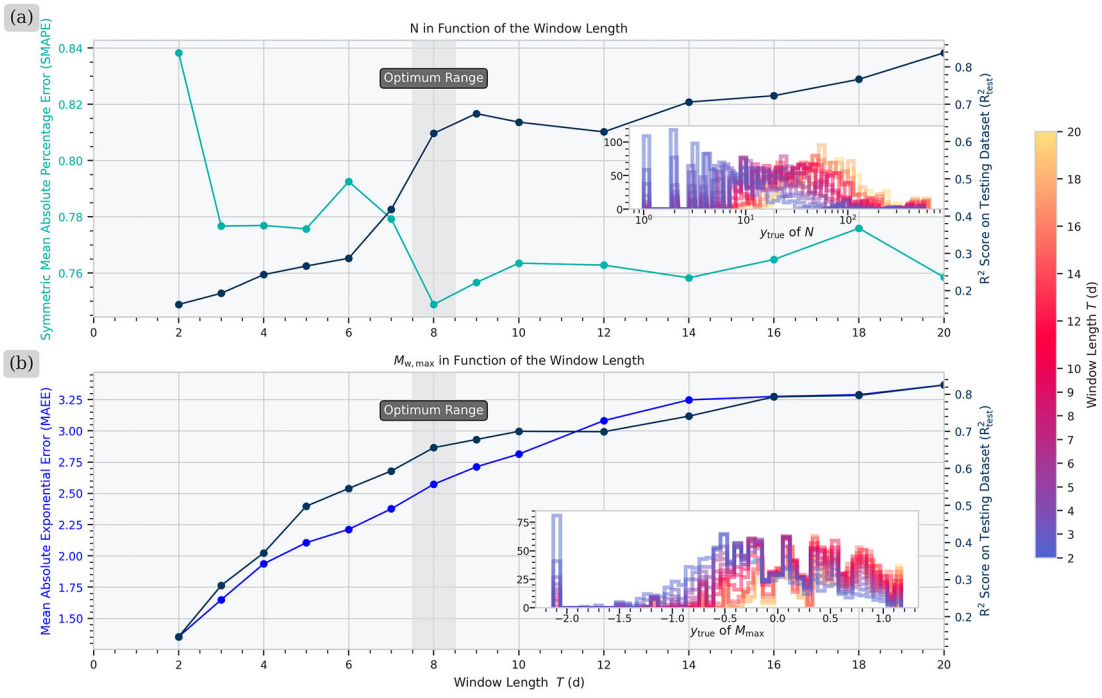

Figure 5. Evaluation of the influence of the window length T on the model performance for the forecast of (a) N and (b) and $M_{w,max}$. (a) For the forecast of N, the trend of the Symmetric Mean Absolute Percentage Error (SMAPE), and R^2_{test} indicate an optimum of 8 days. (b) The increasing Mean Absolute Exponential Error (MAEE) and the plateau of R^2_{test} suggest 8 days as the optimal T for the forecast of $M_{w,max}$.

Similarly to the forecast of N, we find that a window length of 8 days is also the best window length for the forecast of $M_{w,max}$. We define the optimal range by comparing the trends of the MAE and the R^2 score in Figure 5b. The MAE continuously increases from 2 days to $T = 20$ days, indicating that smaller windows yield a better MAE performance. This trend correlates with the shifting distribution towards an increasing number of larger $M_{w,max}$ values for a growing T, which is indicated by the color shift in the inset of Figure 5b. The shift leads to a decreasing Standard Deviation (STD) of the ground truth dataset of y_{true}, which influences the MAE results. Although the susceptibility of MAE to STD is not optimal for an objective comparison, it is the metric that shows the best balance between the influence of STD on the one hand, while penalizing deviations of y_{pred} from especially large magnitudes y_{true} (see a detailed comparison for the performance metrics MAE, MAE, MAPE and SMAPE in Extended Data S5 at Walter (2025)[62]). Similarly to the MAE, the R^2_{test} score doubles between $T = 2$ days and $T = 8$ days, reaching a value of $R^2 = 0.78$. Although it continues to increase to $R^2 = 0.82$ at $T = 20$ days, the slope remains much shallower beyond 8 days. As noted above, shorter window lengths are generally preferable, as they lead to less smoothing when partitioning measured data to generate y_{true} (see Extended Data S6 at Walter (2025)[62]). For these reasons, we select $T = 8$ days as the optimal window length, as it balances a low MAE, a sufficiently high R^2_{test}, and the shortest possible window length that satisfies both criteria.

3.3. Implications and limitations of our approach

The main innovation of this work is the use of continuous model updating through sequential time windows, which enables near-real-time forecasting from the early stages of injection.

Static hindcasting models, such as in Qin et al. (2022),[57] link input features and target variables within the same time window and are thus unable to perform forecasting because they cannot assign measured input features to a future window for forecasting the target variable. We address this challenge by associating features measured in a previous window with a target variable in a future window, storing both in the same row of the dataset, which enables the model to forecast (see Section 5.1). Additionally, our method updates the model architecture and retrains it as new data become available, further enhancing model performance. The continuous forecasting workflow also offers insight into the temporal evolution of feature importance.

Through these insights into our model performance, we observe that the importance of wellhead pressure p_{well} as a predictor of seismic activity in the RF model declines over time. Consistent with the findings of Feng et al. (2023),[64] our results indicate that wellhead pressure p_{well} and flow rate q_{well} are the most important features in the early injection phase. After approximately 100 days, however, our results show that the model increasingly relies on data from the seismic catalog (see in Section 2). Moreover, the R^2 scores for forecasting N and $M_{w,max}$ decline after day 400 (see Figure 1 and Figure 2). Both trends, the decreasing distinctiveness of feature importance and the declining forecast performance,

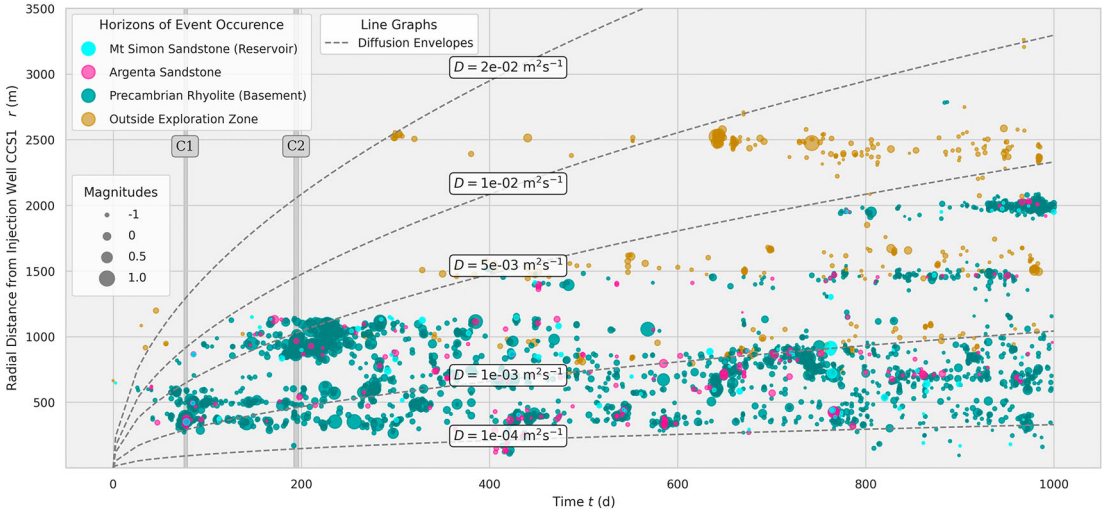

Figure 6. Spatiotemporal distribution of seismic events at the IBDP from the seismic catalog of Dando et al. (2021)[67] in correlation to radially symmetric pore pressure diffusion envelopes.[36] Most of the events happen in the Precambrian basement (teal). The radial diffusion envelopes (dashed lines) indicate that the seismicity locations can be correlated with a diffusion coefficient of $D = 1 \times 10^{-4} 1 \times 10^{-2} m^2/s$. The light-brown events could not be assigned to a horizon since they are located outside the boundaries of the reservoir model of Zaluski and Lee (2021).[68]

suggest that the model becomes progressively less capable of linking hydraulic and seismological measurements to the target values of N and $M_{w,max}$.

The primary reason for the declining forecasting performance over time may be that our workflow does not account for the spatiotemporal correlations among the input features. We illustrate these correlations in Figure 6, which shows each seismic event as a function of time t and radial distance r from the injector CCS1. The figure clearly reveals that seismic events occur at increasing distance from the injection well over time, which can generally be attributed to the migration of the fluid pressure front into the far field.

To quantify this trend, we include a radially symmetric diffusion envelope based on the formula $r = \sqrt{4\pi Dt}$ from Shapiro et al. (2002),[36] where D is the hydraulic diffusivity in $m^2\ s^{-1}$. The increasing distance of seismicity from the well during the injection period implies that the temporal delay between the pressure signal at the injection well p_{well} and the resulting seismic response increases over time. This delay hinders the model's ability to accurately forecast N and $M_{w,max}$.

Another factor that increases the difficulty of correct forecasting is the spatial heterogeneity of the reservoir. As illustrated in Figure 6, the majority of the observed seismicity occurred within the Precambrian basement. However, the observed distance-time relationship does not align with the measured diffusivity of this formation. Specifically, the diffusivity range of $D = 1 \times 10^{-4} 1 \times 10^{-2} m^2/s$ that is suggested in Figure 6 differs by approximately three orders of magnitude from the measured D of the Precambrian Rhyolite with $D = 1.6 \times 10^{-7} m^2/s$ (Table 2).[65] An explanation for this discrepancy is that horizontal pressure diffusion primarily occurs within the Mt. Simon Sandstone, from which fluid pressure then migrates vertically into the Precambrian basement along preexisting faults, partially critically stressed faults.[66]

To improve the forecasting ability of our suggested framework, it would be required to obtain the fluid pressure of the exact time and location of each seismic event. Since direct measurements of these data are not available, they need to be obtained from a reservoir model. Due to fault dominated seismicity at the IBDP, a simple radially symmetric reservoir model would not be sufficient. Instead, a 3D reservoir model would be needed, such as the hydro-mechanically coupled models of Silva et al. (2024)[66] or Luu et al. (2022),[69] which include geological strata and fault structures. Following, our machine learning framework could learn the relationship between the modeled state variables, such as fluid pressure and stress conditions, and the observed seismicity. In a second step, seismicity forecasts for future windows could be performed based on the reservoir model's output for planned injection scenarios. This would allow GCS operators to identify the optimal operational response in dependence of forecasted seismicity.

Due to characterization uncertainties of the reservoir with respect to fault locations and reservoir properties, the numerical reservoir model should be continuously updated as the operation evolves. Currently, most existing numerical modeling codes are too computationally expensive for updating reservoir models in near real-time. However, recent advances in high-performance multi-physics simulators are bringing such models within reach. For instance, Alkhimenkov et al. (2021)[70] developed a coupled multiphysics model capable of simulating wave propagation in a poroelastic medium on 1.5×10^9 grid cells in near-real time. Similarly, Castelletto et al. (2024)[71] provide the open-source software 'Geos', which has been used by Ju et al. (2024)[72] as input for generating a 2D surrogate model of a fractured CO_2 reservoir. A surrogate model of this kind would provide an ideal basis for training our forecasting model with spatially resolved data of state variables and for testing future injection scenarios in a computationally efficient way. In parallel, the continuously updating evaluation metrics of our forecasting framework provide transparent insights on the most important training features as well as into the forecasting performance.

4. Conclusions

This study proposes a continuously updating near-real-time forecasting workflow for induced seismicity. The workflow is based on an adaptive random forest model that is trained on hydraulic and seismological data. It forecasts the number of seismic events N and the maximum moment magnitude $M_{w,max}$ from the early stages of injection, without requiring a large dataset for initial training.

For the presented case study of the Illinois Basin Decatur Project (IBDP), we find that the forecast for N increased in accuracy after the first 500 observed events. Continuous evaluation metrics indicate that hydraulic features, such as well pressure p_{well} and flow rate q_{well}, are most important during the early injection phase, while seismological features gain importance at later times. Similarly, the forecast performance for $M_{w,max}$ improves as more data become available, particularly after the initial cluster of events and up to the mid-term injection period (around 400 days). Again, p_{well} and q_{well} are the dominant features until day 100, after which seismological features become more influential. In addition, we find that our model shows similar results when only using hydraulic input features, although the combination of hydraulic and seismological input results in a slightly higher accuracy. In both cases, the model underestimates $M_{w,max}$ in only 15%

of the forecast windows, primarily due to the exponential scaling of the target variable during training. This is a favorable result, as underestimating $M_{w,max}$ could more seriously impact the project, potentially causing it to stop, depending on the TLS threshold values. Although seismological features did not play a significant role in the forecast of the rather contained seismicity at the IBDP, it is expected that it may make a significant difference in reservoirs with rather cascading earthquake sequences.

Both forecasts for N and $M_{w,max}$ show a decreasing model performance after approximately 400 days of injection, likely due to the 1D nature of our model, which does not account for fluid pressure diffusion into the far field of the reservoir. Therefore, a future development of this workflow would focus on linking observed seismicity to the spatiotemporal distribution of the state variables that are obtained from reservoir models.

In general, we find that our approach is a first important step towards replacing fixed threshold values of R and $M_{w,max}$ in TLS decision schemes by directly forecasting these variables in near real-time. Beyond providing novel insights into forecast performance, our findings underscore the importance of incorporating spatiotemporal distribution of state variables to further improve forecasts of seismological variables.

5. Methods

5.1 Moving window approach with sequential time windows

In this study, we present a near-real-time forecasting workflow that consists of a window data-assimilation block and a random forest (RF) regression model to forecast the number of seismic events N and the maximum moment magnitude $M_{w,max}$. Apart from the relatively simple training procedure, one main advantage of using RF is that it can handle high-dimensional data and provides insights into the importance of each input feature for the forecast of the target feature. Another merit for near-real-time forecasting is that available code implementations of RF in Python are highly developed and are capable of processing large amounts of data in a short time.[73] However, RF is not specifically developed for time-series forecasting. Therefore, we propose a data partitioning scheme (Figure 7) and a corresponding modeling workflow (Figure 8) that continuously updates the training dataset of the RF via a moving window approach. In general, RF regression is an ensemble learning technique that is based on regression trees. For a given dataset, it learns the relationship between numerical values of given input features with the numerical value of a target feature, in our case N or $M_{w,max}$. Each regression tree of the RF is initialized randomly, which means that the location of each feature inside the tree is randomly assigned. The training of the RF model is performed on a training dataset by optimizing the decision threshold at each node within each decision tree until the error between the input training data and the target training data reaches a local minimum.[74] For finding the optimal model architecture, a hyperparameter tuning of the model is performed on an exclusive validation dataset. In a traditional RF modeling workflow, like the one of Qin et al. (2022),[57] the input variables I of each time window w_1 are associated to the target variable O of the same window, resulting in $I(w_1) \rightarrow O(w_1)$. This technique is suitable for evaluating a dataset *a posteriori* where the influence of variables like q_{well}, p_{well}, and poroelastic

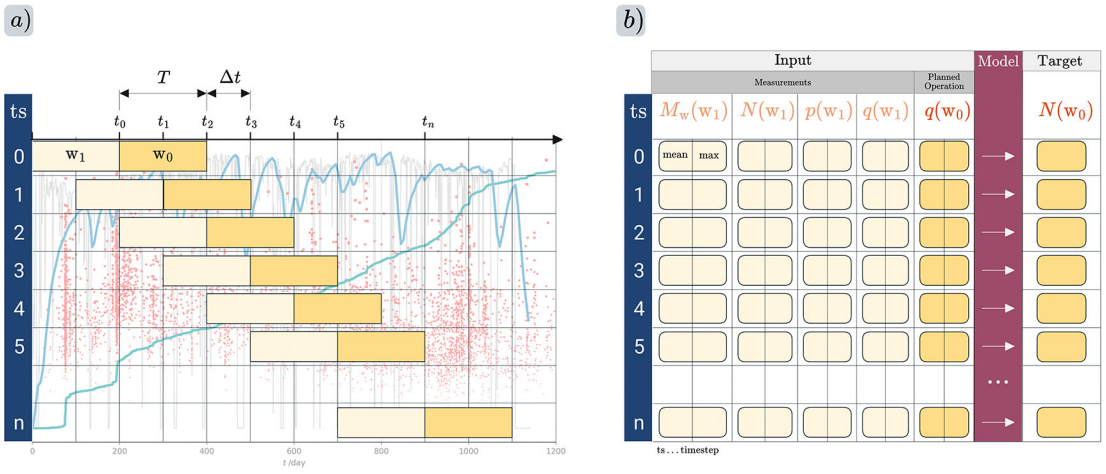

Figure 7. (a) Windowing approaches used to partition the training and testing data from the IBDP site. For each timestep, measurements are assigned to window w_1 while q_{well} for the future pseudo-planned operation and the seismological target variable, either N or $M_{w,max}$, are assigned to window w_0. Statistical metrics are calculated for each variable, including the mean and maximum of each window, as well as optionally the cumulative sum or difference over time, among other metrics. (b) Location of the model layer (dark red) that learns the relationship between the input features and the target features.

stress rate $\partial\sigma'/\partial t$ are related to seismological target features such as $M_{w,max}$ or N. However, this approach is not suitable for forecasting because the window w_1 must lie in the past before the measurement of $I(w_1)$ can be associated with the target variable $O(w_1)$ that is supposed to lie in the future. Therefore, we propose to apply a shifted moving window approach that continuously evolves with injection. Here, the RF model learns the dependency between the measured input variables $I(w_1)$ and input variables from the pseudo-planned operation $I(w_0)$, and the forecasting variable of the future window $O(w_0)$, which can be written as

$$[I(w_1), I(w_0)] \rightarrow O(w_0). \tag{1}$$

As shown schematically in Figure 7a, we define two windows for each timestep, labeled w_0 for the prediction and w_1 for the past window. We decide for this numbering scheme because it can be expanded to incorporate a greater number of previous windows while maintaining a practical logic (e.g., w_2, w_3, etc.).[58] Each of the windows has an equal length T, while the window is shifted by Δt after each iteration of the timestep, which we assume in our work as $\Delta t = 1d$. The window w_1 represents the past window at each time step which groups all observed variables, for example $M_w(w_1)$, $N(w_1)$, $p(w_1)$, and $q(w_1)$. Since most variables have several data points for each window, we calculate statistical parameters from them, such as the mean value, the maximum value, or optionally also the cumulative sum or difference over time for each variable. Apart from using primary measured variables as input features, we extend the dataset by also using post-processing variables, such as moving b-values, because such indirect measures might prove useful for prediction.[75–77] Each statistical value of every variable is assigned to a distinct column in the training dataset, indicated by a light peach color in Figure 7b. In addition to the observed data in window w_1, we group future data in window w_0 where we mainly consider the pseudo-planned flow rate $q(w_0)$. Finally, we assign the target variable in w_0, which in our study is written as $N(w_0)$ and $M_{w,max}(w_0)$. Hence, the data partitioning in shifted windows allows us to use the measured data of window w_1 and the pseudo-planned flow rate $q(w_0)$ to forecast the target variable of the next window w_0. In case the step length Δt is smaller than T, the partition windows overlap. In our case of *a posteriori* analysis, windows that are just located after the current forecast window would leak the measured values of the target variable into the forecast. Therefore, we label these windows as 'shadow' and do not use them for training (see Extended Data S7 at Walter (2025)[62]).

This preprocessed training dataset is used within our forecasting workflow that is capable of continuously assimilating data, (re-) training the RF model, and forecasting the target variables N and $M_{w,max}$. As described in Figure 8, the

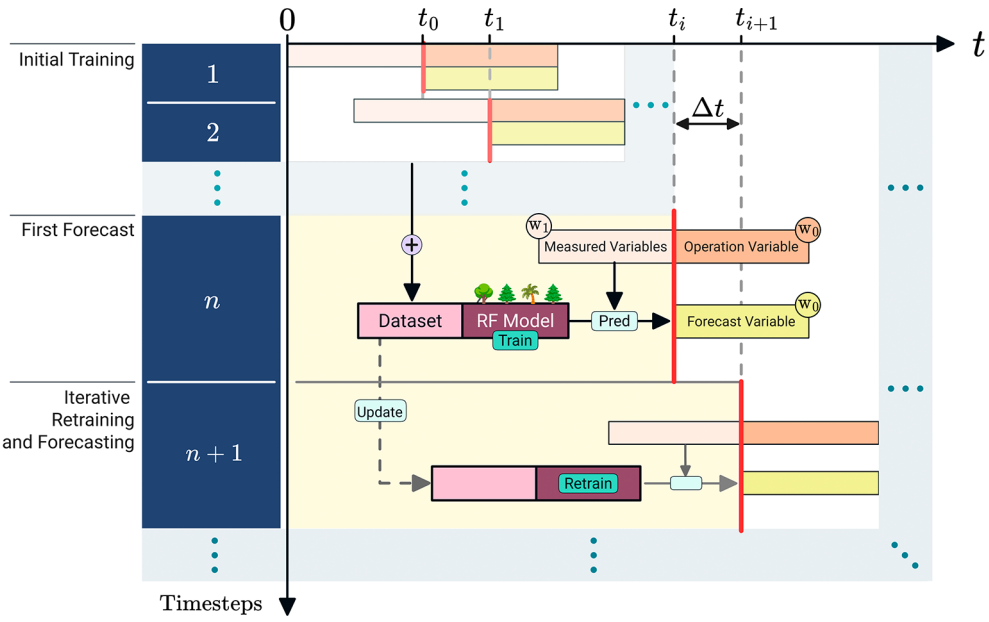

Figure 8. Real-time workflow for data assimilation and seismicity forecasting. The workflow starts with an initial training with timestep n where the training dataset consists of $n-1$ data points. The first forecast is made based on the measurements and the operation variable obtained until time t_i. The random forest model receives measured variables from window w_1, and an operation variable from window w_0 as input features and learns their relationship to a forecast variable, for example, the cumulative number of seismic events N_c, for window w_0. Subsequently, the workflow continuously retrains the RF model with the latest obtained data for every new timestep.

Table 1. Comparing model evaluation metrics when using only hydraulic input features from the injection (RF_H) versus also using the seismic catalog (RF_{HS}). GT is the ground truth.

	N		$M_{w,max}$	
	RF_H	RF_{HS}	RF_H	RF_{HS}
Input Features:				
$q_{well}(w1)$ [mean]	x	x	x	x
$dq_{well}/dt(w1)$ [mean]	x	x	x	x
$p_{well}(w1)$ [mean]	x	x	x	x
$dp_{well}/dt(w1)$ [mean]	x	x	x	x
$p_{well}(w0)$ [mean]	x	x	x	x
$dp_{well}/dt(w0)$ [mean]	x	x	x	x
$N(w0)$		x		x
$M_w(w0)[max]$		x		x
$M_{w,diff}(w0)[mean]$		x		x
$M_{w,diff}(w0)[max]$		x		x
$b-value(w0)$		x		x
$t_{ie}(w0)[mean]$		x		x
$t_{ie}(w0)[max]$		x		x
Evaluation Metrics:				
Underestimate of GT	40.9	35.6	14	15
Overestimation of GT	59.1	64.4	44	42
GT in Interval	–	–	40	43
MAE	45	43	0.47	0.48
SMAPE (%)	39.15	37.15	–	–
AEE	–	–	2.59	2.57

modeling workflow consists of an initial training stage and subsequent continuous forecasting stages. The initial training starts at timestep n at time t_i after the start of the operation and includes $n-1$ rows in the training dataset. We train the RF model on these data where we use the squared error $(y_i-\widehat{y_i})^2$ as optimization criterion. In the following, we obtain a forecast for the next timestep by passing the measurements of the window w_1 and the pseudo-planned hydraulic parameter $q_{well}(w_0)$ to the trained model. Note that the forecast is provided for exactly one window length T, which is one of the key features of our shifted moving window approach. After the first forecast, we continue with the timestep $n+1$ at time t_{i+1}. We update the training dataset with the input data from the previous timestep n and retrain the model. Then, we use the newly measured data and the pseudo-planned $q_{well}(w_0)$ to forecast the next window of the forecast

Table 2. Hydraulic material parameters of the seismically relevant horizons of the IBDP. The parameters are calculated based on the study of Bondarenko et al. (2022)[65] (See full table in Extended Data S8 at Walter (2025)[62]).

Property	Mt. Simon Sandstone (Reservoir)	Argenta Sandstone	Precambrian Rhyolite, Intact (Basement)	Precambrian Rhyolite, Fractured (Basement)
Intrinsic permeability k [m^2]	1e-16	1e-18	1e-21	1e-18
Porosity ϕ [–]	0.115	0.138	0.005	0.020
Specific Storage S_s [m]	6.8e-07	8.5e-07	5.8e-08	1.2e-07
Diffusivity D [$m^2\ s^{-1}$]	1.4e-03	1.1e-05	1.6e-07	7.9e-05

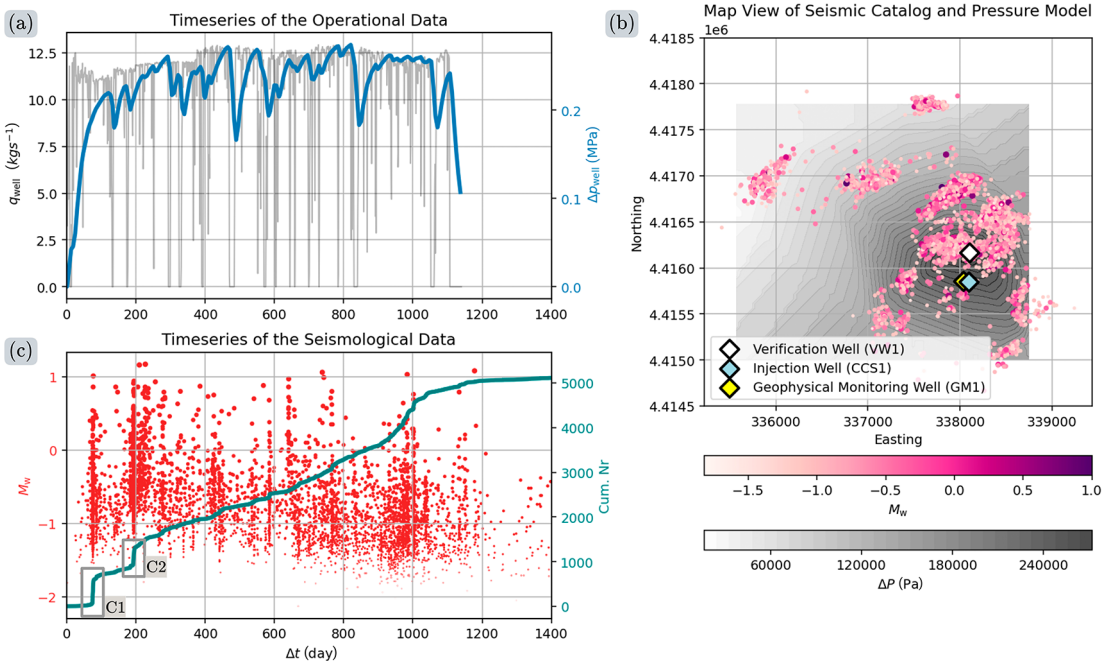

Figure 9. Arrangement of the input data for our model from the IBDP. (a) Hydraulic data q_{well} and p_{well} over the 1100 day duration of the project. We use the differential pressure at $z = 1943mBSL$ from the pressure model of Luu et al. (2022).[69] (b) Map of the injection site including the injection well CCS1 (blue), the verification well VW1 (white), the geophysical monitoring well GM1 (yellow), the spatial distribution of M_w (pink to purple), and the fluid overpressure p_{well} ranging from $0.02MPa$ (white) to $0.27MPa$ (dark gray). (c) Temporal evolution of seismic event moment magnitudes M_w (red) and the cumulative number of seismic events N_c (teal). Two seismic clusters of interests are highlighted with inset boxes labeled C1 and C2 (adapted with data from Dando et al. (2021)[67] and Luu et al. (2022)[69]).

variable. In this way, the training dataset grows continuously with the succession of the operation. The size of the training dataset also influences the optimum model architecture. Therefore, we perform a two-stage grid search with a coarse and a narrow parameter spacing after every iteration to identify the optimal hyperparameters of the RF model. This tuning includes the number of trees, the depth of each tree, the minimum samples per branch and the minimum samples per leaf. Although this optimization adapts the model architecture to the data size, we expect the forecast to have relatively low forecasting capability at the early timesteps due to the data scarcity. However, the forecast improves with every timestep iteration as the training dataset grows.

We evaluate the model's performance by calculating the error of the near-real-time forecast and by assessing its fit on the training and testing data by[1]calculating the error of the near-real-time forecast and[2] assessing its fit on the training and testing data. We use different metrics to evaluate the forecast of N and $M_{w,max}$, as each dataset has a distinct range and distribution of values. This leads to varying penalization of under- and overestimation depending on the chosen error metric. First, we evaluate the forecast for N at each window by calculating the sape

$$SAPE = \frac{|(y_{true} - y_{pred})|}{|y_{pred}| + |y_{true}|}, \tag{2}$$

for the model output y_{pred} and the y_{true} for each window. We summarize the forecast performance over the whole injection time using the SMAPE

$$SMAPE = 100\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y^i_{true} - y^i_{pred}}{|y^i_{pred}| + |y^i_{true}|}\right|, \tag{3}$$

and the MAE

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y^i_{true} - y^i_{pred}|. \tag{4}$$

We justify this selection with a detailed comparison of all evaluation metrics in Extended Data S4 at Walter (2025).[62]

For the forecast of $M_{w,max}$, we use the MAEE

$$MAEE = \frac{1}{N}\sum_{i=0}^{N}(|10^{y_{true}} - 10^{y_{pred}}|), \tag{5}$$

in order to penalize especially deviations from large magnitudes (see detailed comparison for metrics in Extended Data S5 at Walter (2025)[62]). Furthermore, we calculate the uncertainty range of the model at every forecast step based on the Standard Deviation (STD) of the returned random forest values from the mean output value. In addition to the forecasting metrics themselves, we analyze several evaluation metrics that give insight into the iterative training of the model. We use the feature importance based on the decrease in impurities in our model.[78] With this metric, we obtain the relative correlation of each input variable to the forecast variable at every timestep. Another proxy of training performance is the R^2 score, given by

$$R^2 = 1 - \frac{\sum_i \left(y^i_{true} - y^i_{pred}\right)^2}{\sum_i \left(y^i_{true} - \hat{y}_{true}\right)^2}, \tag{6}$$

which we compute in a 2-fold cross-validation scheme. Throughout the continuously evolving modeling workflow, we track both the feature importance and the R^2 score over time, providing an objective basis for evaluating and comparing the model's forecast performance. For the same reason, we use these metrics in the following Section 2 and Section 3 to assess and compare different versions of our model.

5.2 Case study: Illinois basin decatur project project

We apply the proposed near-real-time forecasting workflow to the dataset of the Illinois Basin Decatur Project (IBDP) to test and validate our method. IBDP was a large-scale GCS demonstration project, injecting $1 \times 10^6 t$ of CO_2 over 3 years between 2011 and 2014. The CO_2 was supplied by a bio-fuel production plant operated by the Archer Daniels Midland Company. This CO_2 was injected through the nearby injection well (CCS1) into the Lower Mount Simon sandstone reservoir at a depth ranging from $1923m - 1932m$ below sea level. The reservoir is bounded by the low-permeability Eau Claire shale as a caprock and the low-permeability Precambrian basement below (see Section S9 in Extended Data at Walter (2025)[62]).[79,80] During injection, the CO_2 flow rate q_{well} in the CCS1 injection well was largely within the range of $11kg^{1s^{-1}} < q_{well} < 12.5kg^{1s^{-1}}$,[81] except for a few maintenance breaks (see Figure 9a).[79] The bottomhole fluid pressure p was measured in CCS1 and in multiple zones in the VW1 verification well until injection ceased after 1107 days. For our work, we use the overpressure p_{well} in the CCS1 injection well at a depth of $z = 1943mBSL$ from the pressure model of Luu et al. (2022)[69] and Will et al. (2016),[81] reaching up to $p_{max} = 0.27MPa$. In addition to the hydraulic parameters q_{well} and p_{well}, microseismic activity was monitored at the IDBP site with an array of three geophones in the injection well CCS1 located near the injection depth and an array of 31 geophones at $624 - 1052mBSL$ in the observation well GM1.[6,80] The magnitudes of microseismic events range from $-2.1 < M_w < 1.1$ with a magnitude of completeness of $M_c = -0.8$. The two main event clusters in the seismic catalog, namely C1 and C2, appeared after day 75 and after day 192, as indicated in Figure 9c.[6] For our work, events in the catalog were relocated with double-difference by Dando et al. (2021).[67]

Ethics and consent

Ethical approval and consent were not required.

Data and software availability

Underlying Data

Repository name: Real-Time Forecasting of Induced Seismicity in Geological Carbon Storage with a Machine Learning Workflow. https://doi.org/10.20350/digitalCSIC/17640.[62]

The project contains the following underlying data:

01_IBDP_RF_Model.py (Main Python code for running the model that reproduces the results from the paper).

02_IBDP_RF_Postproc.py (Python code to visualize the results).

README.md (Step-by-step guide for installation and running the case study).

venv_requirements.txt (File to create a virtual environment in conda; see README.md).

00_readme_figures/ (Example results that can be obtained with the code).

01_data/decatur_pressure.hdf5 (Reservoir pressure from the reservoir model of Luu et al. (2022)[69]).

01_data/IBDP_injection.csv (Well pressure and injection rate of Luu et al. (2022)[69]).

01_data/decaturSeismic.hdf5 (Seismic catalog from Dando et al. (2021)[67]).

02_model_output/ (Directory that receives all output from finished model runs, named by timestamp, e.g., 02_model_output/model_20251009_180354).

02_plots/ (Directory to collect miscellaneous plots from the modeling runs that provide insights into the data).

/utils/ (Directory containing supporting code for the main code files 01_IBDP_RF_Model.py and 02_IBDP_RF_Postproc.py).

Data are available under the GNU General Public License 3 or later (GPL-3.0).

Extended data

Repository name: Real-Time Forecasting of Induced Seismicity in Geological Carbon Storage with a Machine Learning Workflow. https://doi.org/10.20350/digitalCSIC/17640.[62]

This project contains the following extended data:

Supplementary Figure 1 in Section S1. Modeling Results for Unscaled $M_{w,max}$

Supplementary Figures 2 and 3 in Section S2. Modeling Results for having only operational input data

Supplementary Figure 4 in Section S3. Shadow Effect of Overlapping windows

Supplementary Figure 5 in Section S4. Comparison of Different Error Metrics for Number of Events N

Supplementary Figure 6 in Section S5. Comparison of Different Error Metrics for Maximum Moment Magnitude $M_{w,max}$

Supplementary Figure 7 in Section S6. Comparison of Binning Results for Different Window Length T for Number of Events N

Supplementary Table 1 in Section S7. Spatiotemporal Evolution of Seismicity

Supplementary Figure 1 in Section S8. 3D Visualization of the Seismic Cloud at the IBDP.

Software availability

- Source code available from: https://github.com/linuswalter/Real-Time-Forecast-IBDP
- Archived software available from: https://doi.org/10.20350/digitalCSIC/17640
- License: GNU General Public License 3 or later (GPL-3.0)

We implement the presented modeling workflow in Python 3.11.9.[82] The random forest model and its evaluation are performed via scikit-learn version 1.5.0,[73] available under the BSD license at https://scikit-learn.org. We create all figures with Matplotlib version 3.9.0,[83,84] available under the Matplotlib license at https://matplotlib.org/. We calculate moving b-values via the "magnitude-distribution" python module of Goebel et al. (2019).[85] All computations are performed on an Ubuntu Server with Intel(R) Xeon(R) CPU 5118, and 64 RAM.

References

1. IEA: *Net Zero by 2050 - A Roadmap for the Global Energy Sector.* Paris: International Energy Agency; 2021.
Reference Source

2. English JM, English KL: **An overview of carbon capture and storage and its potential role in the energy transition.** *First Break.* 2022; **40**(4): 35–40.
Publisher Full Text

3. Kumar A, Tiwari AK, Milani D: **Decarbonizing hard-to-abate heavy industries: Current status and pathways towards net-zero future.** *Process Saf. Environ. Prot.* 2024; **187**: 408–430.
Publisher Full Text

4. Calvin K, *et al.*: *IPCC, 2023 - Climate Change 2023 - Synthesis Report. Summary for Policy Makers.* Intergovernmental Panel on Climate Change (IPCC); 2023.
Publisher Full Text

5. Global CCS Institute: *Global status of CCS 2024: Collaborating for a net-zero future.* Global CCS Institute; 2024.
Reference Source

6. Will R, Smith V, Leetaru HE, *et al.*: **Microseismic monitoring, event occurrence, and the relationship to subsurface geology.** *Energy Procedia.* 2014; **63**: 4424–4436.
Publisher Full Text

7. Peng JX, *et al.*: **Modeling of carbon dioxide measurement on cement plants.** *Adv. Mater. Res.* 2012: 610, 2120–613, 2128.
Publisher Full Text

8. Li H, Cang D, Guo Y, *et al.*: *The analysis of energy structure of coal using and CO2 emission of a typical steel industry.* Shanghai, China: IEEE; 2011.
Publisher Full Text

9. Bachu S: **CO2 storage in geological media: Role, means, status and barriers to deployment.** *Prog. Energy Combust. Sci.* 2008; **34**(2): 254–273.
Publisher Full Text

10. Benson SM, Cole DR: **CO2 sequestration in deep sedimentary formations.** *Elements.* 2008; **4**(5): 325–331.
Publisher Full Text

11. Terzaghi K: **Stress conditions for the failure of saturated concrete and rock.** *Proceedings of the American Society for Testing Materials.* 1932; **45**: 777–801.

12. Vilarrasa V, Carrera J, Olivella S, *et al.*: **Induced seismicity in geologic carbon storage.** *Solid Earth.* 2019; **10**(3): 871–892.
Publisher Full Text

13. Vilarrasa V, Carrera J: **Geologic carbon storage is unlikely to trigger large earthquakes and reactivate faults through which CO2 could leak.** *Proc. Natl. Acad. Sci.* 2015; **112**(19): 5938–5943.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Rutqvist J, *et al.*: **Fault activation and induced seismicity in geological carbon storage – lessons learned from recent modeling studies.** *J. Rock Mech. Geotech. Eng.* 2016; **8**(6): 789–804.
Publisher Full Text

15. Vilarrasa V, De Simone S, Carrera J, *et al.*: **Unraveling the causes of the seismicity induced by underground gas storage at castor, spain.** *Geophys. Res. Lett.* 2021; **48**(7): 19.
Publisher Full Text

16. Cesca S, *et al.*: **Seismicity at the castor gas reservoir driven by pore pressure diffusion and asperities loading.** *Nat. Commun.* 2021; **12**(1): 4783.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. Williams-Stroud S, *et al.*: **Analysis of microseismicity and reactivated fault size to assess the potential for felt events by CO2 injection in the illinois basin.** *Bull. Seismol. Soc. Am.* 2020; **110**(5): 2188–2204.
Publisher Full Text

18. Bommer JJ, *et al.*: **Control of hazard due to seismicity induced by a hot fractured rock geothermal project.** *Eng. Geol.* 2006; **83**(4): 287–306.
Reference Source

19. Karvounis DC, Gischig VS, Wiemer S: *Towards a real-time forecast of induced seismicity for enhanced geothermal systems.* Pittsburgh, Pennsylvania, United States: American Society of Civil Engineers; 2014.
Publisher Full Text

20. Mignan A, Broccardo M, Wiemer S, *et al.*: **Induced seismicity closed-form traffic light system for actuarial decision-making during deep fluid injections.** *Sci. Rep.* 2017; **7**(1): 13607.
Publisher Full Text

21. Schultz R, Beroza GC, Ellsworth WL: **A strategy for choosing red-light thresholds to manage hydraulic fracturing induced seismicity in north america.** *J. Geophys. Res. Solid Earth.* 2021; **126**(12): 20.
Publisher Full Text

22. Verdon JP, Schultz R, Edwards B: **Tolerable magnitudes for induced seismicity at offshore carbon capture and storage projects.** *Int. J. Greenh. Gas Con.* 2025; **142**: 104335.
Publisher Full Text

23. Ader T, *et al.*: **Design and implementation of a traffic light system for deep geothermal well stimulation in finland.** *J. Seismol.* 2020; **24**(5): 991–1014.
Publisher Full Text

24. Verdon JP, Bommer JJ: **Green, yellow, red, or out of the blue? An assessment of traffic light schemes to mitigate the impact of hydraulic fracturing-induced seismicity.** *J. Seismol.* 2021; **25**(1): 301–326.
Publisher Full Text

25. Verdon JP, Eisner L: **An empirically constrained forecasting strategy for induced earthquake magnitudes using extreme value theory.** *Seismol. Res. Lett.* 2024; 17.
Publisher Full Text

26. Kwiatek G, *et al.*: **Controlling fluid-induced seismicity during a 6.1-km-deep geothermal stimulation in finland.** *Sci. Adv.* 2019; **5**: 12.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Häring MO, Schanz U, Ladner F, *et al.*: **Characterisation of the basel 1 enhanced geothermal system.** *Geothermics.* 2008; **37**(5): 469–495.
Publisher Full Text

28. Grigoli F, *et al.*: **The november 2017 mw 5.5 pohang earthquake: A possible case of induced seismicity in south korea.** *Science.* 2018; **5**(June): 1003–1006.
Publisher Full Text

29. Kim K-I, *et al.*: **Protocol for induced microseismicity in the first enhanced geothermal systems project in pohang, korea.** *Renew. Sust. Energ. Rev.* 2018; **91**: 1182–1191.
Publisher Full Text

30. Ellsworth WL, Giardini D, Townend J, *et al.*: **Triggering of the pohang, korea, earthquake (mw 5.5) by enhanced geothermal system stimulation.** *Seismol. Res. Lett.* 2019; **90**(5): 1844–1858.
Publisher Full Text

31. Gaucher E, *et al.*: **Induced seismicity in geothermal reservoirs: A review of forecasting approaches.** *Renew. Sust. Energ. Rev.* 2015; **52**: 1473–1490.
Publisher Full Text

32. Wiemer S: **Introducing probabilistic aftershock hazard mapping.** *Geophys. Res. Lett.* 2000; **27**(20): 3405–3408.
Publisher Full Text

33. van der Elst NJ, Page MT, Weiser DA, *et al.*: **Induced earthquake magnitudes are as large as (statistically) expected.** *J. Geophys. Res. Solid Earth.* 2016; **121**(6): 4575–4590.
Publisher Full Text

34. Bachmann CE, Wiemer S, Woessner J, *et al.*: **Statistical analysis of the induced basel 2006 earthquake sequence: Introducing a probability-based monitoring approach for enhanced geothermal systems: Probability-based monitoring approach for EGS.** *Geophys. J. Int.* 2011; **186**(2): 793–807.
Publisher Full Text

35. Kroll KA, Brudzinski MR: **Evaluating the aftershock duration of induced earthquakes.** *Bull. Seismol. Soc. Am.* 2024; **114**(2): 895–905.
Publisher Full Text

36. Shapiro SA, Rothert E, Rath V, *et al.*: **Characterization of fluid transport properties of reservoirs using induced microseismicity.** *Geophysics.* 2002; **67**(1): 212–220.
Publisher Full Text

37. Shapiro SA, Krüger OS, Dinske C, *et al.*: **Magnitudes of induced earthquakes and geometric scales of fluid-stimulated rock volumes.** *Geophysics.* 2011; **76**(6): 9.
Publisher Full Text

38. Langenbruch C, Weingarten M, Zoback MD: **Physics-based forecasting of man-made earthquake hazards in oklahoma and kansas.** *Nat. Commun.* 2018; **9**(1): 3946.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

39. Boyet A, De Simone S, Ge S, *et al.*: **Poroelastic stress relaxation, slip stress transfer and friction weakening controlled post-injection seismicity at the basel enhanced geothermal system.** *Commun. Earth Environ.* 2023; **4**(1): 104.
Publisher Full Text

40. Boyet A, De Simone S, Vilarrasa V: **Physics-based modeling to understand and to propose forecasting methods of induced seismicity.** *Seismol. Res. Lett.* 2023; **94**(6): 2666–2678.
Publisher Full Text

41. Boyet A, De Simone S, Vilarrasa V: **To bleed-off or not to bleed-off?.** *Geophys. Res. Lett.* 2024; **51**(23): e2023GL107926.
Publisher Full Text

42. Tangirala SK, Parisio F, Vaezi I, *et al.*: **Effectiveness of injection protocols for hydraulic stimulation in enhanced geothermal systems.** *Geothermics.* 2024; **120**: 103018.
Publisher Full Text

43. Andrés S, Santillán D, Mosquera JC, *et al.*: **Delayed weakening and reactivation of rate-and-state faults driven by pressure changes due to fluid injection.** *J. Geophys. Res. Solid Earth.* 2019; **124**(11): 11917–11937.
Publisher Full Text

44. Andrés S, Santillán D, Mosquera JC, *et al.*: **Hydraulic stimulation of geothermal reservoirs: Numerical simulation of induced seismicity and thermal decline.** *Water.* 2022; **14**(22): 3697.
Publisher Full Text

45. Dieterich J: **A constitutive law for rate of earthquake production and its application to earthquake clustering.** *J. Geophys. Res.* 1994; **99**(B2): 2601–2618.
Publisher Full Text

46. Scholz CH: **On the stress dependence of the earthquake b value.** *Geophys. Res. Lett.* 2015; **42**(5): 1399–1402.
Publisher Full Text

47. Dieterich JH, Richards-Dinger KB, Kroll KA: **Modeling injection-induced seismicity with the physics-based earthquake simulator RSQSim.** *Seismol. Res. Lett.* 2015; **86**(4): 1102–1109.
Publisher Full Text

48. Kroll K, Sherman C: 2023; *ORION: Operational FoRecastIng of INduced seismicity.* National Energy Technology Laboratory - Energy Data eXchange; NETL.
Publisher Full Text

49. Dempsey D, Suckale J: **Physics-based forecasting of induced seismicity at groningen gas field, the netherlands.** *Geophys. Res. Lett.* 2017; **44**(15): 7773–7782.
Publisher Full Text

50. Hill RG, Trugman DT, Weingarten M: **Separating injection-driven and earthquake-driven induced seismicity by combining a fully coupled poroelastic model with interpretable machine learning.** *Geophys. Res. Lett.* 2024; **51**(18): e2024GL109802.
Publisher Full Text

51. Spagnuolo E, Nielsen S, Violay M, *et al.*: **An empirically based steady state friction law and implications for fault stability.** *Geophys. Res. Lett.* 2016; **43**(7): 3263–3271.
PubMed Abstract | Publisher Full Text | Free Full Text

52. Rouet-leduc B, *et al.*: **Machine learning predicts laboratory earthquakes.** *Geophys. Res. Lett.* 2017; 7.
Publisher Full Text

53. Karimpouli S, *et al.*: **Explainable machine learning for labquake prediction using catalog-driven features.** *Earth Planet. Sci. Lett.* 2023; **622**: 118383.
Publisher Full Text

54. Karimpouli S, Kwiatek G, Martínez-Garzón P, *et al.*: **Unsupervised clustering of catalogue-driven features for characterizing temporal evolution of labquake stress.** *Geophys. J. Int.* 2024; **237**(2): 755–771.
Publisher Full Text

55. DeVries PMR, Viégas F, Wattenberg M, *et al.*: **Deep learning of aftershock patterns following large earthquakes.** *Nature.* 2018; **560**(7720): 632–634.
Publisher Full Text

56. Saad OM, *et al.*: **Earthquake forecasting using big data and artificial intelligence: A 30-week real-time case study in china.** *Bull. Seismol. Soc. Am.* 2023; **113**(6): 2461–2478.
Publisher Full Text

57. Qin Y, Chen T, Ma X, *et al.*: **Forecasting induced seismicity in oklahoma using machine learning methods.** *Sci. Rep.* 2022; **12**(1): 9319.
PubMed Abstract | Publisher Full Text | Free Full Text

58. Li Z, Eaton DW, Davidsen J: **Physics-informed deep learning to forecast m_max during hydraulic fracturing.** *Sci. Rep.* 2023; **13**(1): 13133.
PubMed Abstract | Publisher Full Text | Free Full Text

59. Yu P, *et al.*: **Association between injection and microseismicity in geothermal fields with multiple wells: Data-driven modelling of rotokawa, new zealand, and húsmúli, iceland.** *J. Geophys. Res. Solid Earth.* 2023; 32.
Publisher Full Text

60. Mignan A, Rinaldi AP, Lanza F, *et al.*: **A multi-LASSO model to forecast induced seismicity at enhanced geothermal systems.** *Geoenergy Science and Engineering.* 2024; **236**: 212746.
Publisher Full Text

61. Picozzi M, Iaccarino AG: **Forecasting the preparatory phase of induced earthquakes by recurrent neural network.** *Forecasting.* 2021; **3**(1): 17–36.
Publisher Full Text

62. Walter L: **Linuswalter/real-time-forecast-IBDP collection.** *DIGITAL.CSIC.* 2025.
Publisher Full Text

63. Karimpouli S, *et al.*: **Forecasting induced seismicity in enhanced geothermal systems using machine learning: Challenges and opportunities.** *Geophys. J. Int.* 2025; **242**(2): ggaf155.
Publisher Full Text

64. Feng Y, Mignan A, Sornette D, *et al.*: **Investigating injection pressure as a predictor to enhance real-time forecasting of fluid-induced seismicity: A bayesian model comparison.** *Seismol. Res. Lett.* 2023; **94**(2A): 708–719.
Publisher Full Text

65. Bondarenko N, Podladchikov Y, Makhnenko R: **Hydromechanical impact of basement rock on injection-induced seismicity in illinois basin.** *Sci. Rep.* 2022; **12**(1): 15639.
PubMed Abstract | Publisher Full Text | Free Full Text

66. Silva JA, *et al.*: **Mechanisms for microseismicity occurrence due to CO2 injection at decatur, illinois: A coupled multiphase flow and geomechanics perspective.** *Bull. Seismol. Soc. Am.* 2024; **114**(5): 2424–2445.
Publisher Full Text

67. Dando B, *et al.*: **Relocating microseismicity from downhole monitoring of the decatur CCS site using a modified double-difference algorithm.** *Geophys. J. Int.* 2021; **227**(2): 1094–1122.
Publisher Full Text

68. Zaluski W, Lee S-Y: *2020 IBDP Final Static Geological Model Development and Dynamic Modelling*. Schlumberger Carbon Services; 2021.

69. Luu K, Schoenball M, Oldenburg CM, *et al.*: **Coupled hydromechanical modeling of induced seismicity from CO2 injection in the illinois basin.** *J. Geophys. Res. Solid Earth.* 2022; **127**(5): e2021JB023496.
Publisher Full Text

70. Alkhimenkov Y, Räss L, Khakimova L, *et al.*: **Resolving wave propagation in anisotropic poroelastic media using graphical processing units (GPUs).** *J. Geophys. Res. Solid Earth.* 2021; **126**(7): e2020JB021175.
Publisher Full Text

71. Castelletto N, *et al.*: *GEOS simulation framework.* Lawrence Livermore National Laboratory; 2024.
Reference Source

72. Ju X, *et al.*: **Learning CO 2 plume migration in faulted reservoirs with graph neural networks.** *Comput. Geosci.* 2024; **193**: 105711.
Publisher Full Text

73. Pedregosa F, *et al.*: **Scikit-learn: Machine learning in python software.** *J. Mach. Learn. Res.* 2011; **12**: 2825–2830.
Reference Source

74. Breiman L: **Random forests.** *Mach. Learn.* 2001; **45**(1): 5–32.
Publisher Full Text

75. Aki K: **Maximum likelihood estimate of b in the formula log n=a-bM and its confidence limits.** *Bull. Earthquake Res. Inst., Tokyo Univ.* 1965; **43**: 237–239.
Reference Source

76. Clauset A, Shalizi CR, Newman MEJ: **Power-law distributions in empirical data.** *SIAM Rev.* 2009; **51**(4): 661–703.
Publisher Full Text

77. Goebel THW, Kwiatek G, Becker TW, *et al.*: **What allows seismic events to grow big?: Insights from b-value and fault roughness analysis in laboratory stick-slip experiments.** *Geology.* 2017; **45**(9): 815–818.
Publisher Full Text

78. Louppe G: **Understanding random forests: From theory to practice.** *arXiv.* 2014.
Publisher Full Text

79. Finley RJ: **An overview of the illinois basin – decatur project.** *Greenhouse Gases: Science and Technology.* 2014; **4**(5): 571–579.
Publisher Full Text

80. Bauer RA, Carney M, Finley RJ: **Overview of microseismic response to CO2 injection into the mt. Simon saline reservoir at the illinois basin-decatur project.** *Int. J. Greenh. Gas Con.* 2016; **54**: 378–388.
Publisher Full Text

81. Will R, Smith V, Lee D, *et al.*: **Data integration, reservoir response, and application.** *Int. J. Greenh. Gas Con.* 2016; **54**: 389–403.
Publisher Full Text

82. Python Software Foundation: **Python: Version 3.11.9.** 2023.
Reference Source

83. Hunter JD: **Matplotlib: A 2D graphics environment software.** *Computing in Science & Engineering.* 2007; **9**(3): 90–95.
Publisher Full Text

84. Team, T. M. D: **Matplotlib: Visualization with python software.** *Zenodo.* 2024.
Publisher Full Text

85. Goebel T: **Magnitude-distribution.** 2029.
Reference Source

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 22 September 2026

https://doi.org/10.21956/openreseurope.25333.r81082

© **2026 Wang W.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Wendong Wang
Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan, China

Recommendation: Major Revision
This manuscript presents an adaptive machine-learning workflow for near-real-time forecasting of induced seismicity during geological carbon storage operations. The topic is timely and relevant, and the continuously updated random-forest framework combined with the shifted moving-window strategy is potentially useful for operational seismic-risk management. The manuscript is generally well organized, and the availability of the underlying data and source code is an important strength.
However, several methodological issues require substantial clarification and, in some cases, additional analysis before the main claims regarding near-real-time forecasting performance can be fully supported. I therefore recommend major revision.

1. The temporal availability of the input variables should be clarified carefully.
Section 5.1 defines w_1 as the previous observation window and w_0 as the future forecasting window. However, Table 1 appears to include several variables from $w0^{w_0}$, including $p_{well}(w_0)$, $N(w_0)$, $M_w(w_0)$, $M_{w,diff}(w_0)$, and the b-value, as model inputs. The authors should explicitly clarify which variables are actually available when each forecast is issued. If these w_0 labels are notation errors, the relevant equations, tables, figures, and text should be corrected consistently. If information from the forecast window is actually used as input, the potential for information leakage should be carefully addressed.
2. The RF workflow should be benchmarked against simpler forecasting baselines.
The manuscript mainly compares the hydraulic-only RF model with the RF model incorporating both hydraulic and seismological variables, while their overall performances are rather similar. This comparison alone does not establish whether the RF framework provides additional forecasting skill beyond simpler alternatives. I suggest including at least one reasonable baseline, such as a persistence forecast or a simple autoregressive/statistical model, using the same sequential forecasting setup and evaluation metrics.
3. The practical meaning of “near-real-time” should be supported quantitatively.
The RF model is repeatedly retrained, and a two-stage hyperparameter grid search is performed

as the dataset evolves. The authors should report the computational time required for model updating and forecasting, for example the approximate runtime per timestep or for the complete case study, and discuss whether this computational cost is compatible with the intended near-real-time application.
4. The interpretation of uncertainty and feature importance should be made more cautious.
The uncertainty range is estimated from the standard deviation of predictions from individual trees, which reflects ensemble dispersion but does not necessarily represent calibrated predictive uncertainty. Similarly, RF feature importance indicates predictive contribution within the fitted model and should not be directly interpreted as evidence of physical or causal mechanisms. The corresponding terminology and interpretations in the Abstract, Results, and Discussion should therefore be revised or better justified.
5. Several aspects of interpretation and presentation should be clarified.
The manuscript should use consistent terminology for seismicity rate, event count N, and cumulative event count N_c. In addition, the selection of the 8-day forecasting window would be better described as a trade-off among forecast error, R^2, and temporal resolution rather than as a unique optimum. The authors should also more clearly acknowledge that the results are derived from a single site and that the selected predictors and window length may not directly transfer to other geological settings. Finally, a careful editorial check is recommended to correct remaining typographical and grammatical issues.
Overall assessment
The manuscript addresses an important problem and proposes a potentially useful adaptive forecasting framework. However, the temporal definition of the model inputs, the lack of benchmarking against simpler approaches, and the current support for the near-real-time and uncertainty-related claims require further attention. Addressing the above issues would substantially strengthen the methodological robustness and interpretation of the study. I therefore recommend major revision.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Geological carbon storage (GCS/CCUS), subsurface multiphase flow, reservoir monitoring, geo-energy engineering, and data-driven/machine-learning applications in subsurface systems.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 12 June 2026

https://doi.org/10.21956/openreseurope.25333.r75243

© 2026 Zhang J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Jize Zhang
Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

The paper presents a methodologically sound workflow.

I have a few comments:

1. If I understand correctly, Table 1 shows that the RF model using only hydraulic features (RF_H) performs nearly identically to the model using both hydraulic and seismological features (RF_HS): SMAPE of 39.15% vs 37.15% for N, and MAE of 0.47 vs 0.48 for M_{w,max}. This near-equivalence is troubling as it could suggest that the model may be primarily learning a pressure-to-seismicity mapping that could be achieved with simpler analytical approaches. The authors attribute this to the "rather contained" seismicity at IBDP, but this seems a bit incomplete and circular.

2. The paper does not benchmark the RF against even simple baselines: persistence forecasts (tomorrow looks like today), linear autoregressive models, or the analytical Shapiro framework that is repeatedly cited. Without such comparisons, it is impossible to determine whether the RF's complexity is justified or whether a simpler model would perform comparably. The comparison between RF_H and RF_HS is useful but insufficient; the real question is whether any ML model outperforms a well-calibrated physics-based or statistical baseline.

3. The hyperparameter search is described as a two-stage grid search performed at every timestep. Given that the RF is retrained at each of >1000 daily timesteps, this represents much excessive computational cost, yet no runtime analysis is provided. For a framework claiming near-real-time capability, wall-clock performance should be reported and discussed.

4. The paper mentions UQ via standard deviation of individual tree predictions, but this is known to be poorly calibrated for RF models and does not constitute rigorous uncertainty quantification.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Machine Learning; Forecasting; Uncertainty Quantification; Natural Hazard; GCS.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.