Induced Seismicity Forecasting with Uncertainty Quantification: Application to the Groningen Gas Field

Hojjat Kaveh*[1,2], Pau Batlle[3], Mateo Acosta[2], Pranav Kulkarni[4], Stephen J. Bourne[5], and Jean Philippe Avouac[1,2]

Abstract

Reservoir operations for gas extraction, fluid disposal, carbon dioxide storage, or geothermal energy production are capable of inducing seismicity. Modeling tools exist for seismicity forecasting using operational data, but the computational costs and uncertainty quantification (UQ) pose challenges. We address this issue in the context of seismicity induced by gas production from the Groningen gas field using an integrated modeling framework, which combines reservoir modeling, geomechanical modeling, and stress-based earthquake forecasting. The framework is computationally efficient thanks to a 2D finite-element reservoir model, which assumes vertical flow equilibrium, and the use of semianalytical solutions to calculate poroelastic stress changes and predict seismicity rate. The earthquake nucleation model is based on rate-and-state friction and allows for an initial strength excess so that the faults are not assumed initially critically stressed. We estimate uncertainties in the predicted number of earthquakes and magnitudes. To reduce the computational costs, we assume that the stress model is true, but our UQ algorithm is general enough that the uncertainties in reservoir and stress models could be incorporated. We explore how the selection of either a Poisson or a Gaussian likelihood influences the forecast. We also use a synthetic catalog to estimate the improved forecasting performance that would have resulted from a better seismicity detection threshold. Finally, we use tapered and nontapered Gutenberg–Richter distributions to evaluate the most probable maximum magnitude over time and account for uncertainties in its estimation. Although we did not formally account for uncertainties in the stress model, we tested several alternative stress models, and found negligible impact on the predicted temporal evolution of seismicity and forecast uncertainties. Our study shows that the proposed approach yields realistic estimates of the uncertainties of temporal seismicity and is applicable for operational forecasting or induced seismicity monitoring. It can also be used in probabilistic traffic light systems.

Cite this article as Kaveh, H., P. Batlle, M. Acosta, P. Kulkarni, S. J. Bourne, and J. P. Avouac (2023). Induced Seismicity Forecasting with Uncertainty Quantification: Application to the Groningen Gas Field, *Seismol. Res. Lett.* **XX**, 1–18, doi: 10.1785/0220230179.

Supplemental Material

Introduction

Stress changes in the earth's lithosphere resulting from activities such as oil and gas extraction or geothermal energy production are capable of triggering or inducing seismicity (Ellsworth, 2013). Much progress has been made recently in the development of a physics-based and computationally efficient model that maps the relation between fluid injection (or extraction) to stress changes and seismicity (Bourne *et al.*, 2014; Bourne and Oates, 2017; Dempsey and Suckale, 2017, 2023; Langenbruch *et al.*, 2018; Candela *et al.*, 2019; Zhai *et al.*, 2019; Dahm and Hainzl, 2022; Kühn *et al.*, 2022; Meyer *et al.*, 2022; Smith *et al.*, 2022; Kim and Avouac, 2023). To create these models, it is necessary to combine a reservoir model, which describes pore pressure diffusion in the subsurface, a geomechanical model, which describes the induced strain and stress, and a seismicity model, which relates the seismicity to stress changes. Such models predict the rate of

---

1. Mechanical and Civil Engineering, California Institute of Technology, Pasadena, California, U.S.A., https://orcid.org/0000-0003-3272-3417 (HK); 2. Geology and Planetary Science Division, California Institute of Technology, Pasadena, California, U.S.A., https://orcid.org/0000-0002-0098-7912 (MA); 3. Computing and Mathematical Science, California Institute of Technology, Pasadena, California, U.S.A., https://orcid.org/0000-0003-4886-058X (PB); 4. Electrical Engineering, California Institute of Technology, Pasadena, California, U.S.A., https://orcid.org/0000-0002-1461-0948 (PK); 5. Shell Global Solutions International B.V., Amsterdam, The Netherlands, https://orcid.org/0000-0003-2925-8411 (SJB)

*Corresponding author: hkaveh@caltech.edu

earthquakes with magnitude above some threshold in space and time.

In reality, the seismicity rate is not observable. The quantity that we can observe is an earthquake catalog, which can be seen as a stochastic realization of the predicted seismicity rate. As a result, the uncertainty in the expected number of earthquakes stems from the epistemic uncertainty of the seismicity rate predicted by the model and from the aleatoric uncertainty due to the realization of the stochastic process. Because of the multiple model layers and the computational cost, optimizing model parameters and quantifying the uncertainties are challenging tasks; however, it is critical that uncertainties be assessed so that the modeling can be used to help plan operations or guide decisions during operations, for example, through a traffic light systems (Baisch *et al.*, 2019; Verdon and Bommer, 2021).

Here, we seek to quantify the uncertainty associated with the predicted number of earthquakes and the maximum magnitude arising from both the epistemic and aleatoric terms. Different approaches have been used in the previous studies. For example, Candela, Pluymaekers, *et al.* (2022) used an ensemble of stress and seismicity models to estimate epistemic uncertainties but did not consider aleatoric uncertainties. A number of studies have used Bayesian methods to estimate epistemic uncertainties (Bourne *et al.*, 2018; Candela, Goncalves Machado, *et al.*, 2022). Some studies (Candela, Goncalves Machado, *et al.*, 2022) discuss epistemic and aleatoric sources of uncertainties but do not provide a method to estimate quantitatively the uncertainty, and the associated confidence level on the predicted number of earthquakes resulting from both the sources. Regarding the prediction of magnitudes, Bourne and Oates (2020) and Zöller and Hainzl (2023) considered the epistemic uncertainty associated with the parametrization of the frequency–magnitude distribution but did not consider the uncertainty of their seismicity model. Dempsey and Suckale (2023) did not consider the uncertainties of the parameters of the magnitude–frequency distribution as well as the parameters in their source model. A method to account for both aleatoric and epistemic sources of uncertainty is missing in the earthquake forecasting literature.

Another point of discussion in the literature regards the choice of the likelihood function that should be used to the merit of a seismicity forecast. Because earthquake catalogs can be considered as the superposition of independent events and events triggered by the previous events, one might consider using either a Poisson (Ogata, 1988; Bourne and Oates, 2017; Dempsey and Suckale, 2017) or a Gaussian (Heimisson *et al.*, 2022; Smith *et al.*, 2022; Dempsey and Suckale, 2023) likelihood. The first one is appropriate in principle if the observed catalog contains a negligible proportion of aftershocks, whereas the second is more relevant in the case of a significant proportion of aftershocks.

We therefore propose an algorithm to account for both aleatoric and epistemic uncertainty that is independent of the stress and seismicity model, and can be used with different likelihood functions. We discuss and quantify these different sources of uncertainty, and propose a general method to determine the confidence intervals on the earthquake number forecast. Although our focus is on estimating earthquake numbers, we also describe how our method can be used to estimate probabilities of observing earthquakes of a certain magnitude in a given time window. This requires a model describing the frequency–magnitude distribution of earthquakes and the associated uncertainty (Zoeller and Holschneider, 2016; Shcherbakov *et al.*, 2019).

We use the novel prior-free methodology in Bajgiran *et al.* (2021) for uncertainty quantification (UQ) together with the Broyden–Fletcher–Goldfarb–Shamno (BFGS) method (Martins and Ning, 2021) for optimizing likelihood functions. Unlike the traditional Bayesian methods, this approach allows us to estimate uncertainties that do not depend on the choice of a prior probability distribution of the model parameters. We apply this approach to the Groningen gas field in the Netherlands, where production from the largest onshore gas field in Western Europe has induced significant and well-documented induced seismicity (Bourne *et al.*, 2014; Dost *et al.*, 2017; Spetzler and Dost, 2017). This case study is particularly suitable for testing the methodology due to the available high-quality information on reservoir characteristics and operations (Jager and Visser, 2017), as well as the well-documented seismicity (Dost *et al.*, 2017; Spetzler and Dost, 2017; Smith *et al.*, 2020). Our modeling framework integrates a reservoir model, which describes the diffusion of pore pressure in the subsurface, a geomechanical model, which describes the induced strain and stress changes both within and outside the reservoir, and a seismicity model, that relates the seismicity rate to the stress changes (Heimisson *et al.*, 2022; Meyer *et al.*, 2022; Smith *et al.*, 2022). The reservoir and geomechanical models were calibrated using pressure well data and surface subsidence measurements (Fig. 1a).

The modeling workflow calculates the seismicity rate in both time and space. The observable seismicity catalog is considered as a stochastic realization of the forecasted seismicity rate. In this study, we differentiate between epistemic sources of uncertainty, which arise from uncertainties in the model parameters, and aleatoric sources of uncertainty, which stem from the stochastic nature of the seismicity process. In principle, both the reservoir and the geomechanical models contribute to the epistemic source of uncertainties, because they jointly determine the stress model used to drive seismicity (Candela, Goncalves Machado, *et al.*, 2022; Meyer *et al.*, 2022). However, we consider here the stress model to be true, because it simplifies the presentation of the method and speeds up the calculation time allowing for the relatively wide range of tests in this study. However, we do present results obtained with different stress forecasts. Integrating these epistemic uncertainties is in principle straightforward but computationally costly.

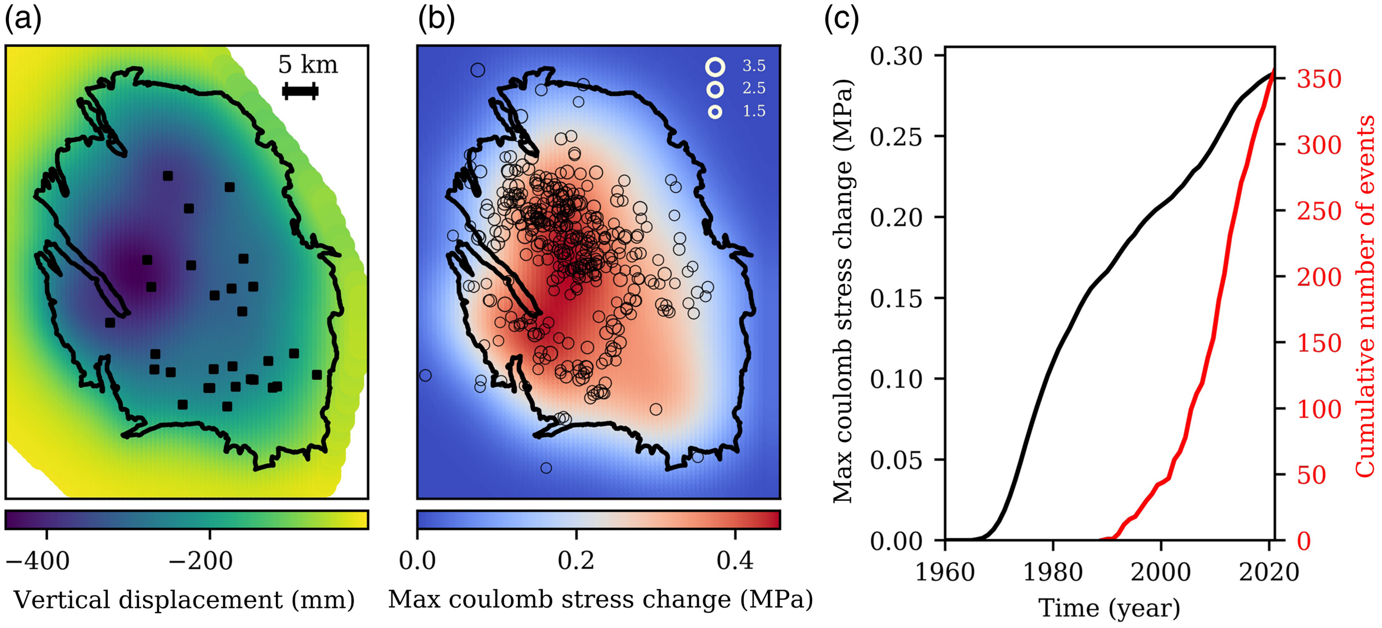

Figure 1. The Groningen field and geomechanical model results. (a) Cumulative vertical displacement since the onset of gas production as of 2019. Black rectangles denote the locations of the wells. (b) Maximum Coulomb stress change at a nominal elevation of 10 m since the onset of gas production and epicenters of $M \geq 1.5$ earthquakes (circles) as of 2021. (c) Time evolution of the spatially averaged maximum Coulomb stress change, and the cumulative number of events with a magnitude greater or equal to 1.5. The color version of this figure is available only in the electronic edition.

A step in that direction is presented by Candela, Pluymaekers, *et al.* (2022), who inverted jointly the parameters of the earthquake nucleation model and the Poisson coefficient associated with the geomechanical model. A full account of epistemic uncertainties would require in principle to include uncertainties on all the parameters of the reservoir model and the geomechanical model, most importantly the reservoir compressibility derived from subsidence measurements.

This article is organized as follows: the Methods section will describe the data and modeling framework used in the study. This section will also introduce two alternative likelihood functions, Gaussian and Poisson, to quantify the fit of the model to the observed number of earthquakes. The algorithm to quantify uncertainties and estimate earthquake magnitude probabilities will also be described. In the Results and Discussion section, we will apply and test the proposed UQ method to the forecast of the number of induced earthquakes at Groningen. The performance of the Poisson and Gaussian likelihood functions will be compared and earthquake magnitude probabilities will be estimated. Using a synthetic catalog, we will assess how much better our seismicity forecasting performance could have been if we had a better seismic network. This analysis can contribute to constraining the cost and benefit balance for investments in improved monitoring strategies. Finally, we will conclude the article by summarizing our findings and discussing their implications for seismicity forecasting, monitoring, and hazard assessment.

Methods

In this section, we first introduce our integrated modeling approach for seismicity forecasting. Second, we define the likelihood functions used in this article. Then, we discuss our algorithm for UQ. Finally, we present a method for estimating the maximum moment magnitude, which leverages the techniques developed in this study.

Integrated modeling framework and seismicity observations

We use a framework that combines reservoir and geomechanical modeling to predict seismicity rates in time and space. The reservoir model (Meyer *et al.*, 2022) and geomechanical model (Smith *et al.*, 2022) allow us to calculate stress changes within and around the reservoir, using data on well extraction rates and pressure. Details on the calculation of the spatial stress distribution are available in Appendix A1, available in the supplemental material to this article. The Coulomb stress can be calculated based on some a priori assumptions on the fault geometry or for the “optimal” orientation yielding the maximum Coulomb stress change. We consider optimally oriented faults, because both options yield very similar seismicity forecasts (Smith *et al.*, 2022). The hypocentral depth distribution shows a peak right above the reservoir top in the anhydrite caprock (Smith *et al.*, 2020). To reduce the computational cost, following Smith *et al.* (2022), we assume a nominal depth of 10 m above the reservoir top. We show in Appendix A2, available in the supplemental material to this article, results obtained with different stress models, including the possibility of the earthquakes nucleating within the reservoir as assumed in most studies (Bourne and Oates, 2017; Dempsey and Suckale, 2017, 2023; Candela *et al.*, 2019). All the model tested yield very similar temporal forecast, but they predict different spatial distributions. Throughout this paper, we use the stress model that gives the best fit to the spatial distributions of earthquakes, but the results presented here are independent of that particular choice (See Appendix 2 for other stress models). The reservoir is represented by cuboids with a vertical extent corresponding to the reservoir thickness and a 500 m × 500 m horizontal extent. The cuboids were designed to match as closely as possible the 3D geometry of the reservoir (Burkitov *et al.*, 2016; Smith *et al.*, 2022). Using the Coulomb stress history, a seismicity model is then used that can calculate the seismicity rate in time and space.

We use the seismicity catalog from the Royal Dutch Meteorological Survey (KNMI; see Data and Resources) to assess the model prediction. See Appendix A3, available in the supplemental material to this article, for more details on the seismicity data. Initially, there were only a few seismic stations around Groningen, but the observation of induced seismicity has led to the deployment of a denser network, resulting in an improvement of the magnitude of completeness of the catalog from about M_c 1.5 in the early 90s to about 0.5 by 2014 (Dost *et al.*, 2017; Smith *et al.*, 2022). For this study, unless specified otherwise, we use the $M_w \geq 1.5$ events from the seismicity catalog of KNMI from the year 1990 to the year 2021. Figure 1b shows the spatial distribution of the change of maximum Coulomb stress (i.e., the maximum value for all possible fault orientations) as of 2021. The seismicity is observed in the zones of higher Coulomb stress changes. However, the consistency between the time evolution of the spatially averaged maximum Coulomb stress changes, and the observed seismicity is not obvious at first glance (Fig. 1c). Despite extraction rates ramping up in the late 60s, causing rapid and measurable subsidence and hence an early increase of Coulomb stress, the seismicity only ramped up in the 1990s (Fig. 1c). The seismicity rate exhibited an accelerating increase initially, despite a relatively steady annual extraction rate. After the production rate was reduced by about 50% following the M_w 3.6 Huizinge earthquake in 2012 (Smith *et al.*, 2022), the seismicity rate decreased gradually within a few years.

The lag between the onset of seismicity and the Coulomb stress change history can be interpreted as an indication that the medium around the reservoir was initially not critically stressed (Bourne and Oates, 2017). This interpretation is compatible with the Groningen area being tectonically quiet since the early Cretaceous (Jager and Visser, 2017).

Alternatively, the lag could result from the time-dependent earthquake nucleation process (Dieterich, 1994; Candela *et al.*, 2019). Because both the explanations may have merit, we use the threshold rate-and-state (RS) model (Heimisson *et al.*, 2022), which includes both the effects. The noninstantaneous nucleation process is represented using the RS friction formalism (Dieterich, 1994) but relaxing the assumption of faults being critically stressed. Heimisson *et al.* (2022) showed that, for a population of faults below steady-state (initially "relaxed"), the rate of seismicity $R(x_1,x_2,t)$ (the rate of earthquakes per unit time and area) depends on the spatial stress history $\Delta S(x_1,x_2,t)$ according to

$$\frac{R(x_1,x_2,t)}{r} = \frac{f_1(x_1,x_2,t)}{f_2(x_1,x_2,t)} \quad t > t_b,$$
$$\frac{R(x_1,x_2,t)}{r} = 0 \quad t < t_b, \tag{1}$$

with

$$f_1(x_1,x_2,t) = \exp\left(\frac{\Delta S(x_1,x_2,t) - \Delta S_c}{A\sigma_0}\right),$$
$$f_2(x_1,x_2,t) = \frac{1}{t_a}\int_{t_b}^{t} f_1(x_1,x_2,t^{'})\mathrm{d}t^{'} + 1, \tag{2}$$

in which x_1 and x_2 are the positions in the east–west and north–south directions, r is the background seismicity rate, and $A\sigma_0$ is a characteristic frictional-stress parameter with A being a constitutive parameter related to the direct effect of RS. t_a is the characteristic time of the nucleation process. ΔS_c is the threshold Coulomb stress, analog to the strength excess in the Coulomb failure model (i.e., the Coulomb stress change needed to initiate fault slip). t_b is a parameter that varies in space and defines the time when at position (x_1,x_2) the stress $(\Delta S(x_1,x_2,t))$ reaches ΔS_c. This parameter is introduced to simplify the expression, but it can be eliminated from equation (1). Equation (1) is a discontinuous function that would make parameter inference challenging. To overcome this issue, we approximated equation (1) with a continuous function. The details are described in Appendix A4, available in the supplemental material to this article.

The vector of model parameters is hereafter defined as

$$\mathbf{u} = [r \quad t_a \quad A\sigma_0 \quad \Delta S_c] \in \mathbb{R}_+^4. \tag{3}$$

The vector of seismicity observation $\mathbf{y}_{M_c}$ that depends on the cutoff magnitude (M_c) is defined by

$$\mathbf{y} = [y_1 \quad y_2 \quad \dots \quad y_T] \in \mathbb{R}^T, \tag{4}$$

in which y_j is the total number of earthquakes in jth time bin, and T is the number of time bins. Finally, the vector of the modeled seismicity rate, which comes from the discretized version of equation (1), is

$$\mathbf{h}(\mathbf{u};\Delta\mathbf{S}) = [h_1(\mathbf{u};\Delta\mathbf{S_1}) \quad h_2(\mathbf{u};\Delta\mathbf{S_2}) \quad \dots \quad h_T(\mathbf{u};\Delta\mathbf{S_T})] \in \mathbb{R}^T, \tag{5}$$

in which $h_j(\mathbf{u};\Delta\mathbf{S_j})$ is the predicted rate of events in the jth time bin. $\Delta\mathbf{S} \in \mathbb{R}^{D_{x_1} \times D_{x_2} \times T}$ is the discretized stress distribution in time and space with D_{x_1} and D_{x_2} as the number of elements in the x_1 and x_2 direction and T as the number of time bins. $\Delta \mathbf{S_j} \in \mathbb{R}^{D_{x_1} \times D_{x_2} \times j}$ is the history of stress distribution up to jth time bin. For details on how to discretize equation (1) and find $\mathbf{h}(\mathbf{u},\Delta\mathbf{S})$, see Appendix A4, available in the supplemental material to this article.

In this article, we use yearly time bins: the average stress in a year at each grid point and the cumulative number of events in each year as the observed seismicity. Because the relationship between the stress variations and the rate of seismicity is nonlinear, one should in principle take into account subannual variations, however, to limit the computational cost we neglect these variations. In addition, Acosta *et al.* (2023) show that the impact of seasonal variations of stress on the predicted annual earthquake rate is negligible. Taking these variations into account can, however, help tighten the constraints on the seismicity model parameters as we show in Appendix A5, available in the supplemental material to this article, using a simple example.

Likelihood functions

In this section, we introduce likelihood functions that are used in this article and discuss their implication in induced seismicity forecasting. We only use temporal-likelihood functions, because we have assumed that the stress model is true, and model parameters are uniform. With these assumptions, equation (1) is only dependent on the space through the spatial distribution of $\Delta\mathbf{S}$ (assumed to be true). Using a spatiotemporal likelihood would be useful if we were testing different stress models, as they predict different spatial distributions of seismicity as we show in Appendix A2, available in the supplemental material to this article. The UQ algorithm presented in the next section is general and does not assume any stress, or seismicity model, and can be used both for temporal and spatial likelihoods.

It is common practice to consider that earthquakes can be triggered either by the previous earthquakes or by other processes such as tectonic loading or stress changes due to subsurface reservoir operations. The most adequate likelihood function depends on the proportion of earthquakes triggered by other earthquakes in the observed seismicity catalog. If the proportion is very small, the seismicity is generally well described by a nonhomogenous Poisson process, meaning that the events are independent but triggered at a rate that can vary in time depending on the loading rate (Ogata, 1988). If the proportion of events triggered by the previous earthquakes is large, then the nonhomogeneous Poisson model is less appropriate. In the context of faults governed by RS friction, coseismic stress changes result in earthquake clusters and aftershock sequences, but the total number of events is unchanged over a time scale much longer than the characteristic nucleation time (Heimisson and Segall, 2018). In that case, if earthquakes are binned over a timestep that is large compared to the typical duration of aftershock sequences, it is probably more adequate to assume that the number of events per time bin results from a Gaussian process (Heimisson *et al.*, 2022).

In the case of seismicity induced at Groningen, Trampert *et al.* (2022) have found that less than 7% of the $M_w > 0.5$ earthquakes are triggered by other earthquakes. Another study suggests possibly as much as 27% of $M_w > 1.3$ events being aftershocks. In any case, an inspection of the catalog shows that aftershock sequences have probably a short duration (smaller than a year; Post *et al.*, 2021). The clustering effect due to earthquake interactions is therefore probably smoothed out when only yearly bins are considered, as is the case when using the Gaussian log likelihood (GLL). However, it could be a source of bias when the Poisson log likelihood (PLL) is used.

In the continuous-time form, the nonhomogeneous PLL can be written (Dempsey and Suckale, 2017) as

$$\mathrm{PLL}(\mathbf{y}|\mathbf{u};\Delta\mathbf{S}) = \sum_{l=1}^{N} \log(\lambda(\mathbf{u};\tau_l)) - \int_0^{\tau_N} \lambda(\mathbf{u};\tau')\mathrm{d}\tau', \quad (6)$$

in which $\lambda(\mathbf{u};\tau)$ is the spatially summed seismicity rate continuously predicted by the model in time (τ), and τ_l is the time when the lth event is observed. N is the total number of observed events. The PLL is discretized as

$$\mathrm{PLL}(\mathbf{y}|\mathbf{u};\Delta\mathbf{S}) \simeq \langle \mathbf{y}, \log(\mathbf{h})\rangle - \sum_{i=1}^{T} \mathbf{h}_i \Delta t, \quad (7)$$

in which $\langle , \rangle$ is the inner product in $\mathbb{R}^T$, and Δt is the size of discretized time, and it is equal to one year in our calculations. In addition, $\log(\mathbf{h})$ is a vector in $\mathbb{R}^T$ for which kth element is $\log(h_k)$.

The GLL, as defined in Heimisson *et al.* (2022) writes

$$\mathrm{GLL}(\mathbf{y}|\mathbf{u};\Delta\mathbf{S}) = -\frac{1}{2}|\mathbf{y} - \mathbf{h}(\mathbf{u};\Delta\mathbf{S})|^2_\Gamma. \quad (8)$$

By $|.|_\Gamma$ we denote the weighted norm defined by $|v|^2_\Gamma = v^\intercal \Gamma^{-1} v$ for a vector $v \in \mathbb{R}^T$, in which T is the number of data points in $\mathbf{y}$. Equation (8) assumes that the difference between the observed and predicted number of events in each year is normally distributed. When all the observations are independent and identically distributed, Γ is a diagonal matrix. We can borrow the concept from a Poisson process, for which the variance equals the mean, to approximate the elements on the main diagonal of Γ (variance) as the number of events in each year. However, in that case because there are some years with zero number of events, Γ would not be invertible. To overcome this issue, Dempsey and Suckale (2023) set the variance at each time bin equal to the number of observed events in that time bin and put a threshold variance for when number of observations is zero, trying to mimic a Poisson process. However, for a Poisson process, the variance is equal to the true rate at each time bin and not the observed rate. Here, we use $\Gamma = \kappa I$, in which κ is the average number of events in $\mathbf{y}$, and $I \in \mathbb{R}^{T\times T}$ is the identity matrix. The additional assumptions about the variance of the Gaussian likelihood both in our work and in Dempsey and Suckale (2023) is one of its weak points compared to the Poisson likelihood, which may cause inaccuracy in estimating maximum likelihood and uncertainties.

Because of the small number of events at the onset of induced seismicity, the Poisson and Gaussian likelihoods can still yield different results, even though the Poisson likelihood approaches the Gaussian likelihood as the number of events becomes large. Alternatively, if aftershocks are present, the Poisson likelihood may lead to biased results (Bourne *et al.*, 2018). Hence, we consider both likelihood functions and analyze the impact of aftershocks on their performance.

Uncertainty quantification and parameter inference

There are several sources of uncertainty to consider when forecasting the future number of earthquakes using our modeling framework. They include the uncertainties in stress distribution calculations (which come from uncertainties in extraction (or injection) rates, pore pressures, and stress model), the uncertainties in the seismicity model parameters, and the uncertainty due to the stochastic nature of the seismicity. We propose an algorithm to bound the range within which the number of events in each time bin is expected at or above a chosen confidence level, accounting for both epistemic and aleatoric sources of uncertainty.

Different methods such as the Markov chain Monte Carlo (Dempsey and Suckale, 2017; Smith *et al.*, 2022) and genetic algorithms (Dempsey and Suckale, 2023) exist to infer the model parameters ($\mathbf{u}$) from the data using different likelihood functions (such as equations 8 or 7) and use them to forecast future earthquakes. Bayesian methods further assume a prior $p(\mathbf{u})$, which is then combined with the data $\mathbf{y}$ and the likelihood $p(\mathbf{y}|\mathbf{u})$ to generate a posterior distribution $p(\mathbf{u}|\mathbf{y})$; their UQ heavily relies on the choice of prior. To overcome this challenge, we use a "prior-free" methodology (Bajgiran *et al.*, 2021) for UQ, along with the BFGS method (Martins and Ning, 2021; their algorithm 4.7) for optimizing the likelihood function. The BFGS algorithm, which uses both first- and second-order derivatives of the cost function, starts from uniformly selected random initial points in the parameter space. One should be careful with the optimization algorithm, as the Poisson- and Gaussian-likelihood functions with the complicated seismicity model and limited data available can potentially have multiple local maxima.

Within this framework, we can separately account for different sources of uncertainty involved in predicting the future events. First, there is uncertainty present in the model parameters ($\mathbf{u}$), which exhibit changes in the expected rate of events (epistemic). Second, there is uncertainty in the stochastic process (aleatoric), which arises from the model of equation (1), in which the rate of events is modeled, rather than the number of events. Both of these uncertainties must be estimated and combined to obtain a final UQ bound. To find the UQ bounds, we first consider a likelihood function with unknown physical parameters $\mathbf{u} \in \mathbb{R}^q$ (in which q is four in this article), and observations from $\mathbf{y} \sim p(\cdot|\mathbf{u})$ ($\mathbf{y}$ is distributed as $p(\cdot|\mathbf{u})$), in which p is either the temporal Gaussian or the Poisson likelihood in this article. Our method is not limited to these specific distributions. For example, p could also represent spatial–temporal likelihoods in a broader sense. Our first step is to determine a UQ region for the rate of future events at the kth time, denoted by $h_k(\mathbf{u}) \in \mathbb{R}$. We consider a region in parameter space around the maximum-likelihood estimate (MLE) for the observed data $\mathbf{y}$, for some $0 \le \alpha \le 1$.

$$\Theta_{\mathbf{y}}(\alpha) := \left\{ \mathbf{u} \in \mathbb{R}^q : \frac{p(\mathbf{y}|\mathbf{u})}{\sup\limits_{\mathbf{u}' \in \mathbb{R}^q} p(\mathbf{y}|\mathbf{u}')} \ge \alpha \right\}, \tag{9}$$

in which $\sup_{\mathbf{u}' \in \mathbb{R}^q} p(\mathbf{y}|\mathbf{u}') = p(\mathbf{y}|\mathbf{u}^{\mathbf{MLE}})$ with $\mathbf{u}^{\mathbf{MLE}}$ as the MLE of the model parameters. The value of α is a function of the level of epistemic uncertainty that we would like to consider. Based on theorem 4.1 in Bajgiran *et al.* (2021), in the asymptotic regime of a large sample of model parameters under regularity conditions over the likelihood function (which are satisfied by the Gaussian and Poisson likelihoods considered here), α relates to the epistemic probability coverage $(1 - \xi)$ with the following equation:

$$\alpha = \exp\left(-\frac{1}{2} \psi_q (1 - \xi) \right), \tag{10}$$

in which ψ_q is the quantile function (inverse of the cumulative distribution function) of a χ_q^2 random variable [with degrees of freedom (q) equal to the dimensionality of the parameter vector $\mathbf{u}$]. The choice of α is made so that the selected interval has the desired probability coverage $1 - \xi$ (on the rate of events).

After maximizing the likelihood function, we find set $\Theta_{\mathbf{y}}(\alpha)$ by selecting all the model parameters $\mathbf{u}$, which have the condition of equation (9). Then, we obtain a confidence interval for $h_k(\mathbf{u})$ as the next equation. The following equation only gives the confidence bound on the rate of events at kth time and not the number of events.

$$\left[\min_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h_k(\mathbf{u}'), \max_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h_k(\mathbf{u}') \right]. \tag{11}$$

We define h_k^- and h_k^+ as the minimum and the maximum rate of events, respectively, with probability $(1 - \xi)$ at kth time as the following:

$$\begin{aligned} h_k^- &= \min_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h_k(\mathbf{u}'), \\ h_k^+ &= \max_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h_k(\mathbf{u}'). \end{aligned} \tag{12}$$

So far, we showed how to put bounds on the *rate* of events (h_k) at kth time (equation 11); the true rate is in the obtained UQ bound with probability $1-\xi$. However, a full UQ bound on the *number* of events should also take into account the stochasticity of the process. To that end, we add aleatoric bounds on top of the epistemic bounds to the quantity of interest function h_k. Because h_k is the rate of events at kth time, the number of events at that time follows a Poisson distribution with its mean equal to h_k. To approximate the confidence limit ($[\phi_-(\mathbf{u}),\phi_+(\mathbf{u})]$) of a Poisson process (number of events), at kth time bin, with a Poisson mean of h_k (rate of events), we use the next equation, which has shown satisfactory performance for different ranges of h_k, especially when $h_k > 4$ (Patil and Kulkarni, 2012):

$$\begin{aligned}\varphi_-(\mathbf{u}) &= \frac{1}{2}\psi_{2h_k}(\gamma/2)\\ \varphi_+(\mathbf{u}) &= \frac{1}{2}\psi_{2(h_k+1)}(1-\gamma/2),\end{aligned} \tag{13}$$

in which ψ_{2h_k} and $\psi_{2(h_k+1)}$ are the quantile function of the χ^2 distribution with $2h_k$ and $2(h_k+1)$ degrees of freedom. When h_k is large (>30), one can approximate the confidence interval of a Poisson process (equation 13) by the confidence interval of a Gaussian process (Brown *et al.*, 2001). When the rate of events (h_k) is smaller than 4, equation (13) is conservative, and one can use other confidence limits introduced in (Patil and Kulkarni, 2012). Here, $1-\gamma$ is the aleatoric coverage, probability for which γ is defined by the practitioner.

Using elements in the set $\Theta_{\mathbf{y}}(\alpha)$, the confidence interval of the number of events, considering both aleatoric and epistemic uncertainties with probability at least $(1-\xi)(1-\gamma)$ is given by:

$$I_k = [\min_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} \varphi_-(\mathbf{u}'), \max_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} \varphi_+(\mathbf{u}')], \tag{14}$$

in which I_k is the confidence interval at kth time bin. The practitioner chooses two parameters ξ and γ that regulate the epistemic and aleatoric uncertainty, respectively. If $\Theta_{\mathbf{y}}(\alpha)$ contains the true parameter with probability $1-\xi$, and the worst-case $1-\gamma$ interval is created on top (equation 14), the combined interval contains the true number of events with a probability of at least $(1-\xi)(1-\gamma)$. There is the possibility of optimizing over ξ and γ such that we obtain the tightest interval for the number of events for a fixed $(1-\xi)(1-\gamma)$.

One practical point to be mentioned is that since the quantile function of χ^2 monotonically increases, we simplify equation (14) as

$$I_k = \left[\frac{1}{2}\psi_{2h_k^-}\left(\frac{\gamma}{2}\right), \frac{1}{2}\psi_{2(h_k^+ +1)}\left(1-\frac{\gamma}{2}\right)\right], \tag{15}$$

in which $\psi_{2h_k^-}$ and $\psi_{2(h_k^+ +1)}$ are the quantile function of χ^2 distribution with $2h_k^-$ and $2(h_k^+ + 1)$ degrees of freedom. We use equation (15) as the final rule for the confidence interval of the number of events at kth point in time, with probability at least $(1-\xi)(1-\gamma)$ (see Appendix A6, available in the supplemental material to this article, for mathematical details). Based on Lemma 1 in Appendix A6, available in the supplemental material to this article, the probability that the number of events lies inside the confidence range in equation (15) is at least $(1-\xi)(1-\gamma)$. As we will see in the simulations, the confidence interval given in equation (15) is a conservative interval, because we have assumed the worst-case scenario for the model parameters with $(1-\xi)$ probability coverage. We summarize the method in Algorithm 1.

ALGORITHM 1

UQ algorithm

```
1. Inputs:
(a) Likelihood model $p(\cdot|\mathbf{u})$
(b) $q = \dim(\mathbf{u})$
(c) Data $\mathbf{y}$
(d) Significance levels $\xi$ and $\gamma$, such that an interval of coverage probability $(1-\xi)(1-\gamma)$ is obtained
2. Find $\mathbf{u}^{\mathbf{MLE}} = \mathrm{argmax}_{\mathbf{u}} p(\cdot|\mathbf{u})$ via an optimization algorithm
3. Set $\alpha = \exp(-\frac{1}{2}\psi_q(1-\xi))$
4. Find the set $\Theta_{\mathbf{y}}(\alpha) := \{\mathbf{u} \in \mathbb{R}^q : \frac{p(\mathbf{y}|\mathbf{u})}{p(\mathbf{y}|\mathbf{u}^{\mathbf{MLE}})} \geq \alpha\}$,
5. For $k$th timestep, find $h_k^-$ and $h_k^+$
    $h_k^- = \min_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h_k(\mathbf{u}')$
    $h_k^+ = \max_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h_k(\mathbf{u}')$
6. The confidence bound on the number of events with probability $(1-\gamma)(1-\xi)$ in $k$th timestep is given by:
    $\left[\frac{1}{2}\psi_{2h_k^-}\left(\frac{\gamma}{2}\right), \frac{1}{2}\psi_{2(h_k^+ +1)}\left(1-\frac{\gamma}{2}\right)\right]$.
```

To help the reader better understand the algorithm and also show why the confidence bound is conservative, we have included some simple pedagogical examples in Appendix A5, available in the supplemental material to this article. The codes of Algorithm 1 applied to those examples are available online (see Data and Resources).

Earthquake magnitude probability

Earlier, we have described how to forecast the number of earthquakes while taking uncertainties into account. To estimate the probabilities of induced earthquake magnitudes, it is essential

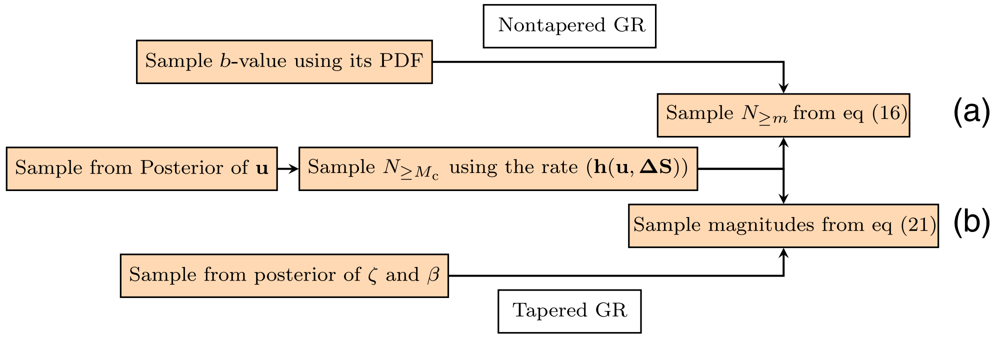

Figure 2. Flowchart illustrating the steps involved in sampling from a Gutenberg–Richter (GR) distribution considering both the aleatoric and the epistemic sources of uncertainty. First, we sample from the posterior distribution of the model parameters $\mathbf{u}$. Then, we sample from a nonhomogeneous Poisson process with the rate $\mathbf{h}(\mathbf{u},\Delta\mathbf{S})$ to generate a realization of $N_{\geq m}$. For the nontapered GR (a), we use the number of events and a realization of the b-value to sample from equation (16) and obtain the magnitudes of the events. For the tapered GR (b), we use the number of events and the realizations of ζ and β. Then, the magnitude of each event is determined randomly using equation (21). For both nontapered and tapered GR distributions, we repeat the process multiple times. The color version of this figure is available only in the electronic edition.

to use a model that describes their magnitude–frequency distribution. Earthquakes in any tectonic setting tend to follow the Gutenberg–Richter (GR) magnitude–frequency distribution, which has different variants. Some of these variants assume truncation at the maximum magnitude beyond which earthquakes are considered physically impossible, whereas others have a gradual tapering (mostly exponential). Such options have been studied in the context of the Groningen gas field (Bourne and Oates, 2020; Dempsey and Suckale, 2023). In this study, we demonstrate our framework using both nontapered and tapered distributions.

Nontapered GR. The number of events with a magnitude greater or equal to m, $N_{\geq m}$ in nontapered GR is given by

$$N_{\geq m} = N_{\geq M_c} 10^{(-b(m-M_c))}, \tag{16}$$

where M_c is the cut-off magnitude, b describes the relative proportion of small-to-large earthquakes, and $N_{\geq M_c}$ is the total number of events with a magnitude greater than or equal to M_c. Here, we consider b and $N_{\geq M_c}$ as random variables with some probability distribution functions, which will be discussed shortly.

There are many methods in the literature for finding the b-value (Marzocchi and Sandri, 2003). We have used the "b-positive (b^+)" method (van der Elst, 2021), because it is insensitive to variations in the magnitude of completeness of the earthquake catalog. Their estimator is calculated only based on positive differences in magnitude between successive events according to

$$b^+ = \frac{1}{(\bar{m}' - M_c')\ln(10)} \qquad m' \geq M_c', \tag{17}$$

in which m' is the difference between the magnitude of two successive earthquakes. M_c' is a constant number that satisfies the $M_c' \geq 2\delta$ condition, in which 2δ is the discretization level of the moments in the catalog. To use equation (17), one should use a moving window of events and find those events for which $m' \geq M_c'$, and then average all m's in the window to find $\bar{m}'$. The standard deviation of the estimated b^+ can be approximated by $\frac{b^+}{\sqrt{N(m' \geq M')}}$, in which $N(m' \geq M')$ is the number of events that satisfy $m' \geq M'$ condition in the time window.

To find a realization of $N_{\geq M_c}$, we sample from a nonhomogeneous Poisson process with rate $\mathbf{h}(\mathbf{u}; \Delta\mathbf{S})$. $\mathbf{h}(\mathbf{u}; \Delta\mathbf{S})$ is the vector of seismicity rates for a desired time span, in which $\mathbf{u}$ is sampled from the posterior distribution of the model parameters. The flowchart of Figure 2a summarizes how one can sample from the nontapered GR distribution while considering both the epistemic and the aleatoric sources of uncertainty.

Using equation (16), the most probable maximum magnitude ($\hat{M}_{\max}$) that would be observed in a sample of size $N \geq M_c$ (Van der Elst *et al.*, 2016) is

$$\hat{M}_{\max} = M_c + \frac{1}{b}\log_{10}(N_{\geq M_c}). \tag{18}$$

It should be noted that due to the distribution's heavy tail, the expected maximum magnitude (mean of the PDF of $M_{\max}$) is larger than the most probable maximum magnitude (mode of the PDF of $M_{\max}$). Throughout this article, we use $\hat{M}_{\max}$ for the most probable maximum magnitude for a nontapered GR distribution.

It is straightforward to find the probability of exceeding any magnitude over any chosen time duration as we describe subsequently. In the limit of large $N(M \geq M_c)$, the confidence level q on the most probable maximum magnitude can be calculated using the following equation:

$$\hat{M}_q = \hat{M}_{\max} - \frac{1}{b}\log_{10}(-\ln(q)). \tag{19}$$

As a result, the probability of having an event with a magnitude greater than $\hat{M}_q$ is $P(\hat{M}_{\max} > \hat{M}_q) = 1 - q$. We can write this quantity as a function of $\hat{M}_{\max}$ and the b-value,

$$P(\hat{M}_{\max} > \hat{M}_q) = 1 - \exp(-10^{b(\hat{M}_{\max} - \hat{M}_q)}). \tag{20}$$

Tapered GR. Based on the tapered GR distribution, the probability of an event with a seismic moment greater than $\mathcal{M}$ is given by the following equation (Kagan, 2002; Bourne and Oates, 2020):

$$P(\geq \mathcal{M}|\mathcal{M} \geq \mathcal{M}_m) = \left(\frac{\mathcal{M}}{\mathcal{M}_m}\right)^{-\beta} e^{-\zeta(\frac{\mathcal{M}}{\mathcal{M}_m}-1)}, \tag{21}$$

in which $\mathcal{M}_m$ is a constant that will be defined shortly, β and ζ are parameters in this model that in principle can be a function of stress or stressing rate. In this study, we consider β and ζ as independent random variables, and we will sample from their posterior distribution. The flowchart of Figure 2b summarizes the sampling scheme from a tapered GR distribution while considering both the epistemic and the aleatoric uncertainties. To generate random realizations from equation (21), the inverse transform sampling method is used (Steinbrecher and Shaw, 2008). Posterior distributions of β and ζ are found by maximizing the following log-likelihood function:

$$l = \sum_{i=1}^{N}\left(\log\left(\beta + \zeta\frac{\mathcal{M}_i}{M_m}\right) - (1+\beta)\log\frac{\mathcal{M}_i}{M_m} - \zeta\left(\frac{\mathcal{M}_i}{M_m} - 1\right)\right), \tag{22}$$

in which N is the number of events. We can relate the seismic moment ($\mathcal{M}$) to the moment magnitude (M) using the following equation:

$$\log_{10}\mathcal{M} = (c + dM), \tag{23}$$

in which $c = 9.1$ and $d = 1.5$. The value $\mathcal{M}_m$ can be found using the following equation:

$$\log\mathcal{M}_m = \left(c + d\left(M_c - \frac{1}{2}\Delta M\right)\right)\log 10, \tag{24}$$

in which M_c is the magnitude of completeness, and ΔM is the size of binned intervals. Based on equation (21), the probability of an event with seismic moment smaller than $\mathcal{M}$ is given by

$$P(< \mathcal{M}|\mathcal{M} \geq \mathcal{M}_m) = 1 - P(\geq \mathcal{M}|\mathcal{M} \geq \mathcal{M}_m). \tag{25}$$

As a result, assuming the magnitude of events is independent and identically distributed according to equation (21), the probability that $N_{\geq M_c}$ events having seismic moment smaller than $\mathcal{M}$ can be found by the following equation:

$$P(< \mathcal{M}, X = N_{\geq M_c}|\mathcal{M} \geq \mathcal{M}_m) = (1 - P(\geq \mathcal{M}|\mathcal{M} \geq \mathcal{M}_m))^{N_{\geq M_c}}, \tag{26}$$

in which X counts the occurrence of a specific outcome. Therefore, the probability of having at least one event out of $N_{\geq M_c}$ realizations with seismic moment greater than $\mathcal{M}$ is given by

$$P(> \mathcal{M}, X \geq 1|\mathcal{M} \geq \mathcal{M}_m) = 1 - (1 - P(\geq \mathcal{M}|\mathcal{M} \geq \mathcal{M}_m))^{N_{\geq M_c}}. \tag{27}$$

Results and Discussion

Uncertainty in the forecast of earthquake numbers

We first comment on the results obtained when the earthquake catalog is split into a training set and a test set (Fig. 3). The training set includes data up to 2008 (shown in white), whereas the test set includes data from 2009 to 2021 (shown in gray in Fig. 3). The maximum Poisson-likelihood model parameters ($\mathbf{u}^{\mathrm{MLE}}$) are estimated using the training set. Once we estimate $\mathbf{u}^{\mathrm{MLE}}$, we run the forward model to obtain the maximum-likelihood estimate (MLE) for the seismicity rate ($\mathbf{h}^{\mathrm{MLE}}$). The observed seismicity follows quite closely the forecast, matching well the peak of seismicity in 2012–2014 and the subsequent decline (Fig. 3). It is noteworthy that the previous models (Candela *et al.*, 2019; Richter *et al.*, 2020) trained on the pre-2009 data were unable to capture the turnover in 2014.

We now seek to bound the ≥90% confidence level ($(1-\gamma)(1-\xi) = 0.9$) for the number of events in each year, in which $(1-\gamma)$ and $(1-\xi)$ are the confidence level in the Poisson process (aleatoric uncertainty) and the model parameters (epistemic uncertainty), respectively. As stated earlier, for a given confidence level ($(1-\gamma)(1-\xi)$), the values of $(1-\gamma)$ and $(1-\xi)$ are not uniquely determined. The confidence bounds slightly change when changing these two parameters while keeping the multiplication constant. We have analyzed the sensitivity of selecting ξ and γ on confidence interval for specific confidence levels in Appendix A7, available in the supplemental material to this article. The balance between these two uncertainties helps to have the tightest interval of the number of events for a given confidence probability. We found that having $1-\gamma$ close to $1-\xi$ is a good first guess. In the case of Figure 3, we selected the combination $1-\gamma = 0.96$ and $1-\xi = 0.94$.

The red lines in Figure 3 show the variability at the 94% confidence level of the forecasted seismicity rate ($\mathbf{h}^{94\%}$). The ≥90% confidence bound on the number of events (black lines in Fig. 3) is obtained by adding the uncertainty due to the Poisson process using equation (15).

Figure 4a compares the observed yearly rate (blue line) of seismicity with the various confidence intervals obtained with our method (red to green lines). We can use that graph to compare the difference between the observed and the predicted yearly number of events with the uncertainties in the model prediction estimated from our method. The comparison shows that, as expected, the misfits lie well within the estimated bounds on the confidence interval (Fig. 4b). The uncertainties are largely overestimated at lower confidence levels and less so at higher confidence levels. For example, 71% of the catalog data points are within the ≥25% confidence interval, and

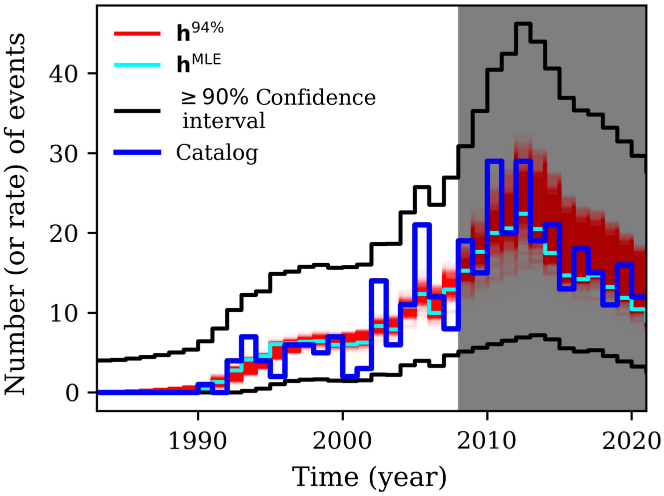

Figure 3. Quantifying uncertainty and forecast up to the year 2021 using the Poisson likelihood: red lines show the seismicity rate with 94% probability, the cyan line is the maximum-likelihood estimate (MLE) of the rate of events, and black lines show ≥90% confidence interval of the number of events with $1-\gamma = 0.96$ and $1-\xi = 0.94$. Seismicity data (with $M_c = 1.5$, blue line) up to the year 2008 is used (white), and the rate is predicted for the years 2009–2021 (gray). The color version of this figure is available only in the electronic edition.

97% of the catalog data points are within the ≥90% confidence interval.

One can understand the results in Figure 4b with the following two extreme cases. First, imagine that we have an infinite number of observations in the training set. In this case, the model parameters are very well constrained, and the uncertainty of the forecast comes primarily from the Poisson process. This is verified in Example 4 of the Appendix A5, available in the supplemental material to this article, for which, in this case, the points in a similar plot as Figure 4b now lie around the $y = x$ line. In the second case, imagine there are very few observations in the training set. Now the uncertainty of the forecast comes from both the lack of knowledge of the model parameters and the aleatoric uncertainty of the Poisson process. As a result, for the same confidence level, the uncertainty range of the forecast is much larger and the aleatoric variations that are observed in the data will remain well within the bounds estimated from our method that considers both the epistemic and aleatoric uncertainties. In other words, the points will fall below the dashed line as seen in Figure 4b.

Other simple examples are shown in Appendix A5, available in the supplemental material to this article, which should help the reader understand the concept.

Comparing likelihoods

In this part, we compare the performance of the nonlocal Gaussian and Poisson likelihood. The effect of likelihood is important in the context of induced seismicity forecasting, because, at the beginning of the seismic activity, the number of events is small (and the Gaussian and Poisson likelihood can be far from equivalent) and clustering is hard to assess. We also use our modeling framework to assess how better seismic monitoring would have helped to predict more accurately and earlier induced seismicity at Groningen.

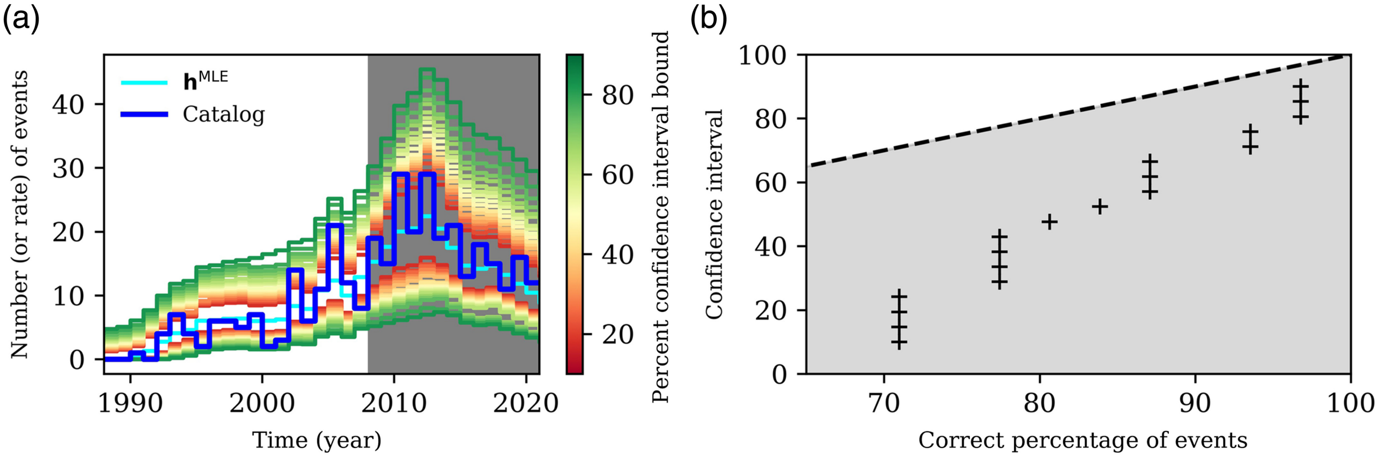

Figure 4. Confidence-bound performance: (a) the white region is used as the training set and the gray region is used as the test set. Dark blue is the seismicity data (with $M_c = 1.5$), and $\mathbf{h}^{\mathrm{MLE}}$ in cyan is the MLE of the rate of events. Different percent confidence interval bounds for the number of events are plotted as well. (b) The percentage of events (x axis) that lie inside a certain confidence interval bound (y axis). The dashed line plots $y = x$ line. Assuming the data comes from a nonhomogenous Poisson Process, and the practitioner's only uncertainty is the Poisson process (aleatoric) and not the model parameters (epistemic), points should lie more or less around this line. In reality, we have epistemic uncertainty, and the confidence bounds are larger than the confidence bounds that only come from a Poisson process. In this figure, for the sake of simplicity, ξ is considered to be equal to γ. The color version of this figure is available only in the electronic edition.

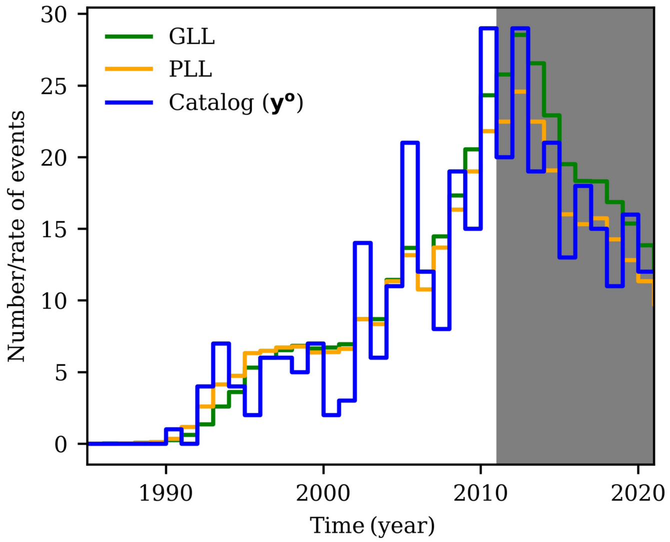

Figure 5. An example of a comparison of the MLE models obtained with the Poisson and Gaussian likelihoods. The white region is used for the training set, and the gray region is used for testing. The blue line is the seismicity data with $M_c = 1.5$, and the orange and the green lines are the MLEs for the rate of events based on the Poisson and Gaussian likelihoods, respectively. In this instance of training-test sets, the Poisson likelihood yields a forecast in better agreement with the observations over the test period than the Gaussian likelihood. The color version of this figure is available only in the electronic edition.

Figure 5 presents a comparison between the MLE models obtained with Poisson and Gaussian likelihoods. In this particular training-test scenario, the Poisson likelihood produces a forecast that is more closely aligned with the observed data during the test period. To assess more broadly the performance of the two likelihoods, we performed systematic tests on data sets with different sizes and different proportions of aftershocks.

Based on Heimisson *et al.* (2022), the overall magnitude of completeness in the KNMI catalog from 1991 to 2021 is almost 1.5. Therefore, we initially used a cut-off magnitude $M_c = 1.5$. To augment the number of events, we also tested using a cut-off magnitude $M_c = 1.2$. To augment further the catalog size and include a known proportion of aftershocks, we also generated a synthetic catalog using the epidemic-type aftershock sequence (ETAS) model. This catalog has almost 10 times more events than the KNMI catalog and would correspond to a magnitude of detection of about $M_c \simeq 0.5$. Therefore, it mimics the catalog that would have been obtained if the seismic network in operation since 2014 had been deployed much earlier. The catalog that is generated based on the ETAS model (denoted by $\mathbf{y}^{\mathbf{ETAS}}_{\mathbf{M_c \approx 0.5}}$) contains 24% of aftershocks—a value consistent with the estimated proportion of aftershocks (Post *et al.*, 2021). For further information on how the synthetic catalog ($\mathbf{y}^{\mathbf{ETAS}}_{\mathbf{M_c \approx 0.5}}$) was created and to view an example of the forecast and UQ algorithm utilized in it, please refer to Appendix A3, available in the supplemental material to this article. By lowering the cut-off magnitude, we should end up with a larger proportion of aftershocks. Therefore, we should expect the Gaussian likelihood to perform better.

For each data set ($\mathbf{y}$), we use a moving boundary between the training set and the test set (Fig. 6a). The moving boundary enables us to use training sets of different sizes. Because we work on different catalogs that have different numbers of events each year, we define a prediction error that is normalized by the number of events as

$$\epsilon(\mathbf{y}^o_{\text{test}}, \mathbf{h}_{\text{test}}) = \sqrt{\frac{1}{N_{\text{test}}} \left\langle \frac{\mathbf{y}^o_{\text{test}} - \mathbf{h}_{\text{test}}}{\mathbf{y}^o_{\text{test}}}, \frac{\mathbf{y}^o_{\text{test}} - \mathbf{h}_{\text{test}}}{\mathbf{y}^o_{\text{test}}} \right\rangle}, \tag{28}$$

in which $\mathbf{y}^o_{\text{test}}$ and $\mathbf{h}_{\text{test}}$ are the vectors of the observed seismicity and the MLE of the predicted seismicity rate in the test set, respectively. N_{test} is the number of data points in the test set. By $(\frac{\mathbf{y}^o - \mathbf{h}}{\mathbf{y}^o})$ we mean element-wise division, and $\langle , \rangle$ is the inner product in $\mathbb{R}^{N_{\text{test}}}$.

Figure 6 shows the performance of the likelihood functions for different data sets and different sizes of training sets. At the first glance, there is a reasonable trend of error reduction as the last year that is used in the training set gets larger. For the data sets with $M_c = 1.5$ and 1.2 (these are from the measured seismicity catalog), the Poisson likelihood generally has a smaller error than the Gaussian likelihood. For the synthetic catalog corresponding to $M_c \approx 0.5$, the Gaussian likelihood yields a better fit to the test set. The large size of the catalog and the large proportion of aftershocks make the Gaussian likelihood more appropriate in that case. In all three cases, the Poisson likelihood has a more stable error that there are fewer jumps in the error of its prediction.

It is important that in comparing the two likelihood functions, the Gaussian likelihood has an additional disadvantage compared to the Poisson likelihood: it has one more parameter, which is its variance. We have assumed equal variance in each observation ($\Gamma = \kappa I$), with κ being the average number of events per year. This is only a first-order approximation that is used in this article for convenience and consistency with the alternative Poisson process. Both our approximation of the variance of the Gaussian likelihood and the one introduced in (Dempsey and Suckale, 2023) can potentially add bias into the MLE and the UQ.

Figure 6b can also be used to assess how better seismic monitoring would have helped detect events earlier and predict more accurately induced seismicity at Groningen. For this purpose, we can compare the error in the forecast for the synthetic catalog ($\mathbf{y}^{\mathbf{ETAS}}_{\mathbf{M_c \approx 0.5}}$) with the error in the forecast of the measured seismicity catalog ($\mathbf{y}^{\mathbf{o}}_{\mathbf{M_c = 1.5}}$ or $\mathbf{y}^{\mathbf{o}}_{\mathbf{M_c = 1.2}}$). As an example, the accuracy of a forecast based on the training set up to 1999 for $\mathbf{y}^{\mathbf{ETAS}}_{\mathbf{M_c \approx 0.5}}$ data set is almost equal to the accuracy of a forecast

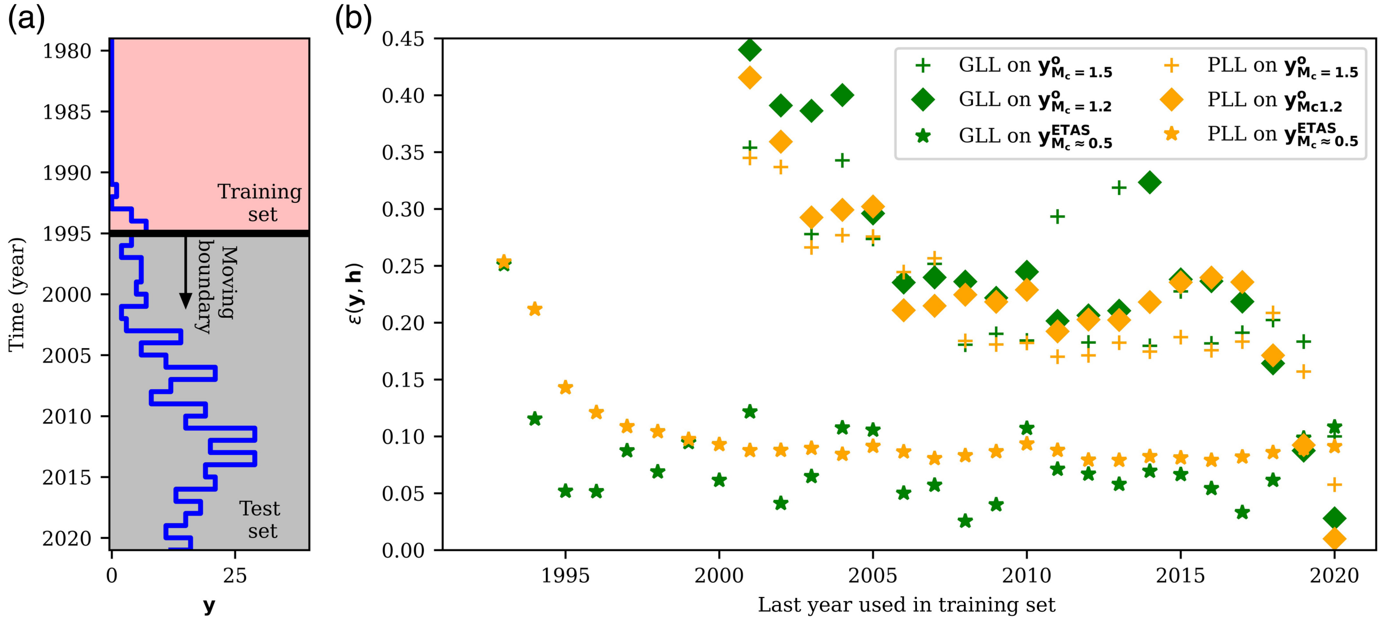

Figure 6. Comparing the performance of Poisson log-likelihood (PLL) and Gaussian log-likelihood (GLL) on different catalogs using different combinations of training and testing sets. (a) The pink region serves as the training set, whereas the gray region is the test set, and different boundaries between the two are examined (b) normalized error as a function of the last year that is used in the training set. The green and orange markers correspond to GLL and PLL, respectively, whereas the plus marker represents $\mathbf{y}^{\mathbf{o}}_{\mathbf{M_c=1.5}}$, the diamond stands for $\mathbf{y}^{\mathbf{o}}_{\mathbf{M_c=1.2}}$, and the star stands for the catalog based on the epidemic-type aftershock sequence (ETAS) model with $M_c \approx 0.5$ ($\mathbf{y}^{\mathbf{ETAS}}_{\mathbf{M_c\approx 0.5}}$). The color version of this figure is available only in the electronic edition.

based on the information up to 2019 for $\mathbf{y}^{\mathbf{o}}_{\mathbf{M_c=1.5}}$. This result suggests that if we had a seismic network with $M_c = 0.5$, we could have reached the same level of accuracy 20 yr earlier.

We note that the errors are smaller when using the catalog with $M_c = 1.5$ compared to the catalog with $M_c = 1.2$, although there are more events in the latter. We interpret this observation as an indication that having an accurate estimate of the magnitude of completeness is important for the performance of the forecast.

It should be noted that the inversions (for both likelihood functions) show a strong trade-off between the background seismicity rate (r) and the characteristic time (t_a). For more information about the inverted model parameters see Appendix A8, available in the supplemental material to this article.

Prospective forecast of earthquake numbers and magnitudes

In this section, we show how our integrated framework and UQ methodology can be used to quantify induced earthquake magnitude probabilities for the Groningen gas field until the year 2030. We use both nontapered and tapered GR distributions. To quantify magnitude probabilities for both distributions, we have followed the steps in Figure 2. To forecast the seismicity after 2021 (the gray region in Fig. 7), we have assumed the "cold winter" scenario for gas extraction (Nederlandse Aardolie Maatschappij [NAM], 2013). Using observations up to the year 2021, the posterior distribution of model parameters ($\mathbf{u}$) is found using PLL. Then, based on the "cold winter" scenario, the time series of stress distribution $\mathbf{\Delta S}$ is calculated. Using this information, we can find the seismicity rate $\mathbf{h}(\mathbf{u},\mathbf{\Delta S})$ for up to the year 2030 (Fig. 7).

The b-value might vary systematically in space, time, or with the stress level (Scholz, 2015; Gulia and Wiemer, 2019; Bourne and Oates, 2020; Muntendam-Bos and Grobbe, 2022). The time evolution of the b-value is calculated using equation (17) and a moving window of 400 events. Figure 8 does not show any monotonic trend that would suggest a dependence on the stress level or time. The b-value seems correlated with the seismicity rate, but the physical and statistical significance of the correlation are not obvious. For simplicity, we assume that the b-value is stationary. We construct a probability density function (PDF) from the distribution obtained by considering successive time windows (Fig. 8). By sampling from this PDF, we find realizations of the b-value for our analysis.

For the tapered GR distribution, to find the posterior distribution of β and ζ, the catalog up to 2021 is used with uniform priors ($0.3 \le \beta \le 1$ and $0 \le \zeta \le 1$). The MLEs for β and ζ are $\beta = 0.62$ and $\zeta = 1.3 \times 10^{-3}$, respectively, with $\mathcal{M}_m = 1.9 \times 10^{11}$ N · m.

We use equations (20) and (18) for the nontapered and equation (27) for the tapered distribution to determine the probability of exceeding a certain magnitude ($P(\hat{M}_{\max} > \hat{M}_q)$). This probability depends on random parameters for both nontapered and tapered GR distributions, and realizations of those parameters are required to accurately calculate it. These

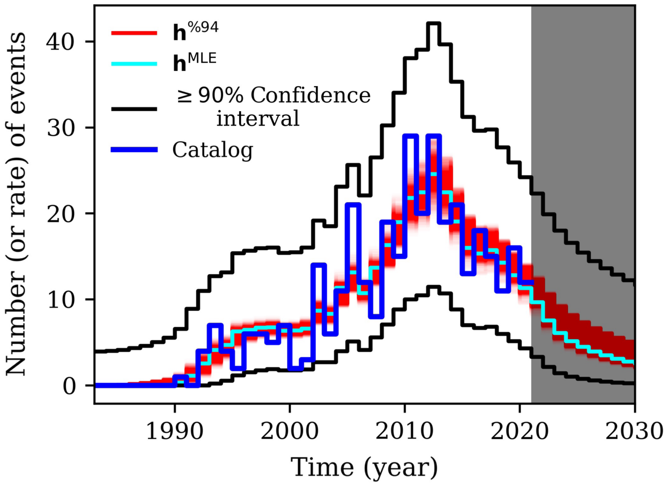

Figure 7. The forecast of the number (or rate) of events for years 2022–2030, the white region is used as the training set to forecast the seismicity in the gray region. The blue line is the seismicity data ($\mathbf{y}^{\mathbf{o}}_{\mathbf{M_c=1.5}}$), cyan is the predicted rate of events based on Poisson likelihood, and red lines are the predicted rate of events with a 94% confidence on the model parameters (epistemic). Black lines are the confidence interval of the number of events (epistemic + aleatoric) with probability at least 0.9, and are generated using Algorithm 1 with $1 - \xi = 0.94$ and $1 - \gamma = 0.96$. The color version of this figure is available only in the electronic edition.

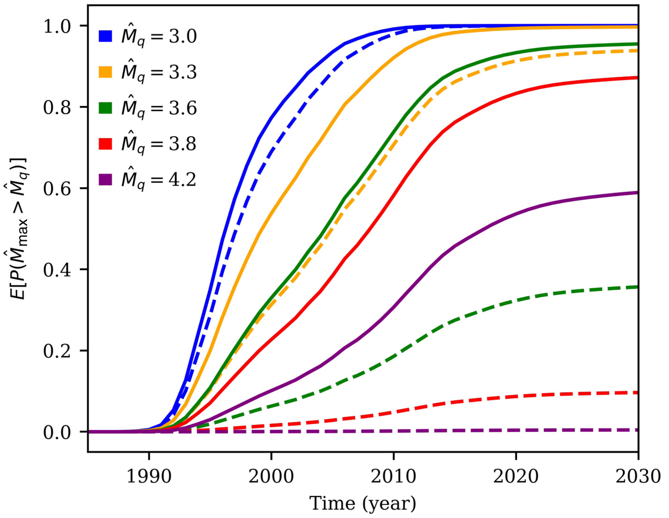

Figure 9. Evolution of the empirical average of the probability of the most probable maximum magnitude ($\hat{M}_{\max}$) being higher than specific magnitudes ($\hat{M}_q$) for different values of $\hat{M}_q$. The solid lines are for the nontapered GR distribution, and the dashed lines are for the tapered GR distributions. The color version of this figure is available only in the electronic edition.

random variables include the model parameter ($\mathbf{u}$), the number of events that is a random realization of a nonhomogenous Poisson process with the rate $\mathbf{h}(\mathbf{u},\Delta\mathbf{S})$, and the b-value for nontapered GR distribution, as well as β and ζ for tapered distribution. By generating multiple realizations from all of these random variables, we find an empirical average for $P(\hat{M}_{\max} > \hat{M}_q)$ for different $\hat{M}_q$. $P(\hat{M}_{\max} > \hat{M}_q)$ is an increasing function of the total number of events, and $P(\hat{M}_{\max} > \hat{M}_q)$ also increases because the total number of events increases over time. Figure 9 illustrates the empirical average evolution of $P(\hat{M}_{\max} > \hat{M}_q)$ over time. Because the event rate has been decreasing after 2012, the slope of $E[P(\hat{M}_{\max} > \hat{M}_q)]$ has also been declining for the 2012–2030 interval. The choice of a frequency–magnitude distribution type, whether tapered or nontapered, has a significant impact on the probability of surpassing a certain magnitude. As a case in point, according to Figure 9, the probability of having an

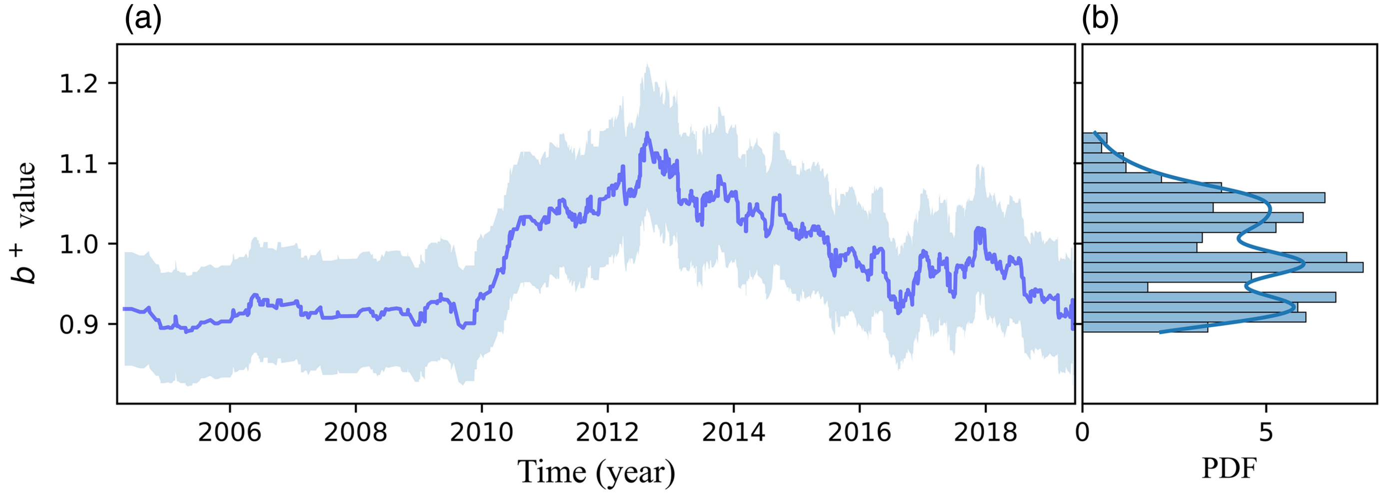

Figure 8. (a) Time series of the b^+ and its 68% confidence interval and (b) probability distribution function (PDF) of b^+. The color version of this figure is available only in the electronic edition.

event with magnitude greater than 4.2 until the year 2030 is ~0.6 based on the nontapered model, but it is almost zero with the tapered model. Determining which distribution is more suitable for the Groningen gas field (Bourne and Oates, 2020; Varty *et al.*, 2021; Muntendam-Bos and Grobbe, 2022; Dempsey and Suckale, 2023) is beyond the scope of this article. Our estimate of the probability of exceeding a certain magnitude is close but slightly larger than the probabilities estimated in Dempsey and Suckale (2023). For instance, according to their analysis using a tapered distribution, the likelihood of an event (up to the year 2030) with a magnitude exceeding 3.6 is about 0.1. In our analysis, this probability is closer to 0.3. The difference between the two estimates could stem from the omission in Dempsey and Suckale (2023) of the epistemic uncertainty of the earthquake rate model, as well as the uncertainty related to parameters of the magnitude–frequency distribution. Another contributing factor might be the utilization of different physical models in these investigations.

Realizations of the nontapered and tapered GR distributions are obtained using the workflow described in Figure 2. The nontapered and tapered GR distributions are displayed in Figure 10 (a and c, respectively). In Figure 10b and 10d, we have plotted the PDF of the $M_{\max}$ (up to 2030) in which we have denoted the most probable maximum magnitude ($\hat{M}_{\max}$), and its 97th and 3rd percentiles as well as the observed maximum magnitude up to 2021. The [3rd, 97th] range of $M_{\max}$ for nontapered distribution up to 2030 is [3.5,5.8] and is [3.3,4] for tapered distribution. The maximum observed magnitude in the field (up to 2021) lies close to the 3rd percentile for the nontapered GR distribution (up to 2030), whereas it is close to the most probable maximum magnitude (up to 2030) for the tapered distribution.

To quantify earthquake magnitude probabilities, we have assumed that the distribution of the b-value in nontapered and β and ζ in tapered GR distributions are stationary in time and uniform in space. This assumption might not be valid, especially for the times and locations that have large fluctuations in fluid extraction. In addition, we have not assigned any uncertainties to the stress model. The validation test shows that the uncertainty bounds are consistent with the observation. However, this might not be necessarily correct in the prospective scenario. The reservoir model ignores the response of the local aquifers to the pressure depletion (Meyer *et al.*, 2022). As a result, the model tends to overestimate the pressure depletion toward the end of the simulation. If this trend continues in the future, it means that our model might be overestimating compaction and the Coulomb stress changes going into the future scenario. Another caveat is that we have assumed a purely poroelastic response of the medium. The assumption is probably a correct approximation over the historical period, because the model fits well with the observed subsidence (Smith *et al.*, 2019). However, there is no guarantee that it will hold true in the future given the possibility of rate-dependent rheology of the reservoir itself and possible viscous flow of the salt layers outside the reservoir (Pruiksma *et al.*, 2015). This might be another cause for our model to overestimate Coulomb stress changes going into the future.

One limitation of our study is that we have not considered the uncertainties associated with the reservoir and the geomechanical models. In that regard, we present in Appendix A2, available in the supplemental material to this article, the results obtained with seven other different stress models. Despite the fact that they predict different spatial distributions of seismicity, they all yield very similar temporal forecasts and uncertainties. Therefore, although we have not tested the full range of possible stress models that could be consistent with data (wellhead pressure, flow rates, and subsidence), this observation supports the claim that it would probably not impact the earthquake forecast much. However, it is important that the best-fitting model parameters differ substantially between the different models, most notably $A\sigma_0$ and ΔS_c (see Appendix A2, available in the supplemental material to this article). This is because these parameters scale with the magnitude of the stress changes used in the seismicity model. The values of these parameters derived from one particular choice of the stress model should, therefore, be considered with caution. Although the true value of these parameters remains highly uncertain because of the uncertainty of the stress input, the forecast is not very sensitive to possible bias introduced by the choice of the stress model, because it is compensated by a bias in the estimate of the model parameters.

One should finally note that, by choosing one particular stress model considered as true, one takes in effect into account the entire dataset that was used to derive the stress model, including data beyond the various training periods considered in this study. The reservoir model employed to compute stress is calibrated using data up to May 2017 (Meyer *et al.*, 2022) and the geomechanical was calibrated using subsidence data from the onset of production until 2019 (Smith *et al.*, 2019). A more rigorous test of the forecasting method should have in principle used a stress model derived from only the data over the training period. This is a minor limitation in the case of this study, given the small sensitivity of the temporal forecast to different stress calculations.

Conclusion

This study describes a method to quantify the epistemic and aleatoric sources of uncertainties in induced earthquake forecasting. The proposed algorithm (Algorithm 1) provides bounds on the confidence intervals. We demonstrated and tested the performance of the approach in the particular case example of induced seismicity at the Groningen gas field where abundant information is available. In principle, the reservoir model, the geomechanical model, and the earthquake nucleation model are all sources of epistemic uncertainties. The method is general enough that it could include all these sources of uncertainties, albeit at a computational cost. To reduce the

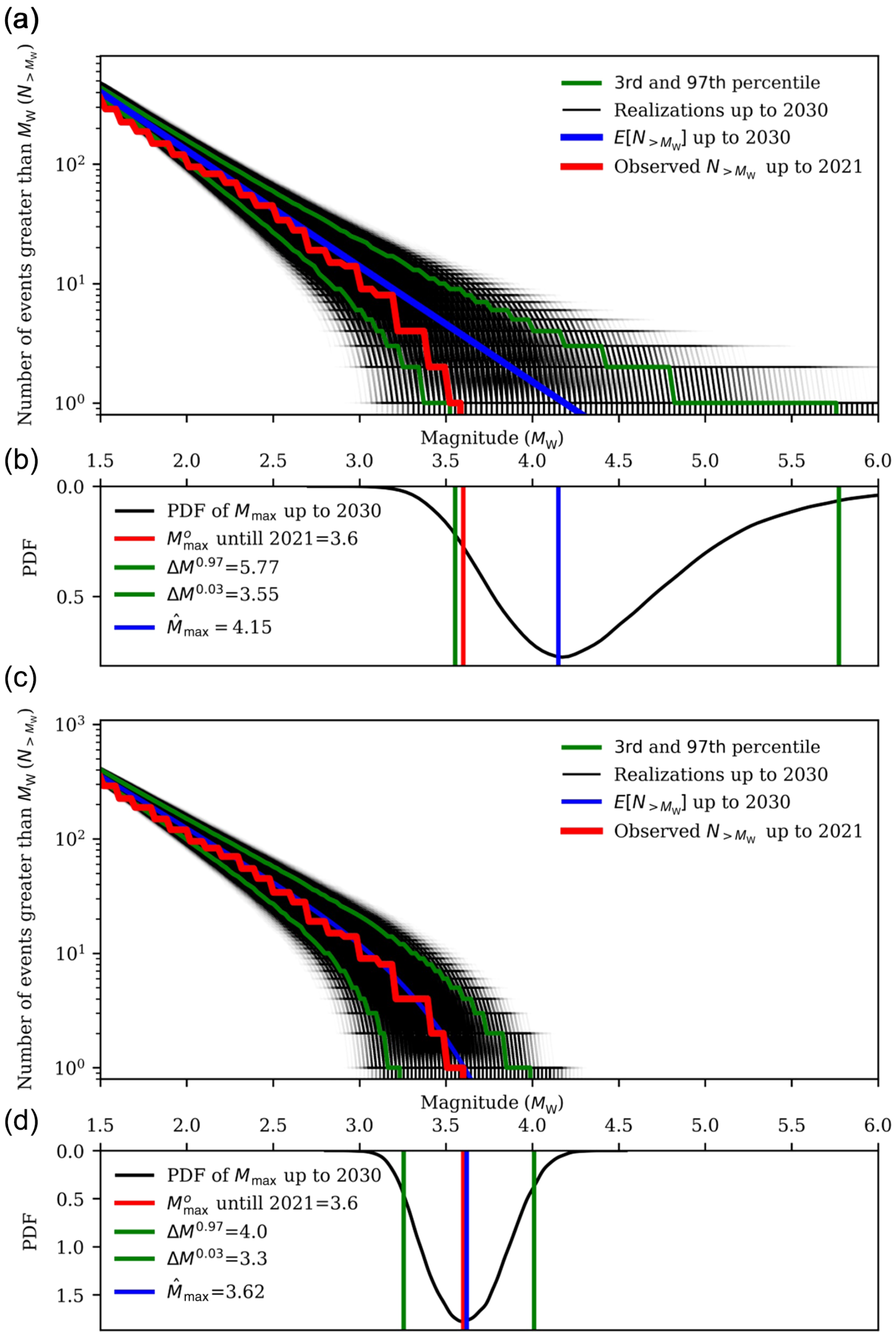

Figure 10. Forecast of the earthquake numbers and magnitudes using the flowchart of Figure 2. Panels (a) and (c) show the nontapered and tapered GR distributions, respectively. For the nontapered GR samples, 20 realizations of the model parameter (**u**), 20 realizations for the number of events (for each model parameter), 50 realizations of b-value, and, finally, 10 realizations of equation (16) are used to generate total 2×10^5 black lines in panel (a). For the tapered GR distribution, five realizations of model parameters **u**, five realizations for the numbers of events, 20 realizations of β, 20 realizations of ζ, and finally 10 realizations from equation (21) are generated for a total of 10^5 realizations. The blue lines in panels (a) and (c) show the expected number of earthquakes above a certain magnitude by 2030 with their 3rd and 97th percentiles in green, whereas the red lines display the recorded number of such events until 2021. Panels (b) and (d) depict the PDF of the maximum magnitude for the nontapered and the tapered distributions. The blue lines represent the most probable maximum magnitude ($\hat{M}_{max}$), and the green lines indicate the 97th and 3rd percentiles. The red lines denote the maximum magnitude recorded until 2021. The color version of this figure is available only in the electronic edition.

computational costs, we assumed the stress model to be true and sampled the stress changes in 2D at only the reference level. We thus ignore the uncertainties due to the reservoir and the geomechanical models. We argued that the choice of the stress model mostly impacts the prediction of the spatial distribution of seismicity, and that the temporal evolution depends mostly on the earthquake nucleation model. We verified this claim by considering different stress models as input of the workflow. We find that the method provides confidence intervals on the number of forecasted earthquakes consistent with the observations for all the different stress models tested. This is because the possible bias introduced in the modeling of the stress changes is compensated via the calibration of the seismicity model (see Appendix A2, available in the supplemental material to this article).

We compare the performance of the Gaussian and the Poisson nonlocal likelihoods for different data sets. We find that the Poisson likelihood is more accurate than the Gaussian likelihood for small datasets with a small proportion of aftershocks (say less than 1000 events over 30 yr and <20% aftershocks). The Gaussian likelihood yields a slightly higher accuracy for a larger dataset and a higher proportion of aftershocks, because the events are not independent anymore, as assumed if a Poisson likelihood is used. However, we find that, in all the cases considered in the study, the Poisson likelihood yields more stable results with smaller fluctuations in the misfit of the forecast.

We use our framework to quantify how the deployment of a more sensitive network early during the field operations would have improved earthquake forecasting. Our simulation shows that with a magnitude of completeness of $M_c \simeq 0.5$, there would have been enough data by 2000 to calibrate the model and reach a forecasting accuracy comparable to the accuracy attained in 2020 with the real seismicity catalog.

We also show how the proposed UQ method can be used to forecast earthquake magnitude probabilities. To that effect, we examined nontapered and tapered Gutenberg–Richter frequency–magnitude distribution until 2030, assuming the "cold winter" scenario for gas extraction (NAM, 2013). We note that our stress model for the future could be biased due to the limitations of our reservoir model (no interactions with adjacent aquifers) and the assumption of purely poroelastic deformation. Nevertheless, although our forecast of maximum magnitudes is close to the previous works, as pointed out correctly in Dempsey and Suckale (2023), the most important source of uncertainty for magnitude prediction is which magnitude–frequency distribution is more suitable for the Groningen, as predictions based on those two methods are significantly different.

Data and Resources

We have used the seismicity catalog from the Royal Dutch Meteorological Survey (KNMI; www.knmi.nl, last accessed May 2022). We have used previously published resources to find the stress distribution (Smith *et al.*, 2019, 2022; Meyer *et al.*, 2022). To forecast for up to 2030, we have used the "cold winter" scenario suggested in NAM (2013). Our codes for Algorithm 1 for some simple examples introduced in Appendix A5, available in the supplemental material to this article, are available on GitHub (https://github.com/hojjatks/UQ-Seismicity-Forecasting, last accessed September 2023). The supplemental material for this article includes a brief formulation of the spatial stress calculation (Appendix A1, available in the supplemental material to this article), analysis of sensitivity of uncertainties to different stress calculations (Appendix A2, available in the supplemental material to this article), a description of the observed and the synthetic catalog (Appendix A3, available in the supplemental material to this article), discretization of equation (1) (Appendix A4, available in the supplemental material to this article), some illustrative examples of our uncertainty quantification (UQ) methodology and impact of the data on the UQ bounds (Appendix A5), Lemma 1 for finding confidence interval bound (Appendix A6), the sensitivity of the total confidence interval to different combinations of epistemic and aleatoric uncertainty levels (Appendix A7), and the likelihood of the model parameters (Appendix A8).

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest available.

Acknowledgments

The authors express their sincere gratitude to the National Science Foundation (NSF) for their financial support of this research project through the Industry-University Collaborative Research Center Geomechanics and Mitigation of Geohazards (Award Number 1822214). In addition, the authors are grateful for the valuable input and discussion provided by Elías Rafn Heimisson, Kyungjae Im, and Jan Van Elk. The authors are also grateful to the anonymous reviewer for their valuable insights and to David Dempsey for his specific and thoughtful comments and suggestions which improved this paper substantially.

References

Acosta, M., J.-P. Avouac, J. D. Smith, K. Sirorattanakul, H. Kaveh, and S. J. Bourne (2023). Earthquake nucleation characteristics revealed by seismicity response to seasonal stress variations induced by gas production at Groningen, *Geophys. Res. Lett.* 50, no. 19, e2023GL105455, doi: 10.1029/2023gl105455.

Baisch, S., C. Koch, and A. Muntendam-Bos (2019). Traffic light systems: To what extent can induced seismicity be controlled? *Seismol. Res. Lett.* **90,** no. 3, 1145–1154.

Bajgiran, H. H., P. B. Franch, H. Owhadi, C. Scovel, M. Shirdel, M. Stanley, and P. Tavallali (2021). Uncertainty quantification of the 4th kind; optimal posterior accuracy-uncertainty tradeoff with the minimum enclosing ball, available at http://arxiv.org/abs/2108.10517 (last accessed March 2023).

Bourne, S. J., and S. J. Oates (2017). Extreme threshold failures within a heterogeneous elastic thin sheet and the spatial-temporal development of induced seismicity within the Groningen gas field, *J. Geophys. Res.* **122,** no. 12, 10–299.

Bourne, S. J., and S. J. Oates (2020). Stress-dependent magnitudes of induced earthquakes in the Groningen gas field, *J. Geophys. Res.* **125,** no. 11, e2020JB020013, doi: 10.1029/2020JB020013.

Bourne, S. J., S. J. Oates, and J. van Elk (2018). The exponential rise of induced seismicity with increasing stress levels in the Groningen gas field and its implications for controlling seismic risk, *Geophys. J. Int.* **213,** no. 3, 1693–1700, doi: 10.1093/gji/ggy084.

Bourne, S. J., S. J. Oates, J. van Elk, and D. Doornhof (2014). A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir, *J. Geophys. Res.* **119**, no. 12, 8991–9015.

Brown, L. D., T. T. Cai, and A. DasGupta (2001). Interval estimation for a binomial proportion, *Stat Sci.* **16,** no. 2, 101–133, doi: 10.1214/ss/1009213286.

Burkitov, U., v. O. Henk, and P. Valvatne (2016). Groningen field review 2015 subsurface dynamic modelling report, available at https://nam-feitenencijfers.data-app.nl/download/rapport/e683753a-e085-417d-995d-b7ae7a9c820f?open=true (last accessed November 2023).

Candela, T., C. Goncalves Machado, O. Leeuwenburgh, and J. Heege Ter (2022). A physics-informed optimization workflow to manage injection while constraining induced seismicity: The Oklahoma case, *Front. Earth Sci.* **10,** doi: 10.3389/feart.2022.1053951.

Candela, T., S. Osinga, J.-P. Ampuero, B. Wassing, M. Pluymaekers, P. A. Fokker, J. D. van Wees, H. A. de Waal, and A. G. Muntendam-Bos (2019). Depletion-induced seismicity at the Groningen gas field: Coulomb rate-and-state models including differential compaction effect, *J. Geophys. Res.* **124**, no. 7, 7081–7104.

Candela, T., M. Pluymaekers, J.-P. Ampuero, J.-D. WeesVan, L. Buijze, B. Wassing, S. Osinga, N. Grobbe, and A. G. Muntendam-Bos (2022). Controls on the spatio-temporal patterns of induced seismicity in Groningen constrained by physics-based modelling with ensemble-smoother data assimilation, *Geophys. J. Int.* **229,** no. 2, 1282–1308.

Dahm, T., and S. Hainzl (2022). A Coulomb stress response model for time-dependent earthquake forecasts, *J. Geophys. Res.* **127,** no. 9, e2022JB024443, doi: 10.1029/2022JB024443.

Dempsey, D., and J. Suckale (2017). Physics-based forecasting of induced seismicity at Groningen gas field, the Netherlands, *Geophys. Res. Lett.* **44,** no. 15, 7773–7782.

Dempsey, D., and J. Suckale (2023). Physics-based forecasting of induced seismicity at Groningen gas field, The Netherlands: Post hoc evaluation and forecast update, *Seismol. Res. Lett.* doi: 10.1785/0220220317.

Dieterich, J. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. Geophys. Res.* **99,** no. B2, 2601–2618.

Dost, B., E. Ruigrok, and J. Spetzler (2017). Development of seismicity and probabilistic hazard assessment for the Groningen gas field, *Neth. J. Geosci.* **96,** no. 5, s235–s245, doi: 10.1017/njg.2017.20.

Ellsworth, W. L. (2013). Injection-induced earthquakes, *Science* **341,** no. 6142, 1225942, doi: 10.1126/science.1225942.

Gulia, L., and S. Wiemer (2019). Real-time discrimination of earthquake foreshocks and aftershocks, *Nature* **574,** no. 7777, 193–199.

Heimisson, E. R., and P. Segall (2018). Constitutive law for earthquake production based on rate-and-state friction: Dieterich 1994 revisited, *J. Geophys. Res.* **123,** no. 5, 4141–4156.

Heimisson, E. R., J. D. Smith, J.-P. Avouac, and S. J. Bourne (2022). Coulomb threshold rate-and-state model for fault reactivation: Application to induced seismicity at Groningen, *Geophys. J. Int.* **228,** no. 3, 2061–2072.

Jager, J. d., and C. Visser (2017). Geology of the Groningen field—An overview, *Neth. J. Geosci.* **96,** no. 5, s3–s15, doi: 10.1017/njg.2017.22.

Kagan, Y. Y. (2002). Seismic moment distribution revisited: I. Statistical results, *Geophys. J. Int.* **148,** no. 3, 520–541, doi: 10.1046/j.1365-246x.2002.01594.x.

Kim, T., and J.-P. Avouac (2023). Stress-based and convolutional forecasting of injection-induced seismicity: Application to the Otaniemi geothermal reservoir stimulation, *J. Geophys. Res.* 128**,** no. 4, e2022JB024960, doi: 10.1029/2022JB024960.

Kühn, D., S. Hainzl, T. Dahm, G. Richter, and I. V. Rodriguez (2022). A review of source models to further the understanding of the seismicity of the Groningen field, *Neth. J. Geosci.* **101,** e11, doi: 10.1017/njg.2022.7.

Langenbruch, C., M. Weingarten, and M. D. Zoback (2018). Physics-based forecasting of man-made earthquake hazards in Oklahoma and Kansas, *Nat. Commun.* **9,** no. 1, 3946, doi: 10.1038/s41467-018-06167-4.

Martins, J. R. R. A., and A. Ning (2021). *Engineering Design Optimization*, Cambridge University Press, ISBN: 9781108833417, doi: 10.1017/9781108980647.

Marzocchi, W., and L. Sandri (2003). A review and new insights on the estimation of the b-value and its uncertainty, *Ann. Geophys.* **46,** no. 6, 1271–1289, doi: 10.4401/ag-3472.

Meyer, H., J. D. Smith, S. Bourne, and J.-P. Avouac (2022). An integrated framework for surface deformation modeling and induced seismicity forecasting due to reservoir operations, *Geol. Soc. Spec. Publ.* **528,** no. 1, SP528-2022, doi: 10.1144/SP528-2022-169.

Muntendam-Bos, A. G., and N. Grobbe (2022). Data-driven spatiotemporal assessment of the event-size distribution of the Groningen extraction-induced seismicity catalogue, **12,** no. 1, 10119, doi: 10.1038/s41598-022-14451-z.

Nederlandse Aardolie Maatschappij (NAM) (2013). Technical addendum to the Winningsplan Groningen 2013 subsidence, induced earthquakes and seismic hazard analysis in the Groningen field, Nederlandse Aardolie Maatschappij (NAM).

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.* **83,** no. 401, 9–27, doi: 10.2307/2288914.

Patil, V. V., and H. V. Kulkarni (2012). Comparison of confidence intervals for the Poisson mean: Some new aspects, *Revstat. Stat. J.* **10,** no. 2, 211–227.

Post, R. A. J., M. A. J. Michels, J.-P. Ampuero, T. Candela, P. A. Fokker, J.-D. van Wees, R. W. van der Hofstad, and E. R. van den Heuvel (2021). Interevent-time distribution and aftershock frequency in non-stationary induced seismicity, *Sci. Rep.* **11,** no. 1, 3540.

Pruiksma, J. P., J. N. Breunese, K. Thienen-Visservan, and J. A. de Waal (2015). Isotach formulation of the rate type compaction model for sandstone, *Int. J. Rock Mech. Min. Sci.* **78,** 127–132, doi: 10.1016/j.ijrmms.2015.06.002.

Richter, G., S. Hainzl, T. Dahm, and G. Zöller (2020). Stress-based, statistical modeling of the induced seismicity at the Groningen gas field, the Netherlands, *Environ. Earth Sci.* **79,** no. 11, 252.

Scholz, C. H. (2015). On the stress dependence of the earthquake b value, *Geophys. Res. Lett.* **42,** no. 5, 1399–1402.

Shcherbakov, R., J. Zhuang, G. Zöller, and Y. Ogata (2019). Forecasting the magnitude of the largest expected earthquake, *Nat. Commun.* **10,** no. 1, 4051, doi: 10.1038/s41467-019-11958-4.

Smith, J. D., J. Avouac, R. S. White, A. Copley, A. Gualandi, and S. Bourne (2019). Reconciling the long-term relationship between reservoir pore pressure depletion and compaction in the Groningen region, *J. Geophys. Res.* **124,** no. 6, 6165–6178.

Smith, J. D., E. R. Heimisson, S. J. Bourne, and J.-P. Avouac (2022). Stress-based forecasting of induced seismicity with instantaneous earthquake failure functions: Applications to the Groningen gas reservoir, *Earth Planet. Sci. Lett.* **594,** 117697, doi: 10.1016/j.epsl.2022.117697.

Smith, J. D., R. S. White, J.-P. Avouac, and S. Bourne (2020). Probabilistic earthquake locations of induced seismicity in the Groningen region, the Netherlands, *Geophys. J. Int.* **222,** no. 1, 507–516.

Spetzler, J., and B. Dost (2017). Hypocentre estimation of induced earthquakes in Groningen, *Geophys. J. Int.* **209,** no. 1, 453–465, doi: 10.1093/gji/ggx020.

Steinbrecher, G., and W. T. Shaw (2008). Quantile mechanics, *Eur. J. Appl. Math.* **19,** no. 2, 87–112, doi: 10.1017/S0956792508007341.

Trampert, J., R. Benzi, and F. Toschi (2022). Implications of the statistics of seismicity recorded within the Groningen gas field, *Neth. J. Geosci.* **101,** e9, doi: 10.1017/njg.2022.8.

van der Elst, N. J. (2021). B-positive: A robust estimator of aftershock magnitude distribution in transiently incomplete catalogs, *J. Geophys. Res.* **126,** no. 2, e2020JB021027, doi: 10.1029/2020JB021027.

Van der Elst, N. J., M. T. Page, D. A. Weiser, T. H. W. Goebel, and S. M. Hosseini (2016). Induced earthquake magnitudes are as large as (statistically) expected, *J. Geophys. Res.* **121,** no. 6, 4575–4590.

Varty, Z., J. A. Tawn, P. M. Atkinson, and S. Bierman (2021). Inference for extreme earthquake magnitudes accounting for a time-varying measurement process, doi: 10.48550/arXiv.2102.00884.

Verdon, J. P., and J. J. Bommer (2021). 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.* **25,** no. 1, 301–326, doi: 10.1007/s10950-020-09966-9.

Zhai, G., M. Shirzaei, M. Manga, and X. Chen (2019). Pore-pressure diffusion, enhanced by poroelastic stresses, controls induced seismicity in Oklahoma, *Proc. Natl. Acad. Sci. Unit. States Am.* 116**,** no. 33, 16,228–16,233.

Zoeller, G., and M. Holschneider (2016). The maximum possible and the maximum expected earthquake magnitude for production-induced earthquakes at the gas field in Groningen, Netherlands, *Bull. Seismol. Soc. Am.* **2016,** S23D-03, doi: 10.1785/0120160220.

Zöller, G., and S. Hainzl (2023). Seismicity scenarios for the remaining operating period of the gas field in Groningen, Netherlands, *Seismol. Res. Lett.* **94,** no. 2A, 805–812.

Manuscript received 7 June 2023