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

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

[1]Mechanical and Civil engineering, California Institute of Technology, Pasadena, CA, USA
[2]Computing and Mathematical Science, California Institute of Technology, Pasadena, CA, USA
[3]Geology and Planetary Science Division, California Institute of Technology, Pasadena, CA, USA
[4]Electrical Engineering, California Institute of Technology, Pasadena, CA, USA
[5]Shell Global Solutions, Amsterdam, Netherlands

Declaration of Competing Interests

The authors declare no competing interests.

Corresponding author: Hojjat Kaveh, hkaveh@caltech.edu

Abstract

Reservoir operations related to natural gas extraction, fluid disposal, carbon dioxide storage, or geothermal energy production, are capable of inducing seismicity. Modeling tools have been developed that allow for quantitative forecasting of seismicity based on operations data, but the computational cost of such models and the difficulty in representing various sources of uncertainties make uncertainty quantification challenging. We address this issue in the context of an integrated modeling framework, which combines reservoir modeling, geomechanical modeling, and stress-based earthquake forecasting. We use the Groningen gas field as a case example of application. The modeling framework is computationally efficient thanks to a 2-D finite-element reservoir model which assumes vertical flow equilibrium, and the use of semi-analytical 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. The model parameters and their uncertainties are estimated using either a Poisson or a Gaussian likelihood. We investigate the effect of the likelihood choice on the forecast performance and we estimate uncertainties in the predicted number of earthquakes as well as in the expected magnitudes. We 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 non-tapered Gutenberg-Richter distributions to evaluate the most probable maximum magnitude over time and account for uncertainties in its estimation. We show that the framework yields realistic estimates of the seismicity model uncertainties and is applicable for operational forecasting or to design induced seismicity monitoring. It could also serve as a basis for probabilistic traffic-light systems.

Plain Language Summary

Some human industrial activities like oil and gas extraction induce earthquakes. These earthquakes are occasionally large enough to cause concern about the possibility of building damage. We address the need for uncertainty quantification in the forecasting of such earthquakes. We use an integrated modeling applied to the largest producing natural gas field in western Europe which is close to the city of Groningen in the Netherlands. Gas production started in the early 1960s and started to induce detectable seismicity 30 years later. We propose and assess the performance of an algorithm for uncertainty quantification in the forecast of earthquake numbers and magnitudes. In ad-

dition, we discriminate between methods for quantifying the forecast performance and measure how much better the forecast could be by deploying better seismic networks from the start of operations. We forecast the number of future earthquakes and an estimate of the most probable maximum magnitude based on a hypothetical future gas extraction scenarios.

1 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; Candela et al., 2018). Much progress has been made recently in the development of a physics-based and computationally efficient model that maps the relation between fluid injection/extraction to stress changes and seismicity (Meyer et al., 2022; Smith et al., 2022; Bourne et al., 2014; Bourne & Oates, 2017; D. Dempsey & Suckale, 2017; Langenbruch et al., 2018; Zhai et al., 2019; Candela et al., 2019; Kim & Avouac, 2023; Kühn et al., 2022; Dahm & Hainzl, 2022; D. E. Dempsey & Suckale, 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.

Because of the multiple model layers and the computational cost, optimizing model parameters and quantifying the uncertainties is a challenging task; it is however critical that uncertainties be assessed so that the modeling can be used to help guide operations for example through a traffic light systems (Baisch et al., 2019; Verdon & Bommer, 2021). Here we focus on quantifying the uncertainty associated with the predicted number of earthquakes and the maximum magnitude. We use the novel prior-free methodology in (Bajgiran et al., 2021) for uncertainty quantification together with the Broyden-Fletcher-Goldfarb-Shamno (BFGS) method (Martins & Ning, 2021) for optimizing likelihood functions. Unlike 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 & Dost, 2017). This case study is particularly suitable for testing the methodology due to

the available high-quality information on reservoir characteristics and operations (Valvatne, n.d.; Jager & Visser, 2017), as well as the well-documented seismicity (Dost et al., 2017; Spetzler & 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, that 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 (Meyer et al., 2022; Smith et al., 2022; Heimisson 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. However, the seismicity rate is not directly observable. Instead, we observe a seismicity catalog, which we consider 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.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 (Shcherbakov et al., 2019; Zoeller & Holschneider, 2016).

This paper is organized as follows. Section 2 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. Section 3 will apply and test the proposed uncertainty quantification 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. Finally, we will conclude the paper by summarizing our findings and discussing their implications.

2 Methods

This section is structured as follows. First, we introduce our integrated modeling approach, which includes a seismicity model as one component. We explain how the seismicity model maps stress distribution to event rate. Second, we define the likelihood functions used in this paper. Then, we discuss our algorithm for uncertainty quantification. Lastly, we present a method for estimating the maximum moment magnitude, which leverages the techniques developed in this study.

2.1 Integrated Modelling 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 stress distribution are available in Appendix A1. 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 since 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 10m above the reservoir top. The reservoir is represented by cuboids with a vertical extent corresponding to the reservoir thickness and a 500m $\times$ 500m horizontal extent. The cuboids were designed to match as closely as possible the 3-D geometry of the reservoir (Burkitov et al., 2016; Smith et al., 2022). Using the Coulomb stress history, a seismicity model is then used to predict the seismicity rate in time and space.

We use the seismicity catalog from KNMI (Royal Dutch Meteorological Survey)[1] to assess the model prediction. 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. The seismicity is ob-

[1] www.knmi.nl, downloaded on May 23, 2022

served in the zones of higher Coulomb stress changes (Fig1b). However, the consistency between the time evolution of the Coulomb stress changes and the observed seismicity is less 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 increased nearly exponentially initially, despite a relatively steady annual extraction rate. After the production rate was reduced by about 50% following the Mw 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 & Oates, 2017). This interpretation is compatible with the Groningen area being tectonically quiet since the early Cretaceous (Jager & Visser, 2017). Alternatively, the lag could result from the time-dependent earthquake nucleation process (Candela et al., 2019; Dieterich, 1994). Because both explanations may have merit, we use the threshold Rate-and-State (RS) model (Heimisson et al., 2022) which includes both effects. The non-instantaneous 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

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

with

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

where x_1 and x_2 are the positions in the East-West and North-South directions, r is the background seismicity rate, $A\sigma_0$ 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 Eq 1. Eq 1 is a discontinuous function that would make parameter inference challenging. To overcome this issue, we approximated Eq 1 with a continuous function. The details are described in Appendix A2.

The vector of model parameters is hereafter defined as

$$\mathbf{u} = \begin{bmatrix} r & t_a & A\sigma_0 & \Delta S_c \end{bmatrix} \in \mathbb{R}^4_+, \tag{3}$$

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

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_T \end{bmatrix} \in \mathbb{R}^T \tag{4}$$

where y_j is the total number of earthquakes in j^{th} time bin and T is the number of time bins. And finally, the vector of modeled seismicity *rate*, which comes from the discretized version of Eq 1 is

$$\mathbf{h}(\mathbf{u}; \mathbf{\Delta S}) = \begin{bmatrix} h_1(\mathbf{u}; \mathbf{\Delta S_1}) & h_2(\mathbf{u}; \mathbf{\Delta S_2}) & \dots & h_T(\mathbf{u}; \mathbf{\Delta S_T}) \end{bmatrix} \in \mathbb{R}^T \tag{5}$$

where $h_j(\mathbf{u}; \mathbf{\Delta S_j})$ is the predicted rate of events in the j^{th} time bin. $\mathbf{\Delta S} \in \mathbb{R}^{D_{x_1} \times D_{x_2} \times T}$ 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. $\mathbf{\Delta S_j} \in \mathbb{R}^{D_{x_1} \times D_{x_2} \times j}$ is the history of stress distribution up to j^{th} time bin. For details on how to discretize Eq 1 and find $\mathbf{h}(\mathbf{u}, \mathbf{\Delta S})$ see Appendix A2.

In this paper, 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. Since the relationship between the stress variations and the rate of seismicity is nonlinear, one should in principle take into account sub-annual variations, however, to limit the computational cost we neglect these variations.

The model parameters are all assumed to be uniform over the reservoir. As a result, all models forecast very similar spatial distributions of earthquake rates (Smith et al., 2022). They however predict substantially different temporal variations of the seis-

195 micity rate when the parameters are varied. We, therefore, use likelihood functions that
196 do not account for the spatial distribution of seismicity but only temporal variations as
197 detailed in the following section.

198 2.2 Likelihood Functions

199 Now we introduce two likelihood functions using the definitions of the vectors, $\mathbf{u}$,
200 $\mathbf{y}_{M_c}$, and $\mathbf{h}(\mathbf{u}, \mathbf{\Delta S})$. It is common practice to consider that earthquakes can be triggered
201 either by previous earthquakes or by other processes such as tectonic loading or stress
202 changes due to subsurface reservoir operations. The most adequate likelihood function
203 depends on the proportion of earthquakes triggered by other earthquakes in the observed
204 seismicity catalog. If the proportion is very small, the seismicity is generally well described
205 by a non-homogenous Poisson process, meaning that events are independent but trig-
206 gered at a rate that can vary in time depending on the loading rate (Ogata, 1988). If
207 the proportion of events triggered by previous earthquakes is large, then the non-homogeneous
208 Poisson model is less appropriate. In the context of faults governed by RS friction, co-
209 seismic stress changes result in earthquake clusters, and aftershock sequences, but the
210 total number of events is unchanged over a time-scale much longer than the character-
211 istic nucleation time (Heimisson & Segall, 2018). In that case, if earthquakes are binned
212 over a time step that is large compared to the typical duration of aftershock sequences,
213 it is probably more adequate to assume that the number of events per time bin results
214 from a Gaussian process (Heimisson et al., 2022).

215 In the case of seismicity induced at Groningen, Trampert et al. (2022) have found
216 that less than 7% of the $M_w > 0.5$ earthquakes are triggered by other earthquakes. An-
217 other study suggests possibly as much as 27% of $M_w > 1.3$ events being aftershocks.
218 In any case, an inspection of the catalog shows that aftershock sequences have proba-
219 bly a short duration (smaller than a year) (Post et al., 2021). The clustering effect due
220 to earthquake interactions is therefore probably smoothed out when only yearly bins are
221 considered, as is the case when using the Gaussian Log Likelihood (GLL). It could how-
222 ever be a source of bias when the Poisson Log Likelihood (PLL) is used.

In the continuous-time form, the non-homogeneous PLL can be written (D. Dempsey & Suckale, 2017) as:

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

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

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

Where $\langle\,,\rangle$ 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. Also, $\log(\mathbf{h})$ is a vector in $\mathbb{R}^T$ whose k^{th} element is $\log(h_k)$.

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

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

where $|.|$ denotes the norm in $\mathbb{R}^T$ where T is the number of data points in $\mathbf{y}$. Eq 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, where 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, since there are some years with zero number of events, Γ would not be invertible. To overcome this issue, we use $\Gamma = \alpha I$, where α is the average number of events in $\mathbf{y}$ and $I \in \mathbb{R}^{T\times T}$ is the identity matrix. The value of α does not change the maximum likelihood estimate because it only scales the log-likelihood. However, it is important for the purpose of uncertainty quantification. This additional assumption about the variance of the Gaussian likelihood is one of its weak points compared to the Poisson likelihood which may cause inaccuracy in estimating uncertainties.

Due to 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.

2.3 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 mostly comes from uncertainties in extraction/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. Although we propose a method to account for uncertainty in stress distribution, we neglect this source of uncertainty in our numerical simulations, considering it to be comparatively insignificant with respect to other sources. This is because varying the assumptions involved in stress calculations, such as the sampling scheme, only results in a re-scaling of the stress field $\mathbf{\Delta S}$ (Smith et al., 2022). As the model prediction depends on the ratio $\frac{\Delta S(x_1,x_2,t)-\Delta S_c}{A\sigma_0}$, the forecast is identical if $A\sigma_0$ is rescaled so that this ratio is constant. The model is therefore insensitive to such re-scaling.

Various methods exist to infer the model parameters $\mathbf{u}$ from the data using different likelihood functions (such as Eqs 8 or 7) and use them to forecast future earthquakes. While Bayesian methods combine a prior $p(\mathbf{u})$ with the data $\mathbf{y}$ and the likelihood $p(\mathbf{y}|\mathbf{u})$ to generate a posterior distribution $p(\mathbf{u}|\mathbf{y})$, their uncertainty quantification heavily relies on the choice of prior. To overcome this challenge, we use a "prior-free" methodology (Bajgiran et al., 2021) for uncertainty quantification, along with the Broyden-Fletcher-Goldfarb-Shamno (BFGS) method (Martins & Ning, 2021) (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.

Within this framework, we can separately account for different sources of uncertainty involved in predicting future events. Firstly, there is uncertainty present in the model parameters ($\mathbf{u}$), which exhibit changes in the expected rate of events (epistemic). Secondly, there is the uncertainty in the stochastic process (aleatoric), which arises from the model of Eq 1, where the *rate* of events is modeled, rather than the *number* of events. Distinguishing and estimating both these uncertainties is crucial, as they can be combined to obtain a final uncertainty quantification (UQ) bound. To achieve this, we consider a likelihood model with unknown physical parameters $\mathbf{u} \in \mathbb{R}^q$ (where q is four in this paper), and observations from $\mathbf{y} \sim p(\cdot|\mathbf{u})$, where p can be either the Gaussian or the Poisson likelihood. Our first step is to determine a UQ region for the rate of future events at the k^{th} 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 $0 \leq \alpha \leq 1$.

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

where $\sup_{\mathbf{u}' \in \mathbb{R}^p} p(\mathbf{y}|\mathbf{u}') = p(\mathbf{y}|\mathbf{u}^{\mathbf{MLE}})$ with $\mathbf{u}^{\mathbf{MLE}}$ as the MLE of the model parameters.

After maximizing the likelihood function and finding set $\Theta_{\mathbf{y}}(\alpha)$, we obtain a confidence interval for $h_k(\mathbf{u})$ as Eq 10. Note that Eq 10 only gives the confidence bound on the *rate* of events at k^{th} time and not the number of events.

$$\left[\min_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h_k(\mathbf{u}'), \max_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h_k(\mathbf{u}') \right]. \tag{10}$$

We define h_k^- and h_k^+ as the minimum and maximum *rate* of events with probability $(1-\xi)$ at k^{th} time as the following:

$$\begin{aligned} h_k^- &= \min_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h(\mathbf{u}') \\ h_k^+ &= \min_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h(\mathbf{u}') \end{aligned} \tag{11}$$

Using 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), to obtain a confidence interval that contains the true parameter with probability $1-\xi$, one can choose $\alpha = \exp\left(-\frac{1}{2}\psi_q(1-\xi)\right)$ with ψ_q the quantile function of a χ^2_q random variable (with degrees of freedom (q) equal to the dimensionality of the parameter vector $\mathbf{u}$). As $\xi \to 1$, the confidence interval collapses to h_k applied to the maximum likelihood estimator of $\mathbf{u}$, and as $\xi \to 0$, the optimal decision becomes a robust estimate over all possible $\mathbf{u}$. The choice of α is made so that the selected interval has the desired probability coverage $1-\xi$ (on the *rate* of events), with ξ chosen by the practitioner.

So far, we showed how to put bounds on the *rate* of events h_k at k^{th} time (Eq 10); 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 quan-

tity of interest function h_k. Since h_k is the rate of events at k^{th} time, the number of events at that time follows a Poisson distribution with its mean equals to h_k. Eq 12 provides the confidence interval ($[\varphi_-, \varphi_+]$) for the *number* of events occurring at a *rate* of $h_k(\mathbf{u})$, with a probability of $1-\gamma$ (Patil & Kulkarni, 2012).

$$\begin{aligned}\varphi_-(\mathbf{u}) &= \frac{1}{2}\psi_{2h_k}(\gamma/2)\\ \varphi_+(\mathbf{u}) &= \frac{1}{2}\psi_{2(h_k+1)}(1-\gamma/2)\end{aligned} \tag{12}$$

where 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 (Eq 12) by the confidence interval of a Gaussian process (Brown et al., 2001). Here $1-\gamma$ is the aleatoric coverage probability where 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 $(1-\xi)(1-\gamma)$ is given by:

$$\left[\min_{\mathbf{u}'\in\Theta_{\mathbf{y}}(\alpha)} \varphi_-(\mathbf{u}'), \max_{\mathbf{u}'\in\Theta_{\mathbf{y}}(\alpha)} \varphi_+(\mathbf{u}')\right]. \tag{13}$$

Note that the practitioner chooses two parameters, ξ, and γ, that regulate the epistemic and aleatoric uncertainty, respectively. As we prove in the following lemma if $\Theta_{\mathbf{y}}(\alpha)$ contains the true parameter with probability $1-\xi$ and the worst-case $1-\gamma$ interval is created on top (Eq 13), 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 for a fixed $(1-\xi)(1-\gamma)$, we obtain the tightest interval for the number of events.

One practical point to be mentioned is that since the quantile function of χ^2 monotonously increases, we simplify Eq 13 as

$$[\frac{1}{2}\psi_{2h_k^-}(\frac{\gamma}{2}), \frac{1}{2}\psi_{2(h_k^+ +1)}(1-\frac{\gamma}{2})] \tag{14}$$

where $\psi_{2h_k^-}$ 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 Eq 14 as the final rule for the confidence interval of the *number* of events at k^{th} point in time, with probability at least $(1-\xi)(1-\gamma)$.

Lemma 1 (Product of probabilities)**.** *Let $\varphi_-(\mathbf{u}), \varphi_+(\mathbf{u})$ be chosen based on Eq 12. Then, based on (Ulm, 1990) for a fixed k (time in the future), any parameters $\mathbf{u}$ and possible data observation (number of earthquakes) $z \sim p_x(\cdot|\mathbf{u})$ Eq 15 holds.*

$$\mathbb{P}(z \in [\varphi_-(\mathbf{u}), \varphi_+(\mathbf{u})]) \geq 1-\gamma \tag{15}$$

Then, if $\Theta_{\mathbf{y}}(\alpha)$ is constructed to contain the true (unknown) parameter $u^\dagger$ with (asymptotic) probability at least $1-\xi$ (by e.g choosing $\alpha = \exp(-\frac{1}{2}\psi_p(1-\xi))$, defining

$$I = \left[\min_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} \varphi_-(\mathbf{u}'), \max_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} \varphi_+(\mathbf{u}')\right], \tag{16}$$

we (asymptotically) have, for all possible values of the true parameter $u^\dagger$

$$\mathbb{P}(z \in I) \geq (1-\xi)(1-\gamma) \tag{17}$$

Proof. We split $\mathbb{P}(z \in I) = \mathbb{P}(z \in I | u^\dagger \in \Theta_{\mathbf{y}}(\alpha))\mathbb{P}(u^\dagger \in \Theta_{\mathbf{y}}(\alpha))$. By construction of $\Theta_{\mathbf{y}}(\alpha))$ the second term is at least $1-\xi$. Since for any $\mathbf{u} \in \Theta_{\mathbf{y}}(\alpha)$, $[\varphi_-(\mathbf{u}), \varphi_+(\mathbf{u})] \subset I$, the first term is at least $1-\gamma$, which completes the proof. □

We summarize the method in Algorithm 1.

2.4 Inclusion of uncertainties on the stress model

While not considered in this work numerically, the methods proposed here can accommodate extra sources of uncertainty. As an example, we can consider uncertainty in the discretized stress distribution $\mathbf{\Delta S}$. The stress changes used as input for the seismicity forecast depends on various factors, including the uncertainties associated with the reservoir model parameters. These parameters were obtained through history matching of the pressure data. The alterations in stress are also influenced by both the reservoir's shape, which was acquired from literature (Burkitov et al., 2016), and the distribution of the reservoir's compressibility, which was determined by adjusting surface subsidence data (Smith et al., 2019). Another factor that affects stress changes is the choice of the stress sampling scheme, although the forecast is insensitive to this to a first-order

Algorithm 1 Uncertainty quantification algorithm

1. Inputs:
   (a) Likelihood model $p(\cdot|\mathbf{u})$
   (b) $q = \dim(\mathbf{u})$
   (c) Data $\mathbf{y}$
   (d) Significance levels ξ and γ, such that an interval of coverage probability $(1-\xi)(1-\gamma)$ is obtained
2. Find $\mathbf{u}^{\mathbf{MLE}} = \arg\max_{\mathbf{u}} p(\cdot|\mathbf{u})$ via an optimization algorithm
3. Set $\alpha = \exp(-\frac{1}{2}\psi_q(1-\xi))$
4. Find the set $\Theta_{\mathbf{y}}(\alpha) := \left\{\mathbf{u} \in \mathbb{R}^q \colon \frac{p(\mathbf{y}|\mathbf{u})}{p(\mathbf{y}|\mathbf{u}^{\mathbf{MLE}})} \geq \alpha\right\}$,
5. For all points in time (k), find h_k^- and h_k^+

$$h_k^- = \min_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h(\mathbf{u}')$$
$$h_k^+ = \min_{\mathbf{u}' \in \Theta_{\mathbf{y}}(\alpha)} h(\mathbf{u}')$$

6. The confidence bound on the number of events with probability $(1-\gamma)(1-\xi)$ is given by:
$$[\frac{1}{2}\psi_{2h_k^-}(\frac{\gamma}{2}), \frac{1}{2}\psi_{2(h_k^++1)}(1-\frac{\gamma}{2})]$$

approximation. The spatial distribution of earthquake forecast is mostly affected by the uncertainties in the spatial distribution of stress changes due to these factors. However, because we have assumed uniform reservoir properties (permeability, porosity) and considered the reservoir geometry as known, the spatial distribution of the forecast is relatively insensitive to the model parameters. Nevertheless, the formalism presented above can be adapted to account for uncertainties in the stress model if necessary. We can write the likelihood models we have been using as $p(\cdot|\mathbf{u}, \mathbf{\Delta S})$. If $\mathbf{\Delta S}$ itself comes from a probability distribution (e.g modeling Gaussian noise) $q(\Delta S)$, one can define a new likelihood function as a function over $\mathbf{u}$ only (that can later be used for optimization) by marginalizing over the random variables:

$$p_{\text{NEW}}(\mathbf{y}|\mathbf{u}) := \int p(\mathbf{y}|\mathbf{u}, \mathbf{\Delta S}) q(\mathbf{\Delta S}) d\mathbf{\Delta S} \tag{18}$$

Eq 18 is the integral of likelihood ($p(\mathbf{y}|\mathbf{u}, \mathbf{\Delta S})$) weighted by $q(\mathbf{\Delta S})$.

This approach would incur a substantial increase of the computational cost of the algorithm and has not been implemented here.

2.5 Earthquake Magnitude probability

Above, we have described how to forecast the number of earthquakes while taking uncertainties into account. To estimate the probabilities of induced earthquake magnitudes, it's essential 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 a maximum magnitude beyond which earthquakes are considered physically impossible, while others have a gradual tapering (mostly exponential). Such options have been studied in the context of the Groningen gas field (Bourne & Oates, 2020; D. E. Dempsey & Suckale, 2023). In this study, we demonstrate our framework using both non-tapered and tapered distributions.

Non-tapered GR

The number of events with a magnitude greater or equal to m, ($N_{\geq m}$) in non-tapered GR is given by:

$$N_{\geq m} = N_{\geq M_c} 10^{(-b(m-M_c))} \tag{19}$$

where M_c is the cut-off magnitude (generally taken to be the magnitude above which the catalog is considered 'complete', say at the $> 90\%$ level), b is the slope of the linear regression line on a log-log plot of earthquake frequency versus magnitude 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 stochastic parameters and we sample from their distributions.

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

$$b^+ = \frac{1}{(\bar{m}' - M_c')\, ln(10)} \qquad\qquad m' \geq M_c' \tag{20}$$

where m' is the difference between the magnitude of two successive earthquakes. M_c' is a constant number that satisfies $M_c' \geq 2\delta$ condition, where 2δ is the discretization level of the moments in the catalog. To use Eq 20, one should use a moving win-

dow of events, and find those events whose $m' \geq M'_c$, and then average all m's in the window to find $\bar{m}'$. The standard deviation of the estimated b^+ can be approximated by $\frac{b^+}{\sqrt{N(m' \geq M')}}$, where $N(m' \geq M')$ is the number of events that satisfy $m' \geq M'$ condition in the time window.

The b-value might vary systematically in space, time, or with the stress level (Gulia & Wiemer, 2019; Scholz, 2015; Muntendam-Bos & Grobbe, 2022; Bourne & Oates, 2020). Here, we assume for simplicity that the b-value is stationary, we construct a probability density function from the distribution obtained by considering successive time windows (Fig 7). By sampling from this PDF, we find realizations of the b-value for our analysis.

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

Using Eq 19, 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 :

$$\hat{M}_{max} = M_c + \frac{1}{b} log_{10}(N_{\geq M_c}). \tag{21}$$

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

It is straightforward to find the probability of exceeding any magnitude over any chosen time duration as we describe below. In the limit of large $N(M \geq M_c)$, the confidence level q on the most probable maximum magnitude can be calculated using Eq 22.

$$\hat{M}_q = \hat{M}_{max} - \frac{1}{b} log_{10}(-ln(q)) \tag{22}$$

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,

$$P(\hat{M}_{max} > \hat{M}_q) = 1 - exp\left(-10^{b(\hat{M}_{max} - \hat{M}_q)}\right). \tag{23}$$

Tapered GR

Based on the tapered GR distribution, the probability of an event with seismic moment greater than $\mathcal{M}$ is given by Eq 24 (Kagan, 2002; Bourne & Oates, 2020):

$$P(\geq \mathcal{M}|\mathcal{M} \geq \mathcal{M}_m) = \left(\frac{\mathcal{M}}{\mathcal{M}_m}\right)^{-\beta} e^{-\zeta(\frac{\mathcal{M}}{\mathcal{M}_m}-1)} \tag{24}$$

where $\mathcal{M}_m$ 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. Note that β is equivalent to β in the non-tapered Gutenberg-Richter law. In the limit $\zeta = 0$ which correspond to no-tapering, we have $\beta = b * log(10)$. in this study, we consider β and ζ as independent stochastic parameters and we will sample from their posterior distribution. The flowchart of Fig 2b summarizes the sampling scheme from a tapered GR distribution while considering both the epistemic and the aleatoric uncertainties. To generate random realizations from Eq 24, the inverse transform sampling method is used (Steinbrecher & Shaw, 2008). Posterior distributions of β and ζ are found by maximizing the following log-likelihood function.

$$l = \sum_{i=1}^{N} \left(log\left(\beta + \zeta \frac{\mathcal{M}_i}{M_m}\right) - (1+\beta) log \frac{\mathcal{M}_i}{M_m} - \zeta(\frac{\mathcal{M}_i}{M_m} - 1) \right) \tag{25}$$

where N is the number of events. We can relate the seismic moment ($\mathcal{M}$) to the moment magnitude (M) using Eq 26.

$$log_{10}\mathcal{M} = (c + dM) \tag{26}$$

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

$$log\mathcal{M}_m = \left(c + d\left(M_c - \frac{1}{2}\Delta M\right)\right) log10 \tag{27}$$

where M_c is the magnitude of completeness and ΔM is the size of binned intervals. Based on Eq 24, the probability of an event with seismic moment smaller than $\mathcal{M}$ is given by:

$$P(< \mathcal{M}|\mathcal{M} \geq \mathcal{M}_m) = 1 - P(\geq \mathcal{M}|\mathcal{M} \geq \mathcal{M}_m) \tag{28}$$

As a result, assuming the magnitude of events are independent and identically distributed according to Eq 24, the probability that $N_{\geq M_c}$ events having seismic moment smaller than $\mathcal{M}$ can be found by Eq 29:

$$P(< \mathcal{M}, X = N_{\geq M_c} | \mathcal{M} \geq \mathcal{M}_m) = (1 - P(\geq \mathcal{M} | \mathcal{M} \geq \mathcal{M}_m))^{N_{\geq M_c}} \tag{29}$$

where X counts the occurrence of specific outcome. So 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:

$$P(> \mathcal{M}, X \geq 1 | \mathcal{M} \geq \mathcal{M}_m) = 1 - (1 - P(\geq \mathcal{M} | \mathcal{M} \geq \mathcal{M}_m))^{N_{\geq M_c}} \tag{30}$$

3 Results and Discussion

3.1 Uncertainty in the forecast of earthquake numbers

As one example of Algorithm 1 for uncertainty quantification, the earthquake catalog is split into a training set and a test set. In this particular instance, the training set includes data up to the year 2008 (shown in white in Fig 3), while the test set includes data from 2009 to 2021 (shown in grey in Fig 3). The maximum likelihood model parameters ($\mathbf{u}^{MLE}$) are estimated using the training set. We used the Poisson likelihood to obtain the inversion results shown in Figure 3, but it's worth noting that our methodology is independent of the choice of likelihood. Once we have estimated $\mathbf{u}^{MLE}$, we run the forward model to obtain the maximum likelihood estimate for the seismicity rate ($\mathbf{h}^{MLE}$), as shown in Fig 3).

We now seek to bound the 90% confidence level ($(1-\gamma)(1-\xi) = 0.9$) for the number of events in each year, where $(1-\gamma)$ and $(1-\xi)$ are the confidence level in the Poisson process (aleatoric uncertainty) and the model parameters (epistemic uncertainty). 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 A3. 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 Fig 3, we selected the combination $1-\gamma = 0.96$ and $1-\xi = 0.94$.

The red lines in Fig 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 Eq 14.

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

3.2 Comparing Likelihoods

In this part, we compare the performance of the non-local Gaussian and Poisson likelihood. When the number of events in a time bin is small (< 30), the Poisson likelihood is in principle more appropriate than the Gaussian likelihood, provided interevent triggering can be neglected. This makes the effect of likelihood important in the context of induced seismicity because, at the beginning of seismic activity, the number of events is small and clustering is hard to assess. We also use our modeling framework to assess how a better seismic monitoring would have helped to predict more accurately and earlier induced seismicity at Groningen.

Fig 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. We therefore 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 ten times more events than the KNMI catalog and would correspond to a magnitude of detection of about $M_c \simeq 0.5$. So it mim-

ics 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 uncertainty quantification algorithm utilized in it, please refer to Appendix A4. By lowering the magnitude cut-off we should end up with a larger proportion of aftershocks. We should therefore 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. Since 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:

$$\epsilon(\mathbf{y}^o_{test}, \mathbf{h}_{test}) = \sqrt{\frac{1}{N_{test}} \left\langle \frac{\mathbf{y}^o_{test} - \mathbf{h}_{test}}{\mathbf{y}^o_{test}}, \frac{\mathbf{y}^o_{test} - \mathbf{h}_{test}}{\mathbf{y}^o_{test}} \right\rangle}, \tag{31}$$

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

Fig 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 $M_c = 1.2$ (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. Moreover, 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 inversion shows a strong trade-off between the background seismicity rate (r) and the characteristic time (t_a). This is because the system has not yet settled back to a regime where the seismicity would result only from the tec-

tonic loading. As a result, the model prediction depends only on the product $r.t_a$ and the forecast is relatively well bounded in spite of the poor constraints on either of these model parameters. Mathematically speaking, in Eqs 1 and 2, when $t > t_b$, in most of the data points, one can approximate:

$$f_2 \approx \frac{1}{t_a} \int_{t_b}^{t} f_1(x_1, x_2, t')dt' \tag{32}$$

then, the forecast of the model of Eq 1 (R) is proportional to the $(r.t_a)$ $(R \propto r.t_a.)$. when $t < t_b$ the model of Eq 1 is not a function of the values of r and t_a. This statement justifies why when using the model in Eq 1 with the current available data set, it is only possible to find the multiplication of r and t_a and not their values respectively. To be able to constrain those parameters we need to have stress distributions such that $f_2 \approx \frac{1}{t_a} \int_{t_b}^{t} f_1(x_1, x_2, t')dt'$ is not valid which is not the case for the current data set. For more information about the inverted model parameters see Appendix A5.

It is important to note 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 ($\Gamma = \alpha I$), with α being the average number of events per year. This is only a first-order approximation that is used in this paper for convenience and consistency with the alternative Poisson process.

Fig 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 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 years earlier.

3.3 Prospective forecast of earthquake numbers and magnitudes

In this part, 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 non-tapered and tapered GR distributions. To quantify magnitude probabilities for both distributions, we have followed the steps in Fig 2. To forecast the seismicity after 2021 (the grey region in Fig 8), we have assumed the 'cold

winter' scenario for gas extraction (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 8).

For the non-tapered GR distribution, we use the b-value determined with the b^+ method (Eq 20) with a moving window of 400 events. The time-series of the b^+ is plotted in 7a. The fact that there is no systematic trend with time, supports our hypothesis of a stationary b-value. We incorporate all the full distribution of b-value measurements to build a PDF (Fig 7b) from which we can sample. For the tapered GR distribution, to find the posterior distribution of β and ζ, the catalog up to 2021 is used with uniform priors ($0.3 \leq \beta \leq 1$ and $0 \leq \zeta \leq 1$). The maximum likelihood estimates for β and ζ are $\beta = 0.62$, correspond to a b-value of 1.4, $\zeta = 1.3 \times 10^{-3}$, with $\mathcal{M}_m = 1.9 \times 10^{11} Nm$.

We use Eqs 23 and 21 for the non-tapered and Eq 30 for the tapered distribution to determine the probability of exceeding a certain magnitude ($P(\hat{M}_{max} > \hat{M}_q)$). This probability depends on stochastic parameters for both non-tapered and tapered GR distributions, and realizations of those parameters are required to accurately calculate it. These stochastic variables include the model parameter ($\mathbf{u}$), the number of events that is a random realization of a non-homogenous Poisson process with the rate $\mathbf{h}(\mathbf{u}, \mathbf{\Delta S})$, and the b-value for non-tapered GR distribution, as well as β and ζ for tapered distribution. By generating multiple realizations from all of these stochastic variables, we find an empirical average for $P(\hat{M}_{max} > \hat{M}_q)$ for different $\hat{M}_q$. Note that $P(\hat{M}_{max} > \hat{M}_q)$ is an increasing function of the total number of events and, since the total number of events increases over time, $P(\hat{M}_{max} > \hat{M}_q)$ also increases. Fig 9 illustrates the empirical average evolution of $P(\hat{M}_{max} > \hat{M}q)$ over time. Since 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 non-tapered, has a significant impact on the probability of surpassing a certain magnitude. However, determining which distribution is more suitable for the Groningen gas field (Bourne & Oates, 2020; D. E. Dempsey & Suckale, 2023; Varty et al., 2021; Muntendam-Bos & Grobbe, 2022) is beyond the scope of this paper.

Realizations of the non-tapered and tapered GR distributions are obtained using the workflow described in Fig 2. The non-tapered and tapered GR distributions are dis-

played in Fig 10 (a1 and b1 respectively). In Fig 10 (a2) and Fig 10 (b2), 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 97^{th} and 3^{rd} percentiles as well as the observed maximum magnitude up to 2021. The maximum observed magnitude in the field lies close to the 3^{rd} percentile for the non-tapered GR distribution whereas it is close to the most probable maximum magnitude for the tapered distribution.

To quantify earthquake magnitude probabilities, we have assumed that the distribution of the b-value in non-tapered 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/locations that have large fluctuations in fluid extraction. In addition, the stress model is not assigned any uncertainties in this analysis. 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 the 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, as the model fits well with the observed subsidence (Smith et al., 2019). There is however 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.

4 Conclusion

This study proposes a method to quantify epistemic and aleatoric sources of errors 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 a particular case example of induced seismicity at the Groningen gas field where abundant information is available. We don't take all possible sources of uncertainties into account. Especially uncertainties associated with the reservoir and geomechanical models are ignored (the stress model is assumed true). The method is general enough that it could include these other sources of uncertainties, albeit at a computational cost. De-

spite ignoring the uncertainty of the stress model, we find that the method provides confidence intervals on the number of forecasted earthquakes consistent with the observations. This is because the possible bias introduced in the modeling of the stress changes is compensated via the calibration of the seismicity model.

We compare the performance of the Gaussian and the Poisson non-local 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 years and less than 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 earlier than 2014 would have improved earthquake forecasting. There would have been enough data by 2000 to calibrate the model and reach an accuracy comparable to the forecasting accuracy attained in 2020 with the real seismicity catalog. We also show how the proposed UQ method can be used to forecast earthquake numbers and magnitude probabilities. To that effect, we examined non-tapered 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 poro-elastic deformation.

Data and resource

We have used the seismicity catalog from KNMI (Royal Dutch Meteorological Survey)[2]. To find the stress distribution we have used previously published resources (Meyer et al., 2022; Smith et al., 2022, 2019). To forecast for up to 2030, we have used the 'cold winter' scenario suggested in NAM (2013).

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 Col-

[2] www.knmi.nl, downloaded on May 23, 2022

laborative Research Center Geomechanics and Mitigation of Geohazards (award #1822214). Additionally, the authors are grateful for the valuable input and discussion provided by Elías Rafn Heimisson, Kyungjae Im, and Jan Van Elk.

References

Baisch, S., Koch, C., & Muntendam-Bos, A. (2019). Traffic light systems: To what extent can induced seismicity be controlled? *Seismol. Res. Lett.*, *90*(3), 1145–1154.

Bajgiran, H. H., Franch, P. B., Owhadi, H., Scovel, C., Shirdel, M., Stanley, M., & Tavallali, P. (2021). Uncertainty Quantification of the 4th kind; optimal posterior accuracy-uncertainty tradeoff with the minimum enclosing ball. *arXiv preprint arXiv:2108.10517*.

Bourne, S. J., & Oates, S. J. (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*(12), 10–299.

Bourne, S. J., & Oates, S. J. (2020). Stress-Dependent Magnitudes of Induced Earthquakes in the Groningen Gas Field. *J. Geophys. Res.*, *125*(11), e2020JB020013. doi: 10.1029/2020JB020013

Bourne, S. J., Oates, S. J., & van Elk, J. (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*(3), 1693–1700. doi: 10.1093/gji/ggy084

Bourne, S. J., Oates, S. J., van Elk, J., & Doornhof, D. (2014). A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir. *J. Geophys. Res.*, *119*(12), 8991–9015.

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval Estimation for a Binomial Proportion. *Stat Sci.*, *16*(2), 101–133. doi: 10.1214/ss/1009213286

Burkitov, U., Henk, v. O., & Per Valvatne. (2016). Groningen Field Review 2015 Subsurface Dynamic Modelling Report.

Candela, T., Osinga, S., Ampuero, J., Wassing, B., Pluymaekers, M., Fokker, P. A., ... Muntendam-Bos, A. G. (2019). Depletion-induced seismicity at the Groningen gas field: Coulomb rate-and-state models including differential compaction effect. *J. Geophys. Res.*, *124*(7), 7081–7104.

Candela, T., Wassing, B., Ter Heege, J., & Buijze, L. (2018). How earthquakes are

induced. *Science*, *360*(6389), 598–600.

Dahm, T., & Hainzl, S. (2022). A Coulomb Stress Response Model for Time-Dependent Earthquake Forecasts. *J. Geophys. Res.*, *127*(9), e2022JB024443. doi: 10.1029/2022JB024443

De Marsily, G. (1986). *Quantitative hydrogeology* (Tech. Rep.). Paris School of Mines, Fontainebleau.

Dempsey, D., & Suckale, J. (2017). Physics-based forecasting of induced seismicity at Groningen gas field, the Netherlands. *Geophys. Res. Lett.*, *44*(15), 7773–7782.

Dempsey, D. E., & Suckale, J. (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*(B2), 2601–2618.

Dost, B., Ruigrok, E., & Spetzler, J. (2017). Development of seismicity and probabilistic hazard assessment for the Groningen gas field. *NETH J GEOSCI*, *96*(5), s235–s245. doi: 10.1017/njg.2017.20

Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, *341*(6142).

Felzer, K. R., Becker, T. W., Abercrombie, R. E., Ekström, G., & Rice, J. R. (2002). Triggering of the 1999 Mw 7.1 Hector Mine earthquake by aftershocks of the 1992 Mw 7.3 Landers earthquake. *J. Geophys. Res.*, *107*(B9), ESE–6.

Gulia, L., & Wiemer, S. (2019). Real-time discrimination of earthquake foreshocks and aftershocks. *Nature*, *574*(7777), 193–199.

Heimisson, E. R., & Segall, P. (2018). Constitutive law for earthquake production based on rate-and-state friction: Dieterich 1994 revisited. *J. Geophys. Res.*, *123*(5), 4141–4156.

Heimisson, E. R., Smith, J. D., Avouac, J.-P., & Bourne, S. J. (2022). Coulomb threshold rate-and-state model for fault reactivation: application to induced seismicity at Groningen. *Geophys. J. Int.*, *228*(3), 2061–2072.

Jager, J. d., & Visser, C. (2017, December). Geology of the Groningen field – an overview. *NETH J GEOSCI*, *96*(5), s3–s15. doi: 10.1017/njg.2017.22

Kagan, Y. Y. (2002, March). Seismic moment distribution revisited: I. Statistical results. *Geophys. J. Int.*, *148*(3), 520–541. doi: 10.1046/j.1365-246x.2002.01594.x

Kim, T. J., & Avouac, J.-P. (2023). *Stress-Based and Convolutional Forecasting of Injection-Induced Seismicity: Application to The Otaniemi Geothermal Reservoir*

Stimulation (preprint). Preprints.

Kuvshinov, B. N. (2008). Elastic and piezoelectric fields due to polyhedral inclusions. *Int J Solids Struct .*, *45*(5), 1352–1384.

Kühn, D., Hainzl, S., Dahm, T., Richter, G., & Rodriguez, I. V. (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., Weingarten, M., & Zoback, M. D. (2018, September). Physics-based forecasting of man-made earthquake hazards in Oklahoma and Kansas. *Nat Commun*, *9*(1), 3946. doi: 10.1038/s41467-018-06167-4

Li, B. Q., Khoshmanesh, M., & Avouac, J. (2021). Surface deformation and seismicity induced by poroelastic stress at the Raft River geothermal field, Idaho, USA. *Geophys. Res. Lett.*, *48*(18), e2021GL095108.

Martins, J. R. R. A., & Ning, A. (2021). *Engineering design optimization.* Cambridge University Press.

Marzocchi, W., & Sandri, L. (2003). A review and new insights on the estimation of the b-valueand its uncertainty. *Ann. Geophys.*.

Meyer, H., Smith, J. D., Bourne, S., & Avouac, J.-P. (2022). An integrated framework for surface deformation modeling and induced seismicity forecasting due to reservoir operations. *Geol. Soc. Spec. Publ.*, *528*(1), SP528–2022.

Muntendam-Bos, A. G., & Grobbe, N. (2022). Data-driven spatiotemporal assessment of the event-size distribution of the Groningen extraction-induced seismicity catalogue. , *12*(1), 10119. doi: 10.1038/s41598-022-14451-z

NAM. (2013). *Technical Addendum to the Winningsplan Groningen 2013 Subsidence, Induced Earthquakes and Seismic Hazard Analysis in the Groningen Field.* Nederlandse Aardolie Maatschappij (NAM).

Oates, S., Landman, A. J., van der Wal, O., Baehr, H., & Piening, H. (2022). Geomechanical, seismological, and geodetic data pertaining to the Groningen gas field: a data package used in the "Mmax II Workshop", on constraining the maximum earthquake magnitude in the Groningen field. *Data publication platform of Utrecht University*.

Ogata, Y. (1988). Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *J Am Stat Assoc*, *83*(401), 9–27. doi: 10.2307/2288914

Patil, V. V., & Kulkarni, H. V. (2012). Comparison of confidence intervals for the poisson mean: some new aspects. *Revstat Stat. J.*, *10*(2), 211–227.

Post, R. A. J., Michels, M. A. J., Ampuero, J.-P., Candela, T., Fokker, P. A., van Wees, J.-D., ... Heuvel, E. R. (2021). Interevent-time distribution and aftershock frequency in non-stationary induced seismicity. *Sci. Rep.*, *11*(1), 1–10.

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

Ross, S. M. (2014). *Introduction to probability models*. Academic press.

Scholz, C. H. (2015). On the stress dependence of the earthquake b value. *Geophys. Res. Lett.*, *42*(5), 1399–1402. (Publisher: Wiley Online Library)

Shcherbakov, R., Zhuang, J., Zöller, G., & Ogata, Y. (2019). Forecasting the magnitude of the largest expected earthquake. *Nat Commun*, *10*(1), 4051. doi: 10.1038/s41467-019-11958-4

Smith, J. D., Avouac, J., White, R. S., Copley, A., Gualandi, A., & Bourne, S. (2019). Reconciling the long-term relationship between reservoir pore pressure depletion and compaction in the Groningen region. *J. Geophys. Res.*, *124*(6), 6165–6178.

Smith, J. D., Heimisson, E. R., Bourne, S. J., & Avouac, J.-P. (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., White, R. S., Avouac, J.-P., & Bourne, S. (2020). Probabilistic earthquake locations of induced seismicity in the Groningen region, the Netherlands. *Geophys. J. Int.*, *222*(1), 507–516.

Spetzler, J., & Dost, B. (2017). Hypocentre estimation of induced earthquakes in Groningen. *Geophys. J. Int.*, *209*(1), 453–465. doi: 10.1093/gji/ggx020

Steinbrecher, G., & Shaw, W. T. (2008, April). Quantile mechanics. *Eur. J. Appl. Math.*, *19*(2), 87–112. doi: 10.1017/S0956792508007341

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

Ulm, K. (1990). Simple method to calculate the confidence interval of a standardized mortality ratio (SMR). *Am. J. Epidemiol.*, *131*(2), 373–375.

Valvatne, P. (n.d.). *Groningen Field Review 2015 Subsurface Dynamic Modelling Re-*

port (Tech. Rep.). Nederlandse Aardolie Maatschappij.

van der Elst, N. J. (2021). B-Positive: A Robust Estimator of Aftershock Magnitude Distribution in Transiently Incomplete Catalogs. *J. Geophys. Res.*, *126*(2), e2020JB021027. doi: 10.1029/2020JB021027

Van der Elst, N. J., Page, M. T., Weiser, D. A., Goebel, T. H. W., & Hosseini, S. M. (2016). Induced earthquake magnitudes are as large as (statistically) expected. *J. Geophys. Res.*, *121*(6), 4575–4590.

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

Verdon, J. P., & Bommer, J. J. (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*(1), 301–326. doi: 10.1007/s10950-020-09966-9

Yang, S., & Wei, J. (2017). *Fundamentals of petrophysics.* Springer.

Yortsos, Y. C. (1995). A theoretical analysis of vertical flow equilibrium. *Transp. Porous Media*, *18*(2), 107–129.

Zhai, G., Shirzaei, M., Manga, M., & Chen, X. (2019). Pore-pressure diffusion, enhanced by poroelastic stresses, controls induced seismicity in Oklahoma. *Proc. Natl. Acad. Sci. U.S.A.*, *116*(33), 16228–16233.

Zoeller, G., & Holschneider, M. (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.

List of email address for each co-author

Hojjat Kaveh: hkaveh@caltech.edu

Pau Battle: pbatllef@caltech.edu

Mateo Acosta: acosta@caltech.edu

Pranav Kulkarni: pkulkarn@caltech.edu

Stephen J Bourne: stephen.bourne@shell.com

Jean Philippe Avouac: avouac@caltech.edu

List of Figure Captions

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 10m 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..

Figure 2: Flowchart illustrating the steps involved in sampling from a GR distribution considering both the aleatoric and the epistemic sources of uncertainty. Firstly, we sample from the posterior distribution of the model parameters $\mathbf{u}$. Then, we sample from a nonhomogeneous Poisson process with the rate $\mathbf{h}(\mathbf{u}, \boldsymbol{\Delta}\mathbf{S})$ to generate a realization of $N_{\geq m}$. For the non-tapered GR (a), we use the number of events and a realization of the b-value to sample from Equation 19 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 Eq 24. For both non-tapered and tapered GR distributions we repeat the process multiple times.

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 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 (grey).

Figure 4: Confidence bound performance: (a) White region is used as the training set and the grey region is used as the test set. Dark blue is the seismicity data (with $M_c = 1.5$) and $\mathbf{h}^{MLE}$ in cyan is the maximum likelihood estimate 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 lies inside a certain confidence interval bound (y-axis). In this figure, for the sake of simplicity, ξ is considered to be equal to γ.

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 grey 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 maximum likelihood estimate 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.

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, while the grey 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, while the plus marker represents $\mathbf{y^o_{M_c=1.5}}$, the diamond stands for $\mathbf{y^o_{M_c=1.2}}$, and the star stands for the catalog based on ETAS model with $M_c \approx 0.5$ ($\mathbf{y^{ETAS}_{M_c\approx 0.5}}$).

Figure 7: (a) Time series of the b^+ and its 68 % confidence interval; there is no strong temporal change in the b-value (b) Probability distribution function (PDF) of b^+.

Figure 8: The forecast of the number/rate of events for years 2022-2030, the white region is used as the training set to forecast the seismicity in the grey region. The blue line is the seismicity data ($\mathbf{y^o_{M_c=1.5}}$), 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 0.9 and are generated using Algorithm 1 with $1-\xi = 0.94$ and $1-\gamma = 0.96$.

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 non-tapered GR distribution and the dashed lines are for the tapered GR distributions.

Figure 10: Forecast of the earthquake numbers and magnitudes using the flowchart of Fig 2. (a1) and (b1) show the non-tapered and tapered GR distributions, respectively. For the non-tapered GR samples, 20 realizations of the model parameter ($\mathbf{u}$), 20 realizations for the number of events (for each model parameter), 50 realizations of b-value, and finally, 10 realizations of Eq 19 are used to generate total 2×10^5 black lines in a1. For the tapered GR distribution, 5 realizations of model parameters $\mathbf{u}$, 5 realizations for the numbers of events, 20 realizations of β, 20 realizations of ζ, and finally 10 realizations from Eq 24 are generated for a total of 10^5 realizations. The blue lines in (a1) and (b1) show the expected number of earthquakes above a certain magnitude by 2030, while the red lines display the recorded number of such events until 2021. (a2) and (b2) de-

826 pict the PDF of the maximum magnitude for the non-tapered and the tapered distri-
827 butions. The blue lines represent the most probable maximum magnitude ($\hat{M}_{max}$), and
828 the green lines indicate the 97^{th} and 3^{rd} percentiles. The red lines denote the maximum
829 magnitude recorded until 2021.

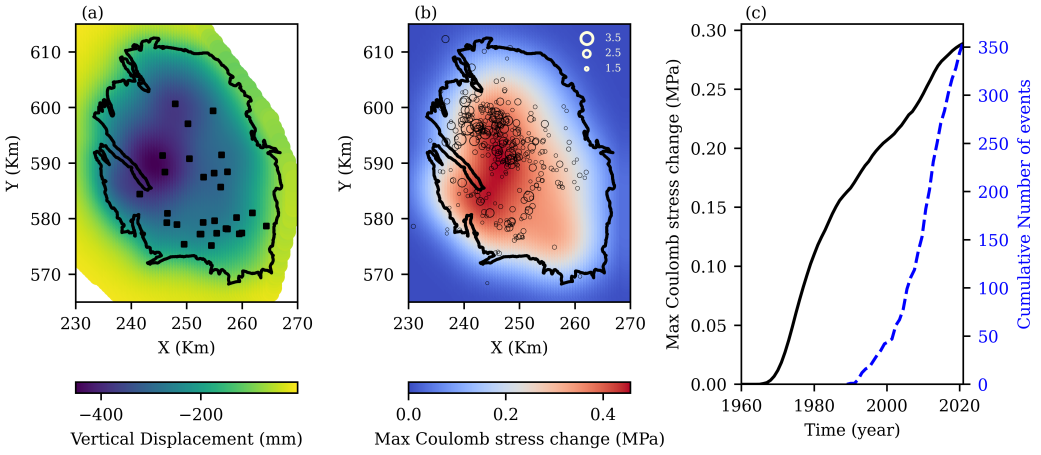

Figure 1. The Groningen field and geomechanical model results. (a) Cumulative vertical displacement since the onset of gas production as of 2019. Black rectangles denote the locations of the wells. (b) Maximum Coulomb stress change at a nominal elevation of 10m since the onset of gas production and epicenters 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.

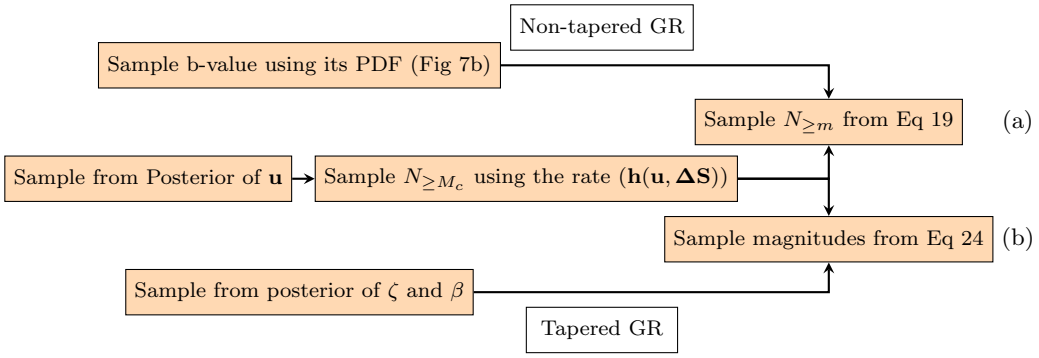

Figure 2. Flowchart illustrating the steps involved in sampling from a GR distribution considering both the aleatoric and the epistemic sources of uncertainty. Firstly, we sample from the posterior distribution of the model parameters $\mathbf{u}$. Then, we sample from a nonhomogeneous Poisson process with the rate $\mathbf{h}(\mathbf{u}, \boldsymbol{\Delta}\mathbf{S})$ to generate a realization of $N_{\geq m}$. For the non-tapered GR (a), we use the number of events and a realization of the b-value to sample from Equation 19 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 Eq 24. For both non-tapered and tapered GR distributions we repeat the process multiple times.

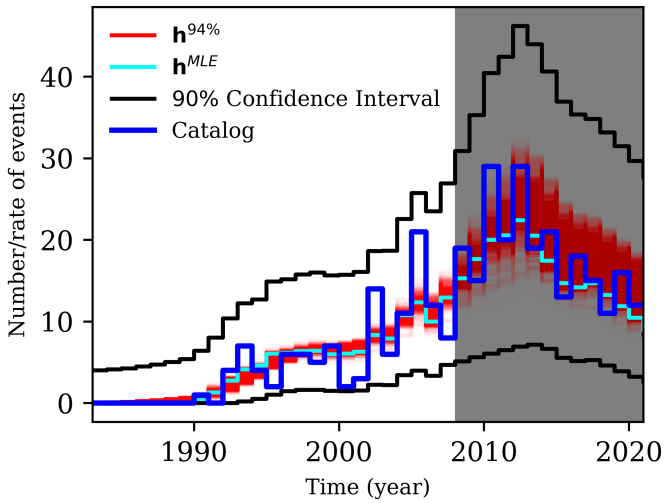

Figure 3. Quantifying uncertainty and forecast up to the year 2021 using the Poisson likelihood: red lines show the seismicity *rate* with 94% probability, the cyan line is the MLE of the *rate* of events, and black lines show 90% confidence interval of the number of events with $1-\gamma = 0.96$ and $1-\xi = 0.94$. 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 (grey).

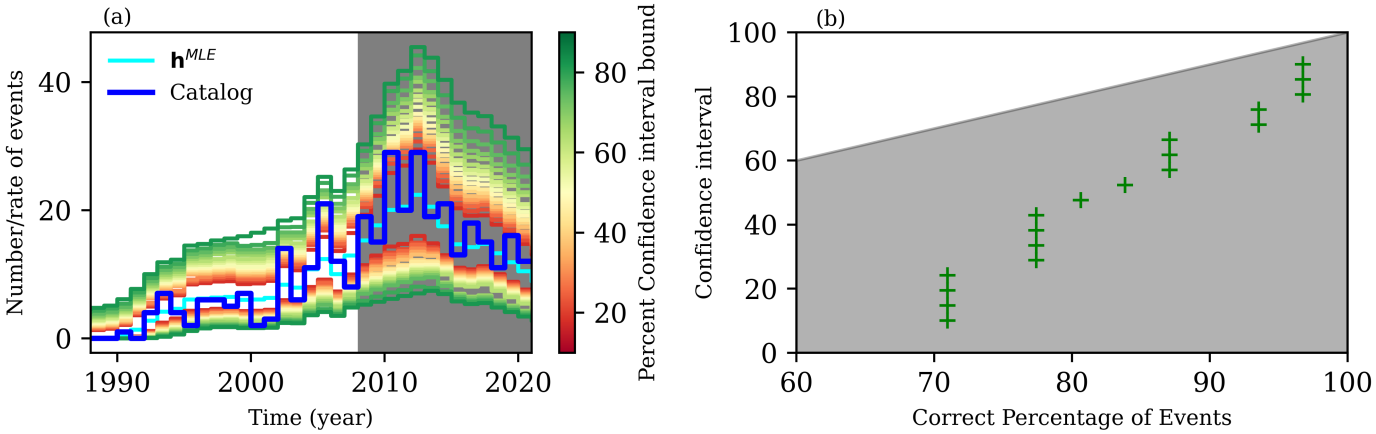

Figure 4. Confidence bound performance: (a) White region is used as the training set and the grey region is used as the test set. Dark blue is the seismicity data (with $M_c = 1.5$) and $\mathbf{h}^{MLE}$ in cyan is the maximum likelihood estimate of the *rate* of events. Different percent confidence interval bounds for the *number* of events are plotted as well. (b) The percentage of events (x-axis) that lies inside a certain confidence interval bound (y-axis). In this figure, for the sake of simplicity, ξ is considered to be equal to γ.

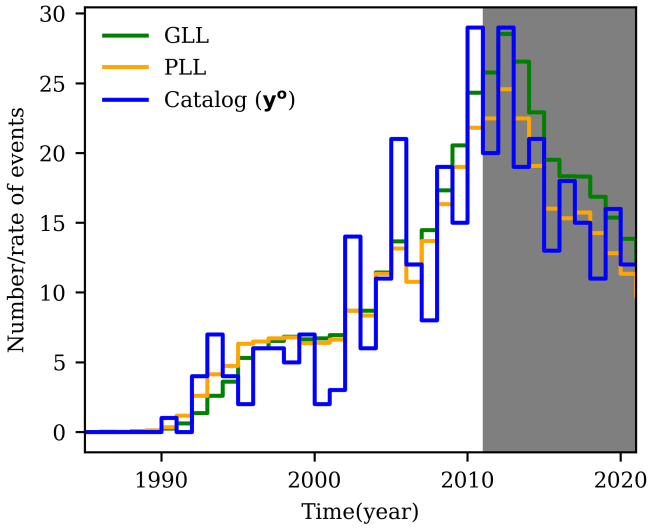

Figure 5. An example of a comparison of the MLE models obtained with the Poisson and Gaussian likelihoods. The white region is used for the training set and the grey region is used for testing. The blue line is the seismicity data with $M_c = 1.5$ and the orange and the green lines are the maximum likelihood estimate 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.

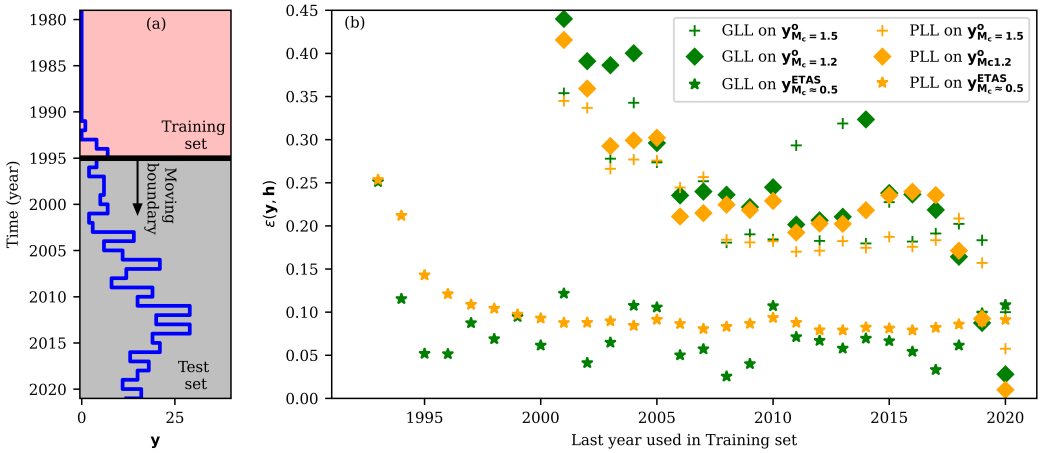

Figure 6. Comparing the performance of Poisson Log-Likelihood (PLL) and Gaussian Log-Likelihood (GLL) on different catalogs using different combinations of training and testing sets. (a) the pink region serves as the training set, while the grey region is the test set, and different boundaries between 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, while 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 ETAS model with $M_c \approx 0.5$ ($\mathbf{y}^{\mathbf{ETAS}}_{\mathbf{M_c\approx0.5}}$).

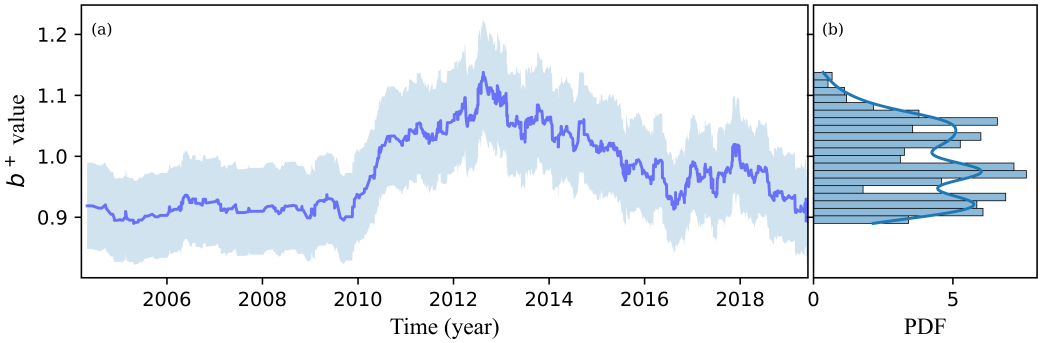

Figure 7. (a) Time series of the b^+ and its 68 % confidence interval; there is no strong temporal change in the b-value (b) Probability distribution function (PDF) of b^+.

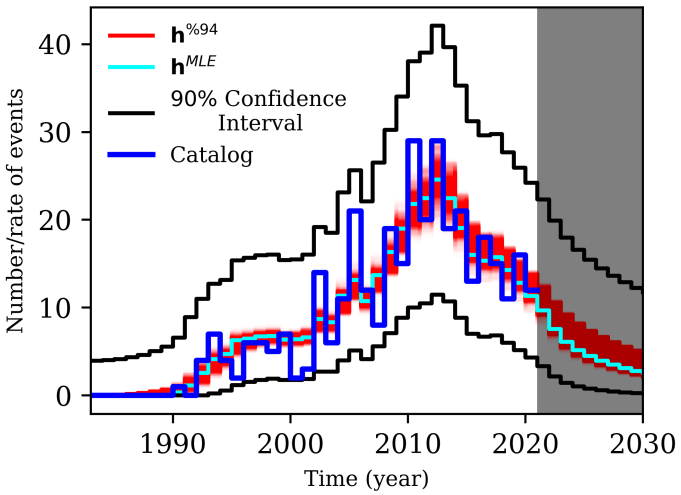

Figure 8. The forecast of the number/rate of events for years 2022-2030, the white region is used as the training set to forecast the seismicity in the grey region. The blue line is the seismicity data ($\mathbf{y}^{o}_{\mathbf{M_c=1.5}}$), cyan is the predicted *rate* of events based on Poisson likelihood, 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 0.9 and are generated using Algorithm 1 with $1-\xi = 0.94$ and $1-\gamma = 0.96$.

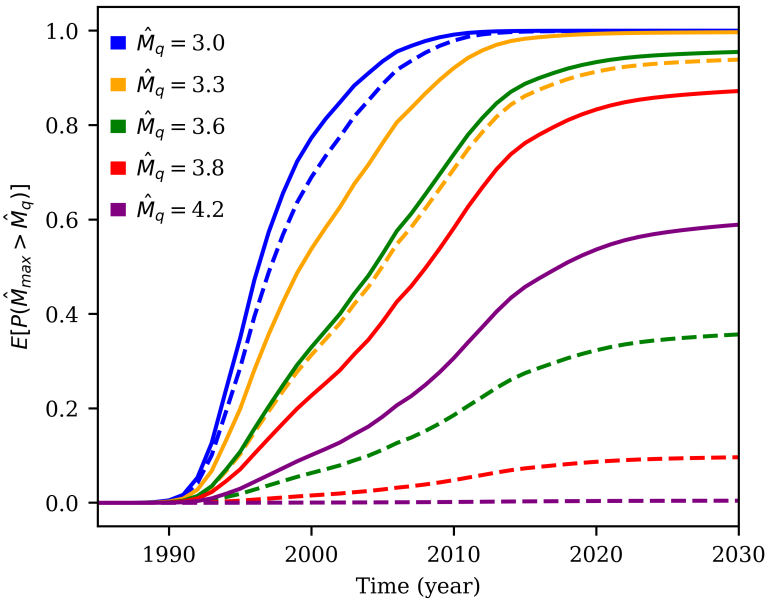

Figure 9. Evolution of the empirical average of the probability of the most probable maximum magnitude ($\hat{M}_{max}$) being higher than specific magnitudes ($\hat{M}_q$) for different values of $\hat{M}_q$. The solid lines are for the non-tapered GR distribution and the dashed lines are for the tapered GR distributions.

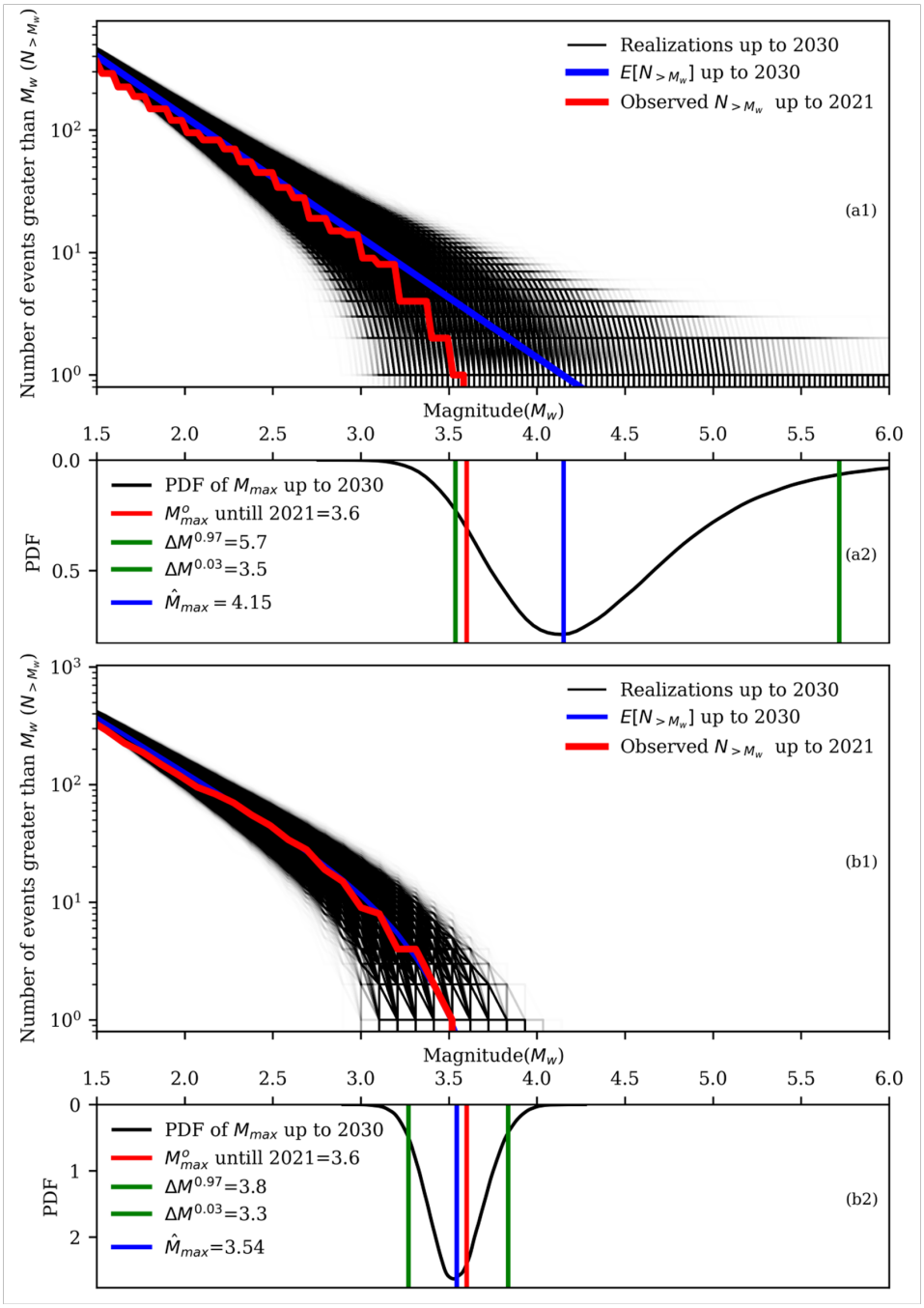

Figure 10. Forecast of the earthquake numbers and magnitudes using the flowchart of Fig 2. (a1) and (b1) show the non-tapered and tapered GR distributions, respectively. For the non-tapered GR samples, 20 realizations of the model parameter (**u**), 20 realizations for the number of events (for each model parameter), 50 realizations of b-value, and finally, 10 realizations of Eq 19 are used to generate total 2×10^5 black lines in a1. For the tapered GR distribution, 5 realizations of model parameters **u**, 5 realizations for the numbers of events, 20 realizations of β, 20 realizations of ζ, and finally 10 realizations from Eq 24 are generated for a total of 10^5 realizations. The blue lines in (a1) and (b1) show the expected number of earthquakes above a certain magnitude by 2030, while the red lines display the recorded number of such events until 2021. (a2) and (b2) depict the PDF of the maximum magnitude for the non-tapered and the tapered distributions. The blue lines represent the most probable maximum magnitude ($\hat{M}_{max}$), and the green lines indicate the 97^{th} and 3^{rd} percentiles. The red lines denote the maximum magnitude recorded until 2021.

Appendix A

A1 From Extraction Rates to Coulomb Stress Changes

To calculate the stress distribution in and around the reservoir, we use the modular and computationally efficient modeling framework that contains reservoir and geomechanical modeling (Meyer et al., 2022; Smith et al., 2022). In this section, we briefly review this framework and discuss how we calculate stress distribution using geodetic and gas extraction information. Overall, this framework is computationally efficient and reproduces stress changes in a 3-D volume.

Assuming homogeneous and constant hydraulic properties, we first compute the fluid pressure diffusion inside the reservoir due to gas extraction using the history and extraction previsions. We use the vertical equilibrium flow model (VFE) from (Meyer et al., 2022). The VFE assumption (Yortsos, 1995) holds when there is a significantly larger vertical fluid diffusion capacity compared to the horizontal one. The assumption is valid when $R_L > 10$ with R_L defined in Eq A1.

$$R_L = (\Delta x/\Delta z).(k_z/k_x)^{1/2} \tag{A1}$$

where Δx and Δz are the horizontal and vertical dimensions of the reservoir and k_x and k_z are the horizontal and vertical permeabilities of the reservoir. In the Groningen gas field case, $\Delta x \approx 35 - 50(Km)$ and $\Delta z \approx 0.1 - 0.3(Km)$. k_x and k_z can be approximated to be of the order of $10^{-13}(m^2)$. As a result, we can assume that the Groningen gas field is a long and thin gas reservoir ($R_L > 117$) and the VFE assumption holds. For this part, we use the reservoir properties from (Oates et al., 2022).

The combination of mass conservation in porous rock (De Marsily, 1986) with Darcy's law gives Eq A2. This equation assumes low matrix compressibility for the Groningen reservoir (Yang & Wei, 2017), and a smooth spatial gradient of the reservoir thickness.

$$\phi \frac{\partial \rho}{\partial p} \frac{dp(x,y,t)}{dt} + \nabla \left(\frac{-\rho(x,y,t) k_x}{\mu(x,y,t)} \nabla p(x,y,t) \right) = \frac{Q(x,y,t)}{\Delta z(x,y) * G_{sat}} \tag{A2}$$

where ϕ is the reservoir's porosity and ρ, μ, p are the fluid's density, viscosity, pressure, and velocity. Q is the flow rate, Δz is the reservoir's thickness and G_{sat} is the gas saturation in the reservoir (Meyer et al., 2022). A history matching procedure on the pressure measurements at the wells yields an error lower than 1 MPa during the whole reservoir history for $\phi = 16\%$, $k_x = 2.5^{-13}\,(m^2)$, and $G_{sat} = 35.7\%$.

We use the computed field of fluid pressure depletion (p), together with the geodetically derived uniaxial compressibility (C_m; (Smith et al., 2019)) such that the reservoir compaction writes $C = C_m(x, y)\Delta p(x, y, t)b(x, y, t)$.

The reservoir is separated into $500(m)\times500(m)\times\Delta z$ offset cuboidal volumes that deform poroelastically and represent first-order variations in reservoir geometry. The relation between compaction and stress uses a semi-analytical Green's function approach combined with a strain-volume formulation (Kuvshinov, 2008) that takes the form:

$$\sigma_{x,y,z} = -\Delta p C_m F_{\sigma_{x,y,z}}(Vertices\,, Observation\,Points) \tag{A3}$$

where σ is the stress, and $F_{\sigma_{x,y,z}}$ is a function of the cuboid's position, and the observation points. For details on this function, the reader is referred to (Kuvshinov, 2008; Li et al., 2021). Due to the resolution on compressibility values in the reservoir (Smith et al., 2019), the strain and stress fields are smoothed to 4 km length using a Gaussian kernel. From the changes in shear stress ($\Delta\tau$) and effective normal stress ($\Delta\sigma'_N = \Delta\sigma_N - \Delta p$), we compute the changes in Coulomb stress (ΔS) from 1960 to 2021. Then, using a cold winter scenario (NAM, 2013) for the gas extraction (after 2021), we extend the Coulomb stress distribution time-series to the year 2030. The changes in Coulomb stress are calculated using Eq A4.

$$\Delta S = \Delta\tau + f\Delta\sigma'_N \tag{A4}$$

with $f \approx 0.6$ the static friction coefficient of the reservoir rock.

The numerous faults offsetting the reservoir are the main cause of the Coulomb stress changes. Only the largest faults are represented in our reservoir model. Numerous secondary faults are ignored or are not even visible in the seismic profiles used to construct the reservoir model. The seismicity is not limited to the zones of high-stress concentration corresponding to the main faults though. To allow for stress increase away from these main faults, the Coulomb stress field is smoothed using a Gaussian filter with 4km standard deviation (Smith et al., 2022). This procedure essentially affects the spatial distribution but not the time evolution of the predicted seismicity rate. The Coulomb stress changes on optimally oriented faults as of 2021 are shown in Fig 1b.

A2 Approximation and Discretization of Eq 1

Eq 1 is discontinuous and it is numerically convenient to approximate it with a continuous function for the purpose of parameter inference. In this section, we present our approximation, as well as the formulation for discretizing the approximate equation. We first start by eliminating the dependent parameter t_b from Eq 1 to form Eq A5.

$$\frac{R(x_1,x_2,t)}{r} = \frac{f_1(x_1,x_2,t)H(\Delta S(x_1,x_2,t)-\Delta S_c)}{\frac{1}{t_a}\int_0^t f_1(x_1,x_2,t')H(\Delta S(x_1,x_2,t')-\Delta S_c)dt'+1} \tag{A5}$$

where H is the Heaviside function. Eq A5 is approximated by Eq A6. Note that, for $\Delta S(x_1,x_2,t) < \Delta S_c$, $f_1(x_1,x_2,t)$ is smaller than one and r is a small number which is in the order of $10^{-5}\, event/Km^2.yr$.

$$\frac{R(x_1,x_2,t)}{r} = \frac{f_1(x_1,x_2,t)}{\frac{1}{t_a}\int_0^t f_1(x_1,x_2,t')H(\Delta S(x_1,x_2,t')-\Delta S_c)dt'+1} \tag{A6}$$

The variables $\Delta S(x_1,x_2,t)$ and R can be discretized in space and time. We define $\mathbf{\Delta S}$ as the discretized stress distribution. Its value in spatial element of x_1^i and x_2^j at time t_k is denoted by $\Delta S(x_1^i,x_2^j,t_k)$. Note that $\mathbf{\Delta S} \in \mathbb{R}^{D_{x_1}\times D_{x_2}\times T}$ where D_{x_1} and D_{x_2} are the number of elements in the x_1 and x_2 direction and T is the number of time snapshots. To simplify equations, we use the notation $\mathbf{\Delta S}_k(x_1^i,x_2^j)$ as the stress history of the element i and j up to time k.

We concatenate all the model parameters into one vector $\mathbf{u} \in \mathbb{R}_+^4$ as defined in Eq A7.

$$\mathbf{u} = \begin{bmatrix} r & t_a & A\sigma_0 & \Delta S_c \end{bmatrix} \in \mathbb{R}_+^4, \tag{A7}$$

For any vector $\mathbf{u}$ we define $\mathbf{G} \in \mathbb{R}^{D_{x_1}\times D_{x_2}\times T}$ whose elements $G_k(\mathbf{u};\mathbf{\Delta S}_k(x_1^i,x_2^j))$ in x_1^i and x_2^j is dependent on the stress history of that element up to time k and is defined as Eq A8.

$$G_k(\mathbf{u};\mathbf{\Delta S}_k(x_1^i,x_2^j)) = \frac{rf_1(x_1^i,x_2^j,t_k)}{\frac{1}{t_a}\left(\sum_{l=0}^{l=k} f_1(x_1^i,x_2^j,t_l)H(\Delta S(x_1^i,x_2^j,t_l)-\Delta S_c)\right)+1} \tag{A8}$$

where H is the Heaviside function. Note that Eq A8 is a discrete version of Eq A6 and G_k is the seismicity rate (per unit area and time) for a specific element x_1^i and x_2^j at specific time t_k. At a given time (t_k), we integrate the spatial seismicity rates (G_k's) over

space, to find the seismicity rate (per unit time) of the whole area of interest. We denote this variable by $h_k(\mathbf{u}; \mathbf{\Delta S_k})$ and is defined in discrete form as the following summation.

$$h_k(\mathbf{u}; \mathbf{\Delta S_k}) = \sum_{i=1}^{D_{x_1}} \sum_{j=1}^{D_{x_2}} G_k(\mathbf{u}; \mathbf{\Delta S}_k(x_1^i, x_2^j)) \Delta x_1 \Delta x_2 \tag{A9}$$

where Δx_1 and Δx_2 are the lengths of the grids along the x_1 and x_2 directions. Then, we construct a vector in $\mathbb{R}^T$ such that it contains seismicity rates in a discrete-time fashion.

$$\mathbf{h}(\mathbf{u}; \mathbf{\Delta S}) = \begin{bmatrix} h_1(\mathbf{u}; \mathbf{\Delta S_1}) & h_2(\mathbf{u}; \mathbf{\Delta S_2}) & \dots & h_T(\mathbf{u}; \mathbf{\Delta S_T}) \end{bmatrix} \in \mathbb{R}^T \tag{A10}$$

The next step is to *compare* the seismicity rate that is defined in A10 with the time-series of a reference seismicity data ($\mathbf{y}$). $\mathbf{y}$ can be either the observed seismicity data ($\mathbf{y^o_{M_c}}$) with a given cut-off (M_c) or a synthetic catalog ($\mathbf{y}^{synth}_{M_c}$).

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_T \end{bmatrix} \in \mathbb{R}^T \tag{A11}$$

A3 Sensitivity of confidence bounds on choosing γ and ξ

As stated in Section 2, for a given confidence level $(1-\xi)(1-\gamma)$ the values of $(1-\gamma)$ and $(1-\xi)$ are not uniquely determined. There is a possibility of finding the optimum combination of γ and ξ to have the tightest interval of the number of events for a fixed confidence level of $(1-\xi)(1-\gamma)$. We have found that in our framework $\xi = \gamma$ is a good first guess. Fig A1 shows different confidence intervals for a fixed confidence level (90%) while changing the value ξ. This figure suggests that there is a small sensitivity to the selection of those parameters close to $\xi = \gamma$.

A4 Seismicity Data and Synthetic Catalog

We have applied our analysis in this paper to different earthquake catalogs. The Groningen seismicity catalog has a time-varying completeness magnitude. Until 1993 the completeness magnitude in the region was 2.5. The seismic network gradually enhanced and it reached below magnitude 0.5 after 2016. Heimisson et al. (2022) assumed a simplified constant magnitude of completeness of 1.5. In this work, we mostly use a cut-off magnitude of 1.5 which gives a total 356 number of events. In one case, where we want to see the effect of aftershocks, we use a cut-off magnitude of 1.2 with 633 total events.

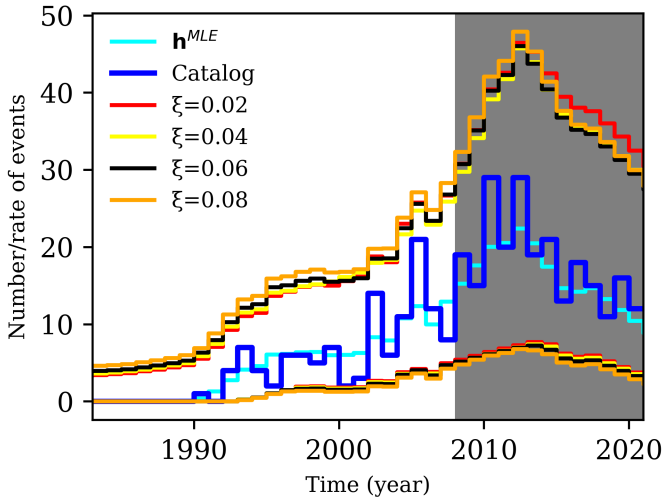

Figure A1. Sensitivity of confidence bounds on choosing γ and ξ while keeping $(1-\xi)(1-\gamma) = 0.9$. 90% confidence bounds of the number of events using different values of ξ. This figure suggests that there is a small change in confidence bounds when changing ξ and γ while $(1-\xi)(1-\gamma) = 0.9$ and close to $\xi = \gamma$.

A simplified constant magnitude of completeness is an assumption that has been used in this work.

To see the effect of aftershocks on the performance of the likelihood functions and also to study the value of a better seismic network, we have generated a synthetic catalog based on the ETAS model. The ETAS catalog is designed to have a completeness magnitude of almost 0.5 (with the b-value ≈ 0.99), which implies ten times more events compared to the catalog with $M_c = 1.5$.

Based on (Post et al., 2021) a mean fraction of 76.6% of events in the Groningen catalog is the background seismicity. Since we want to amplify the *number* of events by almost ten times, we need to generate the *rate* of events with almost 7.6 times more than the rate of events in the original catalog (with $M_c = 1.5$). Then, when the effect of aftershocks is added, we will have a catalog with almost ten times more events than the catalog with $M_c = 1.5$.

We first need to invert for the model parameters $\mathbf{u}$ in Eq 1 for a catalog with $M_c = 1.5$ and data up to the year 2021. After inverting for the model parameters, we forward simulate the seismicity model (Eq 1) to find the rate of events corresponding to maximum likelihood estimates ($\mathbf{h}(\mathbf{u}^{\mathbf{MLE}})$). Assuming a nonhomogenous Poisson process, we amplify this *rate* using the so-called *thinning* process (Ross, 2014) with a factor of 7.6. This is the *rate* of events corresponding to a catalog with $M_c \approx 0.5$. Finally, we add

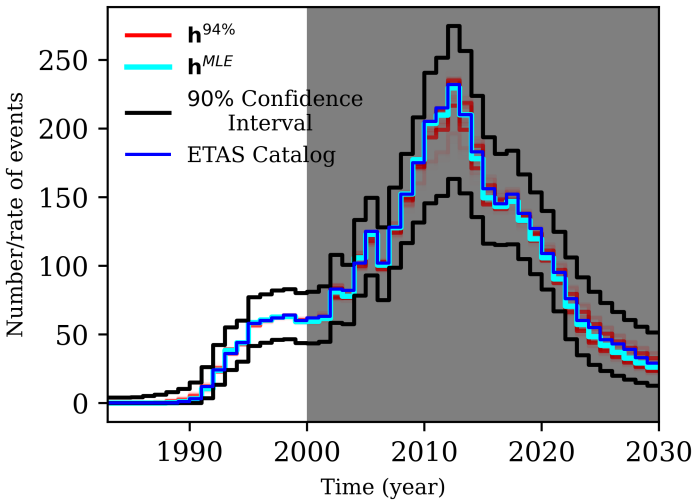

Figure A2. Forecast of the synthetic catalog ($M_c \approx 0.5$) shows that in the case of having a better seismic network, we could predict the number of events much earlier. The data up to the year 2001 is used for the training set and the grey part is the test set. The blue line is the number of events in each year in the ETAS synthetic catalog. The cyan line is the MLE for the rate of events, the red lines are the rates in the 94% confidence interval of the model parameters, and the black lines are the 90% confidence interval for the number of events.

aftershocks to this catalog using the ETAS parameters introduced in (Post et al., 2021):

$$\psi_{ETAS} = 1/\tau(t) + \sum_{i=1}^{n(t)} \frac{K}{(c+t-v_i)^{1+\theta}} 10^{\alpha(m_i - m_{min})} \tag{A12}$$

where $\psi(t)$ is the hazard of temporal ETAS model, $1/\tau(t)$ is the background rate, $n(t)$ number of events before time t, m_i and v_i are the moment magnitude and time of the i^{th} event. Based on (Post et al., 2021), $c = 6$ hours, $\theta = 0.1$, $\alpha = 0.3$ and $g = 0.4$; where $g = \frac{K}{c^\theta}\frac{b}{b-\alpha}\frac{1}{\theta}$ is defined as the branching ratio. To add the effect of aftershocks, the seismicity rate from the *thinning* process and the mentioned ETAS parameters are used in the code that was provided by (Felzer et al., 2002). The catalog that is synthetically generated by the ETAS model has 3359 events from 1979 up to 2021. Fig A2 shows the forecast of seismicity using the data set generated by the ETAS model. This figure suggests that we could predict the number of earthquakes much earlier if we had a better seismic network.

A5 Parameter Space

In the previous parts, we showed that the maximum likelihood model parameters ($\mathbf{u^{MLE}}$) can be used to find the maximum likelihood of the rate of events ($\mathbf{h}^{MLE}$) (e.g

cyan line in Fig 3). We have also shown how to find a set of event *rates* for a certain confidence level (e.g red lines Fig 3). In this part, we focus on the distribution of the model parameter space (**u**). In our inversion of Eq 1, between the parameters of the study, $A\sigma_0$ and ΔS_c are well-constrained. However, the values of the likelihoods of r and t_a are not well constrained because of the specific form of the Eq 1.

We can examine the distribution of the likelihood in the space of the model parameters. Fig A3 shows the distribution of the Poisson likelihood as a function of model parameters when information up to the year 2021 is used. As stated earlier, the inversion is almost insensitive to the individual values of r and t_a, while keeping the multiplication of those constant.

We have delved into a careful examination $r-t_a$ plot of the Fig A3 in Fig A4. The black dots in Fig A4 are the points whose likelihood ratios to the maximum likelihood are almost one (greater than 0.99). Our analysis reveals that our model/observation cannot constrain t_a and r independently. Specifically, it is possible for a very small r with a big t_a to have the same likelihood as a very big r with a small t_a. We employed the least square method to derive a relation between these two model parameters. Our analysis yields the following relationship:

$$t_a(K.yr) = \frac{4.5 \times 10^{-5} (\frac{K.event}{Km^2})}{r(\frac{event}{Km^2.yr})} \tag{A13}$$

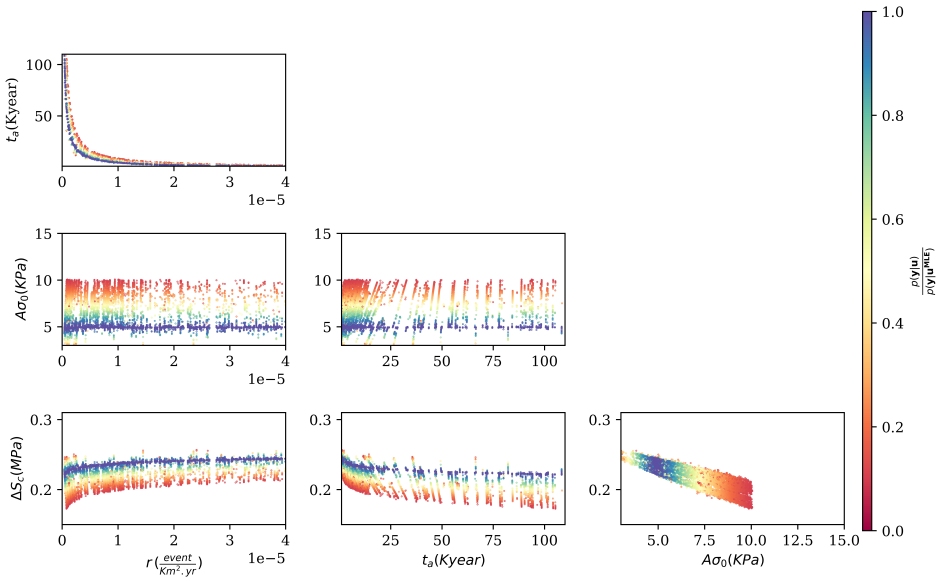

Figure A3. Ratio of the (Poisson) likelihood to the maximum (Poisson) likelihood estimate (Data until the year 2021 have been used). This graph suggests that with the current data, we cannot constrain r and t_a separately. However, the $A\sigma_0 - \Delta S_c$ plot shows the convergence of $A\sigma_0$ and ΔS_c.

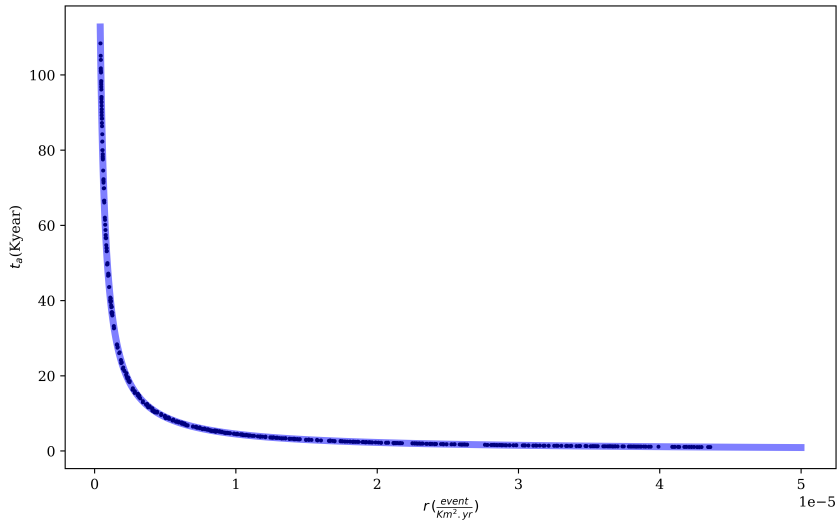

Figure A4. Relationship between r and t_a. The black dots are the points whose likelihoods are close to the maximum likelihood (their likelihood ratio to the maximum likelihood is greater than 0.99). Using the least square method we can fit a $y = \frac{a}{x}$ line to the data. The resulting blue line represents the fitted line, indicating that we can only invert for the multiplication of $t_a(Kyear)$ and $r(\frac{event}{Km^2.yr})$, as given by the Eq A13.