

Physics-Based Forecasting of Induced Seismicity at Groningen Gas Field, The Netherlands: Post Hoc Evaluation and Forecast Update

David E. Dempsey, Jenny Suckale

Author's accepted manuscript (peer-reviewed, before publisher typesetting)

Published as: Dempsey, David E., Suckale, Jenny (2023). Physics-Based Forecasting of Induced Seismicity at Groningen Gas Field, The Netherlands: Post Hoc Evaluation and Forecast Update. Seismological Research Letters. <https://doi.org/10.1785/0220220317>

Accepted manuscript. The version of record, published by the Seismological Society of America in Seismological Research Letters, is at <https://doi.org/10.1785/0220220317>. Not for commercial redistribution.

Open copy, summary and machine-readable versions:
<https://ddempsey.github.io/papers/2023-dempsey-physics-forecasting-induced-seismicity-groningen/>

1 Physics-based forecasting of induced seismicity at
2 Groningen gas field, the Netherlands: post hoc
3 evaluation and forecast update
4

5 **Authors**

6 D. E. Dempsey¹, J. Suckale²

7 ¹Dempsey Geomodelling Ltd., Christchurch, New Zealand

8 ²Department of Geophysics, Stanford University, Stanford, California, USA

9 **Corresponding author**

10 David Dempsey, dempseygeomodelling@gmail.com, 39 Victoria Avenue, Whanganui,
11 4500, New Zealand

12 **Abstract**

13 Induced seismicity due to natural gas extraction from the Gronin-
14 gen reservoir, the Netherlands, has been occurring since the early 1990s.
15 Prospective forecasts of the possible maximum earthquake magnitude for
16 different gas extraction scenarios have previously been published. How-
17 ever, their accuracy and continued relevance is unknown. Here, we evalu-
18 ate our prior Groningen seismicity forecast and propose a revised model.

19 We compare the forecast to earthquakes that occurred after its publica-
 20 tion, between 1 February 2017 and 1 July 2022. We used volume-scaling
 21 to account for actual gas extraction being 35% less than considered in the
 22 forecast. After this correction, we found that the forecast slightly under-
 23 estimated the number of $M \geq 2.5$ events. However, it was consistent with
 24 the largest event in that period, an M 3.4. Then, we updated the forecast
 25 model for a new depressurization scenario, testing its pseudo-prospective
 26 (out-of-sample) performance using L- and N-tests. This revision consid-
 27 ered two possible magnitude frequency distributions (MFD): (1) tapered
 28 Gutenberg-Richter (GR), where induced ruptures are confined within the
 29 reservoir, and (2) unbounded GR, where ruptures can propagate into the
 30 underburden. A tapered GR is the more likely scenario due to clear
 31 rollover in the MFD and a lack of events beneath the reservoir. However,
 32 we cannot exclude the possibility of an unbounded GR, because direct
 33 stress measurements in the underburden are not currently available. For
 34 the period 2021 to 2030, this new forecast suggests a 1% likelihood that
 35 the largest event will exceed M 3.8 for the tapered model, and M 5.4 for
 36 the unbounded model.

37 **Keywords**

38 Induced seismicity, Groningen, maximum magnitude, prospective forecast, post
 39 hoc evaluation.

40 **Introduction**

41 Around the world, hydrocarbon production and associated activities have led to
 42 sharp increases in seismicity, e.g., Oklahoma, USA [Keranen and Weingarten,
 43 2018]; Alberta, Canada [Schultz et al., 2020]; and Groningen, the Netherlands
 44 [van Thienen-Visser and Breunese, 2015]. The seismic hazard arising from hy-

45 hydrocarbon production highlights the need to anticipate and manage induced
46 earthquakes. But hazard assessments of induced seismicity face some unique
47 challenges [Bommer et al., 2015, Bommer, 2022] as compared to standard haz-
48 ard assessments for natural seismicity. For example, a recent re-evaluation of
49 ground shaking in Oklahoma in 2018 [Petersen et al., 2018] revealed a higher
50 chance of minor and moderate damage compared to some regions of California
51 near the San Andreas Fault.

52 Standard probabilistic assessments of seismic hazard assume that the stochas-
53 tic occurrence of earthquakes is governed by a stationary distribution [Baker
54 et al., 2021]. The mean of this distribution is the regional earthquake rate, ide-
55 ally constrained by a long earthquake catalog that averages out the stochasticity.
56 However, for induced seismicity, the rate of earthquakes can vary dramatically
57 over operational time scales: weeks-to-months for well stimulation [e.g., Häring
58 et al., 2008, Schultz et al., 2020] and years-to-decades for reservoir production
59 [Keranen et al., 2014, Majer et al., 2007, Muntendam-Bos et al., 2022]. To cap-
60 ture this year-on-year variability, the United States Geological Survey forecasts
61 seismic hazard annually [Petersen et al., 2016, 2017, 2018] rather than attempt-
62 ing to compute a single, static hazard map. Apart from natural stochasticity,
63 several production-related processes contribute to the short-term variability in
64 triggered earthquake rate including pore-pressure diffusion [e.g., Hsieh and Bre-
65 dehoeft, 1981, Shapiro et al., 2002, Dempsey and Riffault, 2019], poroelastic
66 stresses [e.g., Chang and Segall, 2016, Goebel et al., 2017, Zhai et al., 2019],
67 and hydraulic fracturing [e.g., Shapiro and Dinske, 2009, Schultz et al., 2020].

68 The prominence of hydrocarbon production as a cause of induced seismicity
69 is both a challenge and an opportunity. It is a challenge because production
70 data can be cumbersome, if not impossible, to assemble and model at a regional
71 scale. But it is an opportunity because carefully considered operational proto-

72 cols could reduce the risk of triggered earthquakes [e.g., Dempsey and Riffault,
73 2019] and mitigate impacts on nearby communities. Realizing this opportunity
74 requires models designed for specific hydrocarbon fields that connect operational
75 protocols to probabilistic forecasts of the associated earthquake rate [e.g., Broc-
76 cardo et al., 2017, Clarke et al., 2019, Li et al., 2022]. One example, and the
77 subject of this study, is the prospective forecast model by Dempsey and Suckale
78 [2017] of Groningen gas field in the Netherlands.

79 Once Europe’s largest producing gas field, concerns about induced earth-
80 quakes have led to year-on-year reductions in Groningen’s output from peak pro-
81 duction in 2013. In 2016, the field operator, Nederlandse Aardolie Maatschappij
82 (NAM), convened a workshop [Coppersmith et al., 2016] to solicit estimates of
83 the maximum earthquake magnitude, $M_{\max}$, that could occur in the field [Bom-
84 mer and Van Elk, 2017]. These estimates, of which the Dempsey and Suckale
85 [2017] forecast was one contribution, were assessed by a separate expert panel
86 who then developed a logic tree for $M_{\max}$. However, static models cannot cap-
87 ture inevitable deviations from planned production levels and evolving sociopo-
88 litical or economic drivers. Nor can such static models take advantage of new
89 data or conceptual understanding to improve production plans. In 2022, NAM
90 organized a second workshop to solicit updated estimates of $M_{\max}$ at Groningen
91 and to develop a new logic tree [Coppersmith et al., 2022], following a similar
92 structure to the 2016 workshop. The study described here was one contribution
93 to the 2022 $M_{\max}$ workshop.

94 Our goal in this paper is to advance the state-of-the-art of induced-seismicity
95 forecasting. We adopt an iterative development philosophy. We start by con-
96 ducting a post hoc analysis of our prior forecasts in Dempsey and Suckale [2017]
97 and then use this analysis as a basis to develop an improved model and a new
98 prospective forecast of induced seismicity for Groningen gas. Our approach is

99 inspired by the value of adopting an inherently dynamic, so-called *living* model
100 [Maldonado et al., 2022] to inform decision-making. Ideally, such models are
101 subject to testing and accuracy checks [Schorlemmer et al., 2007, Király-Proag
102 et al., 2016], post hoc evaluation [Strader et al., 2017, Dempsey et al., 2022] and
103 improvement.

104 A prospective forecast is a probabilistic statement about a specific type of
105 outcome (e.g., maximum earthquake magnitude, total number of earthquakes)
106 happening within a well defined future time period. At Groningen, prospective
107 forecasts have been published by Bourne et al. [2014], Dempsey and Suckale
108 [2017], Bourne and Oates [2017a], Richter et al. [2020], Bourne and Oates [2020]
109 and Limbeck et al. [2021]. A retrospective forecast, sometimes called a hindcast,
110 fits a model to past seismicity data, hence building confidence that the model
111 is accurately capturing seismogenic processes [e.g., Bourne and Oates, 2017b,
112 Candela et al., 2019, Smith et al., 2022, Heimisson et al., 2022]. A third type
113 of forecast is termed pseudo-prospective, wherein a forecast model is trained
114 on one part of the available data (hindcasting) and then evaluated against the
115 withheld (out-of-sample) portion [e.g., Bourne and Oates, 2017b, 2020, Richter
116 et al., 2020, Limbeck et al., 2021, Smith et al., 2022]. This study makes use of
117 all three forecast types.

118 Geological Setting

119 The 900 km² Groningen gas field is located in the northwest of the Nether-
120 lands, extending partially offshore (Fig.1a). The main gas-bearing layer is the
121 Rotliegend Sandstone, located at an average depth of 3 km and varying in thick-
122 ness from southeast to northwest [Bourne et al., 2014]. The reservoir is overlaid
123 by Zechstein salt deposits up to 2 km thick and underlain by Carboniferous
124 limestone. The reservoir is extensively faulted with groups of N-S, E-W, and

125 NE-SW trending normal faults (Fig.1a) with offsets ranging up to the thickness
126 of the reservoir.

127 Natural gas has been produced from the Groningen reservoir since the 1960s.
128 By July 2022, a total of 2250 billion cubic meters (bcm) had been produced,
129 peaking at more than 80 bcm/yr in the late 1970s (Fig.1b). Production peaked
130 again in 2013 at 54 bcm/yr but has since been declining. The mean reservoir
131 pressure has declined more than 25 MPa from its initial value of 35 MPa (Fig.1c).
132 Depressurization has been accompanied by mean reservoir compaction of 30 cm.

133 *Induced seismicity at Groningen*

134 Earthquakes were first detected in the Groningen field in 1991 (Fig.1b). A
135 network extension in 1995 and further station deployment between 2014 and
136 2016 have steadily improved monitoring [Dost et al., 2017, Muntendam-Bos
137 et al., 2022]. By July 2022, a total of 840 $M \geq 1.0$ events have been detected in
138 the field. The largest event to date was a M 3.6 in 2012 [Dost and Kraaijpoel,
139 2013]. While Spetzler and Dost [2017] and Willacy et al. [2019] have suggested
140 that the majority of events were located within the reservoir, a later reanalysis
141 by Smith et al. [2020] has suggested that a substantial proportion of events could
142 have occurred in the overburden, specifically the Zechstein basal anhydrite. All
143 three studies found little evidence for earthquake nucleation in the underburden
144 at Groningen. However, underburden seismicity has been observed in other
145 producing gas fields [Segall, 1989, Bardainne et al., 2008].

146 The nucleation depths of earthquakes is important, because it indicates what
147 physical mechanism could be loading the faults. Earthquake nucleation within
148 the reservoir can be explained by poroelastic stresses induced by depressuriza-
149 tion [Segall and Fitzgerald, 1998, Dempsey and Suckale, 2017] or by compaction
150 gradients that impose shear strains [Bourne and Oates, 2017b]. Earthquake nu-

151 cleation in the overburden can also be explained using poroelastic models with
 152 a different set of parameters [Smith et al., 2022]. Stress changes that favor
 153 nucleation may also locally promote rupture propagation [e.g., Dempsey et al.,
 154 2016, Buijze et al., 2017, Galis et al., 2017]. Ruptures whose dimensions are
 155 confined to the stress-perturbed region are commonly called “induced” to dis-
 156 tinguish them from “triggered” ruptures that propagate without limit and may
 157 hence have a larger event magnitude.

158 One line of evidence for dimensionally-constrained earthquake ruptures is a
 159 relative deficit of large magnitude events in the magnitude frequency distribution
 160 (MFD) [Shapiro et al., 2011, Dempsey et al., 2016]. This effect is sometimes
 161 described as a rollover of the MFD at a corner magnitude – an event size where
 162 limited fault sizes or stress conditions within the field could restrict earthquake
 163 propagation. However, statistically confirming a rollover is difficult when only
 164 a few large events are in the catalogue [Main, 2000]. Rollover of the MFD can
 165 be modelled empirically using variants of the Gutenberg-Richter (GR) power-
 166 law, either by introducing a truncating maximum magnitude [Utsu, 1999] or an
 167 exponential taper [Bourne and Oates, 2020].

168 The MFD properties of Groningen earthquakes have been extensively stud-
 169 ied. Bourne et al. [2014] analysed all $M \geq 1.5$ events between 1995 and 2012, and
 170 found that statistical uncertainty was too great to suggest evidence of a rollover
 171 or taper. They estimated a GR b -value between 0.90 and 1.19. Muntendam-
 172 Bos and Grobbe [2022] provided a more recent analysis and suggested there
 173 are variations in b -value across the field (between 0.8 and 1.5) but that these
 174 are stable in time. They also tested whether bounded GR distributions could
 175 describe the data and found evidence for a corner magnitude of 3.5 for a subset
 176 of events. In a subsequent study by Bourne and Oates [2020], a range of stress-
 177 dependent b -value and stress-dependent taper models were tested. Although

178 the stress-dependent taper had the best fit, the authors caution that a single
179 model cannot be definitively preferred.

180 When the existence or nature of a taper is unclear, then a conservative al-
181 ternative scenario is to consider an unbounded GR distribution. This approach
182 implies that earthquake ruptures are not restricted to the reservoir and can prop-
183 agate into underlying formations. We emphasize that this is not the most likely
184 scenario for earthquakes at Groningen. For example, in addition to evidence
185 for an MFD taper, Smith et al. [2020] estimated the likelihood that relocated
186 hypocentres were below the reservoir to be less than 12%.

187 *Prior models of induced seismicity*

188 Several physical models have been proposed to explain earthquake occurrence at
189 Groningen [Bourne et al., 2014, Dempsey and Suckale, 2017, Bourne and Oates,
190 2017b, Buijze et al., 2017, Candela et al., 2019, Smith et al., 2022, Heimis-
191 son et al., 2022]. Many assume that prior to gas production, ambient shear stresses
192 within the reservoir were well below the critical limit of frictional failure [Jaeger
193 et al., 2009, Zoback, 2010]. Evidence for this relatively low pre-stress is the 25
194 years and 1000 bcm of production at Groningen before the first earthquake was
195 detected (Fig. 1b). An alternative explanation for the delay is that it reflects
196 a long earthquake nucleation time scale that can be calibrated into a rate-and-
197 state friction model. However, Heimis-son et al. [2022] have found that such
198 models struggle to capture the reduction in seismicity following its 2014 peak,
199 and that a stress threshold is needed to correct the issue.

200 To explain induced seismicity at Groningen, models generally need to ac-
201 count for how gas production causes stress to accumulate on faults. The pre-
202 vailing view is that depressurization of the reservoir has caused substantial com-
203 paction (Fig. 1c), which due to heterogeneity of material properties is spatially

204 variable across the field [Bourne et al., 2014]. Lateral gradients of compaction
 205 induce shear strains on faults embedded in the reservoir [Bourne and Oates,
 206 2017b]. These strains are exacerbated on faults that offset reservoir compo-
 207 nents [Buijze et al., 2017, Candela et al., 2019].

208 An alternative model is that depressurization induces a poroelastic contrac-
 209 tion within the reservoir, and this induces large extensional stresses [Segall and
 210 Fitzgerald, 1998, Dempsey and Suckale, 2017]. Smith et al. [2022] developed
 211 a more realistic poroelastic and compaction model that could estimate stress
 212 changes within and above the reservoir. They found that stress changes within
 213 the reservoir were stabilizing whereas changes above were destabilizing, which is
 214 consistent with their hypocenter inversion pointing to earthquakes in the over-
 215 burden [Smith et al., 2020]. Other models have also considered whether rapid
 216 cycling of production wells could induce poroelastic transients that nucleate
 217 earthquakes [Segall and Lu, 2015] but were inconclusive and remain untested
 218 for Groningen [Postma and Jansen, 2018].

219 Once a suitable stress description has been selected, a model is needed for
 220 earthquake nucleation. A popular choice is the rate-and-state friction model of
 221 Dieterich [1994], which has few parameters. Candela et al. [2019] and Richter
 222 et al. [2020] have both applied this model to capture the onset and increase
 223 of seismicity at Groningen. A difficulty with this model is that it assumes
 224 faults begin as critically stressed, which is unlikely to be the case at Groningen.
 225 However, a modified rate-and-state formulation introduced by Heimisson et al.
 226 [2022] has addressed this issue.

227 Coulomb failure is an alternative description of earthquake nucleation. In
 228 this formulation, earthquakes nucleate when shear stress exceeds fault shear
 229 strength. Then, the seismicity rate, λ , is assumed to be proportional to the
 230 stress loading rate, i.e., $\lambda(t) \propto \dot{\tau} - f(\dot{\sigma}_n - \dot{p})$, where τ and σ_n are shear and

normal stresses, p is fluid pressure, and f is a friction coefficient. This model has been used by Bourne and Oates [2017b], Dempsey and Suckale [2017], and Smith et al. [2022] to provide an acceptable fit to the historical earthquake rate in Groningen.

Once an earthquake is nucleated, a model is needed to estimate its magnitude. A straightforward approach is to sample this directly from an assumed MFD for the field [e.g., Bourne and Oates, 2017b, Heimisson et al., 2022]. An alternative is to simulate the relevant fault friction and stress transfer physics [Dempsey and Suckale, 2017, Buijze et al., 2017], which can provide insights into limits on rupture dimension and hence MFD rollover. To be useful for forecasting, the simulation approach needs to be fast, which tends to favor either empirical models or highly reduced descriptions of earthquake propagation physics.

Methodology

Below, we summarize the particular model parameterizations we have used to approximate seismogenesis at Groningen. These include descriptions of the tectonic stress field, seismogenic faults, a poroelastic model that links gas production to fault loading, and simulations of source nucleation and size (Fig. 2). Additional model details can be found in Dempsey and Suckale [2017] or in the Electronic Supplement to this article. For brevity, we denote the Dempsey and Suckale [2017] study as DS17, and the updated model as DS22.

To characterize the pre-production stress, we assumed an extensional tectonic regime in which the vertical stress, σ_1 , is given by depth integration of a constant overburden density (2500 kg/s) to the average reservoir depth (3 km). The minimum horizontal stress, σ_3 , is rotated an angle θ clockwise from North. Its lowest possible value, σ_{crit} , is limited by brittle failure of the crust with fric-

257 tion coefficient, f [Jaeger et al., 2009]. Its actual value is parameterized by h ,
 258 where $h = 0$ is a critically stressed crust and $h = 1$ is an isotropic stress state.
 259 The magnitude of the maximum horizontal stress, σ_2 , is parameterized by H to
 260 be between the vertical and minimum horizontal stresses, with $H=0$ denoting
 261 $\sigma_2 = \sigma_3$ and $H=1$ denoting $\sigma_2 = \sigma_1$.

262 Summarizing, we impose a homogeneous pre-production stress fully de-
 263 scribed by four parameters: f , h , H , and θ . This controls the initial shear
 264 and normal stresses resolved on individual faults. We then impose further stress
 265 changes based on gas extraction and the associated pressure decline in the reser-
 266 voir, $\Delta p(\mathbf{x}, t)$ (Fig. 2a). Depressurization has been modeled by NAM using a
 267 two-component (water and gas) reservoir simulator that accounts for historic ex-
 268 traction and has been history matched to pressure and subsidence data. Three
 269 future depressurization scenarios were provided by NAM as an input to the DS17
 270 model, corresponding to extraction of 21, 27 and 33 bcm/yr between February
 271 2017 and 2025.

272 We used the poroelastic model of Segall and Fitzgerald [1998] to link de-
 273 pressurization to reservoir stress changes. We approximated the reservoir as an
 274 Eshelby [1957] ellipsoidal inclusion that would contract as pressure drops were
 275 it not elastically coupled to the surrounding rock. Instead, the reservoir de-
 276 velops an extensional strain within and subsidence occurs above (Fig. 2c). We
 277 calculated the extensional strain locally via a poroelastic stress path coefficient,
 278 A [Segall and Fitzgerald, 1998, Zoback, 2010] and added this perturbation to
 279 the tectonic stress, $\Delta \sigma_h(\mathbf{x}, t) = A \Delta p(\mathbf{x}, t)$ where $h=2$ or 3 .

280 Faults embedded in the regional stress field experience a time and space
 281 varying load (Fig. 2b). We have approximated faults as straight line elements.
 282 The i^{th} fault has a normal vector, $\hat{\mathbf{n}}_i$, related to its strike, φ_i , and dip, ϕ_i , and
 283 a length, L_i . The evolving stress field causes normal, $\sigma_n(s, t)$, and shear stress,

284 $\tau(s, t)$, on the fault to vary with position and time (s is a coordinate along the
 285 strike of the fault). Our models included the 325 longest faults in Groningen
 286 (Fig. 1a), with a minimum fault length of 2 km.

287 In the DS17 model, the total poroelastic loading across all faults defined
 288 a time-varying seismicity rate for the field, $\lambda(t)$. Individual poroelastic loads
 289 were then imposed as inputs to a 1D fracture mechanics simulator that approxi-
 290 mated earthquake nucleation and rupture on heterogeneous faults with constant
 291 vertical thickness. The output of this simulator was an ensemble of possible in-
 292 duced earthquakes for each fault, each with a different hypocenter, nucleation
 293 time, and magnitude larger than a minimum of M 1.0. Simulated earthquake
 294 were obtained for each of the 325 modeled faults and combined to create syn-
 295 thetic catalogs. Interactions between earthquakes on separate faults were not
 296 considered.

297 *Improvements in the DS22 model*

298 The major updates we have introduced are: (i) a revised depressurization sce-
 299 nario provided by NAM (see Electronic Supplement), (ii) representation of the
 300 faults' vertical dimension, which varies with reservoir thickness, (iii) replacement
 301 of the fracture-mechanics earthquake simulator with two empirical MFDs, (iv)
 302 model calibration that incorporates events locations, and (v) pseudo-prospective
 303 testing to estimate forecast accuracy.

304 *Earthquake rate model*

305 We define the rate density, ξ , as the seismicity rate per along-strike fault length.
 306 We assumed each fault had a time and space varying rate density of $M \geq 1.0$
 307 seismicity, $\xi_i(s, t)$, which is proportional to the local stressing rate:

$$\xi_i(s, t) = B \times CFS_i(s, t), \quad CFS_i(s, t) = \dot{\tau}_i(s, t) - f(\dot{\sigma}_{n,i}(s, t) - \dot{p}_i(s, t)), \quad (1)$$

where i denotes the fault index, s is an along-strike coordinate, CFS is the Coulomb Failure Stress that parameterizes the proximity of the fault to frictional failure, and $CFS > 0$ is the threshold condition for earthquake nucleation. The magnitude and evolution of stress components on a fault depends on the fault's orientation within the background stress field and its location within the depressurizing reservoir. The parameter B is a constant of proportionality determined later.

Faults in the DS22 model had a vertical dimension that was set by the thickness of the Groningen reservoir, increasing approximately linearly from southeast to northwest. This is important because faults in the thicker part of the reservoir have a larger surface area exposed to stressing. Hence, in our model, they have a higher rate of earthquake nucleation for the same stressing rate.

We incorporated reservoir thickness as a weight, $w(s)$, multiplied to each fault's local rate density, $\xi(s, t)$. We set the weight equal to one when the local thickness, $h(s)$, was equal to the reference thickness of 223 m, i.e.

$$w(s) = \frac{h(x(s), y(s))}{h(250, 592)}, \quad h(x, y) = -3.04(x - 250) + 4.11(y - 592) + 223. \quad (2)$$

We derived this expression by fitting to the reservoir thickness map presented in Bourne and Oates [2017b].

A new depressurization scenario, $\Delta p(x, y, t)$, was provided by NAM for the construction of forecasts out to 2030. We applied this as a loading that produced

328 a rate density for the i^{th} fault. We calculated a total seismicity rate for the fault
 329 by integrating its rate density over its length

$$\lambda_i(t) = \int_0^L \xi_i(s, t) ds. \quad (3)$$

330 The seismicity rate in a spatial bin can be calculated by modifying the integral
 331 limits in Eq. (3) to just that section of a fault within the bin.

332 We calculated the seismicity rate for the entire reservoir by summing the
 333 contributions of all faults

$$\lambda(t) = \sum_{i=1}^{325} \lambda_i(t). \quad (4)$$

334 We determined the constant of proportionality, B , by ensuring the total
 335 modeled seismicity over the training period, $[1960, t_{\text{train}}]$, matched the observed
 336 number of $M \geq 1.0$ earthquakes in that period, N_{train}

$$B = N_{\text{train}} / \int_{1960}^{t_{\text{train}}} \lambda(t) dt. \quad (5)$$

337 For pseudo-prospective tests, we set t_{train} to 1 February 2017 and N_{train} to
 338 670. For the DS22 prospective forecast model, we set t_{train} to 1 January 2021
 339 and N_{train} to 820.

340 We estimated the expected number of earthquakes, $\bar{N}_{\text{fcst}}$, in the forecast
 341 period, $[t_{\text{start}}, t_{\text{end}}]$, through a similar integral

$$\bar{N}_{\text{fcst}} = \int_{t_{\text{start}}}^{t_{\text{end}}} \lambda(t) dt. \quad (6)$$

342 For pseudo-prospective tests, we set t_{start} to 1 February 2017 and t_{end} to 1
 343 July 2022, i.e., the DS17 evaluation period. For the DS22 prospective forecast,
 344 we set t_{start} to 1 January 2021 and t_{end} to 1 January 2030.

345 *Synthetic catalogs and forecasts*

346 A useful forecast product is the exceedance probability curve, which quantifies
347 the likelihood that, in a defined time interval, there will be an earthquake whose
348 characteristics exceed a certain level. We developed exceedance curves by con-
349 sidering the statistical properties of synthetic earthquake catalogs obtained from
350 our model. Each catalog represents a possible future realization of seismicity at
351 Groningen.

352 To construct a synthetic catalog, first, we calculated its size, N_{fcst} , by sam-
353 pling a random integer from a Poisson distribution with mean, $\bar{N}_{\text{fcst}}$. Thus,
354 future catalogs have an uncertain size but with an average determined by the
355 future seismicity rate, $\lambda(t)$. Second, for each event in the catalog, we assign it a
356 magnitude, $M_i = [M_1, M_2, \dots, M_{N_{\text{fcst}}}]$, by sampling from the earthquake simula-
357 tor ensemble (for DS17) or an empirical MFD (for DS22). Third, we record the
358 largest event in a catalog, $M_{\text{max}} = \max(M_i)$, and the number of $M \geq 2.5$ events,
359 $N_{2.5} = \text{count}(M_i \geq 2.5)$. Fourth, we construct 10^5 such catalogs and calculate
360 a cumulative distribution function for the set of values M_{max} and $N_{2.5}$. These
361 are the forecast exceedance curves.

362 *Empirical magnitude model*

363 The DS17 model could only represent earthquakes whose ruptures were confined
364 to the reservoir layer. However, if earthquakes are able to rupture outside of the
365 reservoir and into basement formations, then they could have larger magnitudes.
366 In the DS17 model, this dimensional constraint manifested as rollover in the
367 modelled MFD. Although this is consistent with observations at Groningen,
368 excluding the possibility of underburden ruptures introduces uncertainty into
369 our forecasts.

370 For DS22, we decided to generate earthquake magnitudes by randomly sam-

371 pling from an empirical MFD. We considered two possible MFDs: (1) tapered
 372 and (2) unbounded GR. They represent two approximate scenarios where earth-
 373 quake source dimensions are either constrained to the reservoir or can rupture
 374 into the basement. We selected the distributions and their parameters from the
 375 Bourne and Oates [2020] study of Groningen seismicity. A detailed description
 376 can be found in the original article and the Electronic Supplement.

377 ***Model history-match and testing***

378 *Temporal history match*

379 We history-matched our forecast models to the induced earthquake catalog
 380 maintained by KNMI (see Data and Resources). The DS17 model was calibrated
 381 to the event time information using a Markov-chain Monte Carlo (MCMC) al-
 382 gorithm [Foreman-Mackey et al., 2013] to minimize model misfit. Model misfit
 383 was approximated by the log-likelihood (LLK) that scores a non-homogeneous
 384 Poisson process (NHP, the modeled seismicity rate, $\lambda(t)$) against a set of earth-
 385 quake times [e.g., Bourne and Oates, 2017b, Dempsey and Riffault, 2019, Richter
 386 et al., 2020]

$$LLK_{\text{nhp}} = \sum_{i=1}^N \log \lambda(t_i) - \int_{t_1}^{t_N} \lambda(t) dt. \quad (7)$$

387 Larger values of LLK_{nhp} denote progressively better fits of the model to the
 388 data. An advantage of LLK_{nhp} is that, when evaluated over parameter space, it
 389 can be interpreted as a parameter posterior. This in turn supports probabilistic
 390 sampling to create ensembles of seismicity rate models.

391 A drawback of temporal history matching is that it does not make use of the
 392 information in earthquake location data. There are advantages in requiring the
 393 model to also honour event locations. For instance, if certain fault orientations

dominate in parts of the field, then hypocenter data may provide clues about which fault orientations were inclined to fail first. This in turn can constrain parameters related to stress field orientation.

Spatio-temporal history match

For the DS22 model, we extended history-matching to account for both earthquake times and locations. To do this, we quantified the model misfit using a modified Gaussian log-likelihood. This approach has also been used by Smith et al. [2022] and Heimisson et al. [2022] to quantify misfit of Groningen seismicity models. The Gaussian LLK over N observations is written, up to a constant:

$$LLK_{\text{gs}} = - \sum_i^N \frac{(\hat{x}_i - x_i)^2}{2\sigma_i^2} \quad (8)$$

where $\hat{x}_i$ is an observation, x_i is the model approximating the observation, and σ_i is the uncertainty associated with the observation.

Our model misfit, S , was comprised of four weighted terms associated with misfits to: (1) S_{t0} , the modelled event rate, λ_t , from the beginning of production in 1960 to just before peak seismicity in 2012, (ii) S_{t1} , modelled event rate during peak seismicity and the post-2014 decline, (3) S_x , modelled spatial event density, ρ_x , in the x direction, and (iv) S_y , the event density in the y direction. A weighting term, ω , allowed for relative prioritisation of the temporal and spatial matches.

$$\begin{aligned}
S &= \omega(S_{t0} + S_{t1}) + (1 - \omega)(S_x + S_y), \\
S_{t0} &= -\frac{1}{51} \sum_{1960}^{2011} \frac{(\hat{\lambda}_t - \lambda_t)^2}{2\hat{\lambda}_t}, \\
S_{t1} &= -\frac{1}{9} \sum_{2012}^{2021} \frac{(\hat{\lambda}_t - \lambda_t)^2}{2\hat{\lambda}_t}, \\
S_x &= -\frac{1}{20} \sum_1^{20} \frac{(\hat{\rho}_x - \rho_x)^2}{2\hat{\rho}_x}, \\
S_y &= -\frac{1}{20} \sum_1^{20} \frac{(\hat{\rho}_y - \rho_y)^2}{2\hat{\rho}_y}.
\end{aligned} \tag{9}$$

We constructed the event rate observations, $\hat{\lambda}_t$, by binning all $M \geq 1.0$ seismicity at one year intervals. We constructed the event density observations, $\hat{\rho}_i$ where $i = [x, y]$, by binning all seismicity at either 2.1 km intervals in the x direction or 2.5 km intervals in the y direction. This corresponded to 20 bins in both directions across the study area. We assumed events in each bin were generated by a local Poisson process whose associated variance was equal to the observation, hence $\sigma_i^2 = \hat{\lambda}_t$ or $\hat{\rho}_i$. We set a minimum of variance of 1 to avoid divide by zero issues in bins with zero observations.

Although certain choices we have made in the construction of S may seem arbitrary, they are all practical decisions to improve model calibration. For instance, separating pre- and post-peak seismicity (S_{t0} and S_{t1}) made it easier for us to find models that capture this abrupt reversal in rate. In this study, we set ω to 0.75, which prioritised models that produced a good temporal match. However, within models with approximately similar temporal matches, we were still able to select those that also captured the spatial distribution. Our decision to separate out matches in time and two space dimensions was intentional to maximise the amount of data in each bin.

430 We calibrated the DS22 seismicity model using the Differential Evolution
 431 algorithm of Storn and Price [1997] as implemented in the Scipy Python library
 432 [Virtanen et al., 2020]. This is a genetic algorithm that minimizes a misfit func-
 433 tion, in this case the expression S (Eq. 9). It does this by sampling parameters
 434 from a bounded space $[0 < h < 1, 0 < H < 1, 0^\circ < \theta < 90^\circ, 0.5 < f < 0.8, 0.8 <$
 435 $A < 1.0]$, saving the models with low misfit, and then recombining their pa-
 436 rameters to search for improved fits. The optimization is global and was used
 437 with default parameter specifications and stopping criteria. The result is a large
 438 ensemble of possible model parameter combinations, each ranked by their misfit
 439 score.

440 *Model testing*

441 To measure the accuracy of our forecast model, we compare modelled and mea-
 442 sured seismicity rates, either visually or quantified by a misfit function. In
 443 minimizing misfit, history-matching increases model accuracy over the training
 444 period, but does not guarantee the model’s accuracy when comparing to fu-
 445 ture seismicity. To understand future performance, we test our model pseudo-
 446 prospectively by history-matching only a subset of the data and quantifying
 447 the model’s accuracy against the withheld data. Schorlemmer et al. [2007] rec-
 448 ommend two tests of forecast accuracy that can be used in prospective and
 449 pseudo-prospective settings. The L-test, which matches the modelled seismicity
 450 rate to the data, and the N-test, which checks for over- or underestimates of
 451 the number of earthquakes. Both tests were trialled by Candela et al. [2019] for
 452 hindcasting of a Groningen model.

453 The L-test considers whether the catalog of N event times, $[t_1, \dots, t_N]$, is
 454 consistent with a particular rate model, $\lambda(t)$. We score the model against the
 455 data using the NHP log-likelihood (Eq. 7). To evaluate model performance,
 456 Schorlemmer et al. [2007] proposed creating synthetic earthquake catalogs from

457 $\lambda(t)$, each with their own event times, $[\hat{t}_1, \dots, \hat{t}_M]$, where M could be different
 458 between catalogs. The synthetic catalogs were also scored with a likelihood. The
 459 distribution of synthetic scores provides a reference for evaluating the observed
 460 score, quantified as a quantile within this distribution. If the quantile is at the
 461 low end (e.g., <0.05), then the model provides a poor fit to the data; quantiles
 462 at the high end are acceptable [Schorlemmer et al., 2007]. In this study, the
 463 quantile is equivalent to $1-p$ where p is a probability along the exceedance curve.

464 In contrast, the N-test checks whether the total number of earthquakes fore-
 465 cast by a model for a given period is consistent with the number of earthquakes
 466 that actually occurred, N . This is achieved by comparing N to the distribution
 467 of catalog sizes, M , for the many synthetic catalogs constructed for the L-test.
 468 A quantile of N at the low and high ends of this distribution corresponds to
 469 models that, respectively, are over- and underestimating the earthquake rate.

470 Results

471 *Post hoc evaluation of DS17 forecast*

472 Dempsey and Suckale [2017] describe three forecasts produced by the DS17
 473 model by extrapolating over the period 1 February 2017 to 1 January 2025
 474 (Fig. 3a). Forecast seismicity was driven by different rates of depressurization
 475 corresponding to gas extraction averaging 21, 27 and 33 bcm/yr. Synthetic cat-
 476 alogs were sampled and used to construct exceedance curves for (1) the number
 477 of $M \geq 2.5$ earthquakes (Fig. 3b), and (2) the largest magnitude event in the
 478 forecast period, $M_{\max}$ (Fig. 3c). For example, for the 21 bcm/yr scenario, there
 479 was a 50% likelihood of 14 or more $M \geq 2.5$ events over the forecast period with
 480 an $M_{\max}$ of 3.4 or higher. There was a 5% likelihood of 21 or more $M \geq 2.5$
 481 events with an $M_{\max}$ of M 3.9 or higher.

There were several challenges in retrospectively evaluating the DS17 forecast. First, actual gas extraction between 1 February 2017 and 1 July 2022 averaged 13.7 bcm/yr, which is about 35% less than the lowest scenario originally considered (21 bcm/yr). Second, we conducted our evaluation for the first 5.5 years of the nearly eight-year forecast period. Finally, the DS17 forecast is a probability distribution over future outcomes specifying the number of $M \geq 2.5$ events or maximum event size. We only have one opportunity to observe the future, but one sample is not sufficient to confirm a probability distribution. At best, if the actual outcome sits well outside the forecast, we might question its quality.

We partly address the reduced extraction and forecast window issues through a scaling argument. It is clear that gas extraction at a larger volumetric rate results in a higher rate of earthquakes (Fig. 3a). An inspection of DS17 modelled earthquake rates in the period 2017 to 2020 shows that the three extraction scenarios at 21, 27, and 33 bcm/yr correspond to ensemble average earthquake rates of about 35, 45 and 55 $M \geq 1.0$ events per year. Thus, the relative earthquake rates are in proportion to the relative gas extraction rates. This assumption is the basis of the Zöller and Holschneider [2016] forecast.

If, in 2017, we had had the foresight to construct a forecast for 13.4 bcm/yr gas extraction over nearly 5.5 years (total extraction of 72.5 bcm), then evaluation would be straightforward. In practice, the closest we have is an eight year forecast at 21 bcm/yr (total extraction of 168 bcm). Applying a volume scaling argument, we should therefore expect the number of earthquakes to be about 43% of the 21 bcm/yr forecast given in DS17. This would correspond to a 50% likelihood of 6 or more $M \geq 2.5$ events and a 5% likelihood of 9 or more. The actual number of $M \geq 2.5$ events between 1 February 2017 and 1 July 2022 was 10 (Table 1), which places this outcome at about the 2% ex-

509 ceedance level of the volume-scaled DS17 forecast. This exercise is equivalent
510 to the N-Test [Schorlemmer et al., 2007] and suggests the DS17 model has been
511 slightly underestimating the rate of $M \geq 2.5$ events.

512 There are a few possible interpretations of this result. The first is that it
513 can be attributed as a plausible, if rare outcome, within acceptable bounds of
514 stochastic uncertainty. In other words, the earthquakes during this period were
515 a 1 in 50 event and such outcomes do occur from time to time. In this case, the
516 N-test is fair and the conclusion should be to accept the forecast. This is fine
517 as a hypothesis, although it cannot be tested without constructing, waiting and
518 retrospectively testing many more forecasts. That is not very practical, nor a
519 particularly satisfying explanation.

520 The second explanation is that the volume scaling argument, although con-
521 sistent with the narrow range of tested extraction scenarios (21 to 33 bcm/yr),
522 is less accurate when extrapolated outside this range, to an effective rate of
523 13.4 bcm/yr. In this case, the N-test is not really fair because the forecast has
524 been misrepresented by volume scaling. The third explanation is that a de-
525 ficiency in the model had lead to an underestimate of $M \geq 2.5$ events in the
526 forecast period. This could be due to limitations of the reservoir simulator that
527 produced the depressurisation scenarios, the simple poroelastic model we used
528 to convert pressure decline into stress changes, or in the earthquake source sim-
529 ulator itself. In this case, the N-test is fair and the conclusion should be to
530 reject the forecast.

531 Another forecast output is the maximum event magnitude. Unlike $M \geq 2.5$
532 events, this quantity is not straightforward to rescale given known variation
533 from the original scenarios, because $M_{\max}$ depends on the statistics of extreme
534 value distributions. It is also influenced by any truncation at high magnitudes
535 in the MFD, which is likely to be relevant for Groningen forecasts. Therefore,

536 the best argument we can offer here is a qualitative one.

537 Considering the $M \geq 2.5$ event forecast for the 21 and 33 bcm/yr extrac-
538 tion scenarios (Fig. 3b), the major parts of their respective exceedance curves
539 span from 10 to 25 $M \geq 2.5$ events. Across this range, there is a difference of
540 about 5 events between the two scenarios. In comparison, exceedance curves
541 for the largest event are more similar across the scenarios, differing by about
542 0.05 magnitude units. We find that in the GR and truncated distribution, the
543 largest events are also the rarest outcomes. An order of magnitude increase in
544 the size of catalog is required to induce a magnitude increment in the largest
545 event. Here, the observed catalog, comprising 10 $M \geq 2.5$ events, is of a similar
546 size to the scenario medians in DS17 (between 15 and 20 events). For this rea-
547 son, the largest event exceedance curves given in DS17 are likely to be only a
548 marginally too-high representation of the 13.7 bcm/yr scenario. In this context,
549 the maximum event size magnitude of M 3.4 in the evaluation period appears
550 consistent with the DS17 forecast.

551 *Updated DS22 forecast*

552 The new forecast by our DS22 model accounts for revised depressurization,
553 history matching to earthquake locations, a choice of two empirical MFDs, and
554 pseudo-prospective testing.

555 *Pseudo-prospective testing*

556 We estimated the forecast skill of the DS22 model under pseudo-prospective
557 conditions. First, we used the earthquake catalog prior to 1 February 2017 to
558 train a minimum misfit seismicity model. This required that we reduce the
559 upper time limit of the term S_{t1} in Eq. (9) from 2021 to 2017 and rebalance the
560 weighting factor from 1/9 to 1/5. Then, we used data between 1 February 2017

561 and 1 July 2022 to evaluate the trained model.

562 Fig. 4(a) shows that the trained model replicates the onset of seismicity in
563 the 1990s, growth through the early 2000s, peak in 2014 and then decline to
564 2017. Seismicity onset is late by about two years, but the pattern is a reasonable
565 match thereafter. During the test period, seismicity declines steadily at a rate
566 that appears consistent with the data. This model had a similar initial stress
567 distribution to DS17 [see Dempsey and Suckale, 2017, fig 2(c-e)] but has a lower
568 friction coefficient. Like the DS17 model, the poroelastic stress path parameter
569 is higher than reported for real reservoirs. In our opinion, this deviation likely
570 reflects missing physics in the model.

571 Fig. 5 presents modelled earthquake area density for the minimum misfit
572 model we used for pseudo-prospective testing. The spatial distribution approx-
573 imately captures two clusters of seismicity in the center and southwest of the
574 field, although small underestimates of the event density are noted.

575 We have performed the L-test for withheld events in the test data set. The
576 observed likelihood had a value of 460, which places it at the high end (4%
577 probability to exceed) of the synthetic distribution of likelihoods (Fig. 4(b)).
578 On this basis, the rate model can be judged consistent with the data.

579 The N-test is a second check on the rate model, in this case to check whether
580 it is over- or underestimating the total number of events (Fig. 4(c)). We con-
581 structed exceedance curves for the number of $M \geq 2.5$ events for both MFD
582 models in the evaluation period. The actual number of $M \geq 2.5$ events that
583 occurred during this 5.5 year period was 10. This outcome had a 9 and 11%
584 probability to exceed, respectively, for the tapered and unbounded GR models.

585 Finally, we considered the exceedance curve for the largest magnitude event
586 (Fig. 4(d)). Both MFD models were consistent with the largest M 3.4 event
587 that occurred during the evaluation period. This outcome had a 34 and 60%

588 probability to exceed, respectively, for the tapered and unbounded GR models.

589 Based on L-test, N-test and $M_{\max}$ exceedance test, we find no clear evidence
590 to reject this model. Therefore, we have proceeded to use it for prospective
591 forecasting.

592 *Prospective forecast: 2021 to 2030*

593 We constructed the prospective DS22 forecast by retraining to obtain a mini-
594 mum misfit model using earthquakes up to 1 January 2021. The resulting model
595 had very similar parameters to the one trained during pseudo-prospective test-
596 ing (Table 2). The extrapolated seismicity rate in the period 2021 to 2030 shows
597 a steadily declining seismicity rate, dropping from 20 to 10 $M \geq 1.0$ events per
598 year.

599 For this model, we constructed synthetic catalogs for the forecast period
600 and used these to define exceedance curves for both MFD scenarios. Both the
601 tapered and unbounded GR models produced similar estimates for the number
602 of future $M \geq 2.5$ events, with a 10% chance to exceed 8, a 1% chance to exceed
603 11, and a 0.1% chance to exceed 13 or 14 (Table 3).

604 In contrast, our estimate of $M_{\max}$ varies considerably between the two MFD
605 models. The tapered GR model has a rollover at M 3.5, which exerts an upper
606 constraint on $M_{\max}$ at very low exceedance probabilities, e.g., 0.1% chance to
607 exceed 3.9. In comparison, the unbounded GR model has no such constraint,
608 indicating larger $M_{\max}$ values of 4.3, 5.4 and 6.5 at 10, 1 and 0.1% exceedance.

609 An analysis of model sensitivity to key parameters has been included in the
610 Electronic Supplement. We found that the model could achieve a reasonable
611 match to the data even when one of its parameters was knowingly assigned an
612 incorrect value. However, this has a minimal effect on forecast outputs, partic-
613 ularly $M_{\max}$, which are largely constrained by the history matching process.

614 Discussion

615 *Possible maximum magnitude*

616 Our goal in this study was to estimate the *possible* maximum magnitude of seis-
617 micity at Groningen. This is different from the *expected* maximum magnitude
618 [Zöller and Holschneider, 2016]. We take the expected maximum magnitude as
619 a distribution expectation, that is, the magnitude associated with the 50% ex-
620 ceedance probability on the cumulative distribution function for $M_{\max}$ (Fig. 6C
621 or the top row of Table 3). For our tapered GR model, this is an M 3.2 event,
622 and for our unbounded GR, it is M 3.5. These values are similar to those ob-
623 tained by Bourne and Oates [2020] (their table 4), who tested a range of MFD
624 models.

625 We take “possible” in possible maximum magnitude to mean an outcome
626 with finite probability. In this sense, all values along the exceedance curve
627 are possible maximum magnitudes, each occurring with increasing degrees of
628 improbability. For our tapered GR model, possible magnitudes increase very
629 little from 10 to 0.1% exceedance, from M 3.6 to 3.9. For our unbounded GR
630 model, possible maximum magnitudes are higher, e.g., M 5.4 at 1% exceedance.
631 This is at the top end of the range of the various taper models tested by Bourne
632 and Oates [2020] for the same exceedance probability.

633 For our analysis, the choice between tapered or unbounded GR model is the
634 most significant determinant of $M_{\max}$. We decided to include an unbounded
635 MFD despite it being less likely for two reasons: First, it is possible that
636 unbounded GR is the appropriate generating MFD for Groningen seismicity
637 to date and the apparent absence of larger magnitudes is an unlikely statisti-
638 cal artefact. Previous studies [Bourne and Oates, 2020, Muntendam-Bos and
639 Grobde, 2022] have judged this situation improbable – we are inclined to agree
640 – but none of them have confidently ruled out the unbounded GR. In the logic

tree constructed by the expert panel of the 2022 Mmax workshop [Coppersmith et al., 2022], a probability of 10% was assigned to the scenario that ruptures could occur within the underburden.

The second argument for considering an unbounded MFD is that, although tapered GR is likely a better description of seismicity to date, we cannot confidently assess that future seismicity will not later switch to an unbounded distribution. For example, deepening or widening of pressure depletion could impose a larger poroelastic loading across a greater underburden volume [Segall, 1992]. Ultimately, such critical uncertainties are best investigated with direct observational evidence such as drilling into the underburden, performing pressurization tests and borehole image analysis to determine in-situ stress conditions, or obtaining drill-core to test for velocity-weakening frictional properties. Until these data are obtained it is, in our view, unwise to categorically rule out the possibility of unbounded ruptures.

Other definitions of the possible maximum magnitude appeal to inherent physical limits of the seismogenic system. For example, Bourne et al. [2014] estimated an $M_{\max}$ of 6.5 corresponding to all induced compaction strain were released in a single event. The longest fault included in the DS17 model could potentially generate an $M_{\max}$ of 6.4, although this never happened in the very large number of catalogs generated in that study. Compared to $M_{\max}$ defined probabilistically, it is less clear how to estimate relative likelihoods for these competing models. An expert-led process that estimates weights in a logic tree [Bommer and Van Elk, 2017] is one way to do this, e.g., the deliberations of the 2022 Mmax expert panel that assigned a 10% weight to the unbounded-rupture branch [Coppersmith et al., 2022].

666 *Correctness of model physics*

667 The models we have presented do not include all possible physics relevant to
668 induced seismicity at Groningen. Arguably, the major physical shortcoming has
669 been our decision not to incorporate differential compaction in the reservoir as
670 an additional load on faults [e.g., Bourne et al., 2014, Bourne and Oates, 2017b,
671 Richter et al., 2020]. Incorporating this effect in future updates has the potential
672 to improve the model’s spatial match to seismicity because compaction has a
673 different pattern than pressure decline [e.g., Bourne et al., 2014, fig. 9].

674 Our study, as well as those of Bourne and Oates [2017b] and Smith et al.
675 [2022] have used Coulomb stressing rates as a proxy for earthquake nucleation
676 rather than implementing a rate-state nucleation model as others have done
677 [Candela et al., 2019, Richter et al., 2020, Heimissson et al., 2022]. Both our
678 model and prior studies [Bourne and Oates, 2017b, Smith et al., 2022] produce
679 a reasonable match to the data, which means model misfit alone is a poor basis
680 for preferring one set of physics over the other. Rate-and-state models may
681 struggle to replicate the sharp decline in seismicity after 2014 (e.g., Candela
682 et al. [2019] fig. 6 and Richter et al. [2020] fig. 9), although this may have been
683 solved with modified formulations that introduce stress thresholds [Heimissson
684 et al., 2022].

685 Despite the different choices available for constructing seismicity rate models,
686 it is not clear these will have a major impact on the output forecast for $M_{\max}$.
687 This is because, in fitting to the same data, all such models are taught to capture
688 first-order features of the seismicity time series: its onset, rise, peak, and decline.
689 Thus, the historical context exercises quite a strong constraint on their future
690 behaviour. Ultimately, we expect that minor differences in modelled earthquake
691 rates arising from different physical formulations will be small compared to the
692 uncertainty in the MFD.

693 We emphasize that our study is one of several published forecast models
 694 for induced seismicity at Groningen [Bourne et al., 2014, Zöller and Holschnei-
 695 der, 2016, Bourne and Oates, 2017b, Shapiro et al., 2017, Candela et al., 2019,
 696 Richter et al., 2020, Limbeck et al., 2021, Smith et al., 2022, Heimisson et al.,
 697 2022]; all combining different physics and forecasting methods. Collectively,
 698 these can be viewed as a model ensemble with certain conferred advantages
 699 [Marzocchi et al., 2012, Taroni et al., 2014]. For quantities such as the post
 700 2014-peak seismicity rate, several of the forecasts have developed consensus,
 701 which outlier models can use as a reference to evaluate their own methods. At
 702 the same time, the consensus has an opportunity to observe the outliers and
 703 consider whether they are revealing any particular blindspots in the conceptu-
 704 alization of the majority.

705 Conclusion

706 We have undertaken a post hoc evaluation of the Dempsey and Suckale [2017]
 707 Groningen seismicity forecast (termed DS17) for the period 1 February 2017 to
 708 1 July 2022. We have also developed a new prospective forecast for 2021 to
 709 2030 (termed DS22). Our goal with both forecast models was to estimate an
 710 exceedance probability curve for the largest event in the forecast period, $M_{\max}$,
 711 using a physics-based model that links gas extraction and depressurization to
 712 earthquake nucleation on reservoir faults.

713 Our updated DS22 model incorporated earthquake locations and a new de-
 714 pressurization plan, and performed well in out-of-sample (pseudo-prospective)
 715 testing. A critical uncertainty that remains was whether future induced ruptures
 716 would be dimensionally bounded by the reservoir. Studies to date [Bourne and
 717 Oates, 2020, Smith et al., 2020, Muntendam-Bos and Grobbe, 2022, Copper-
 718 smith et al., 2022] have concluded that rupture into underlying Carboniferous

719 limestone is unlikely, but none have ruled it out as a possibility. This uncertainty
720 is particularly consequential to the forecast, with a 1% probability to exceed M
721 3.8 for bounded ruptures and M 5.4 for unbounded, between 2021 and 2030.
722 We adopt the consensus view that bounded ruptures ($M_{\max}$ 3.8) should be the
723 preferred model. We recommend that direct observation and testing of the
724 underburden could be a practical approach for reducing this uncertainty.

725 Data and Resources

726 The earthquake catalog used in this study can be obtained from KNMI at http://cdn.knmi.nl/knmi/map/page/seismologie/all_induced.csv. Gas extrac-
727 tion volumes can be obtained from NAM at <https://www.nam.nl/gas-en-olie/gaswinning.html>. The depressurization scenario used in the updated forecast
728 has been included in the Electronic Supplement. All other data are available
729 from cited references in the text. The code needed to recreate the results of
730 this study is available at <https://doi.org/10.5281/zenodo.7109337>. The
731 full dataset made available to participants of the Mmax workshop is available
732 at <https://public.yoda.uu.nl/geo/UU01/RHHRPY.html>. Supplemental Ma-
733 terial for this article includes additional details about the seismological model
734 and a sensitivity analysis.

737 Declaration of Competing Interests

738 This research was completed and published under contract to NAM as a contri-
739 bution to the Second Workshop on Mmax for Seismic Hazard and Risk Analysis
740 in the Groningen gas field.

741 **acknowledgments**

742 The authors acknowledge Nederlandse Aardolie Maatschappij (NAM) for sup-
743 port and data that enabled this study. J Bommer and J van Elk are acknowl-
744 edged for their organisation of the 2016 and 2022 Mmax workshops. S Oates is
745 acknowledged for compiling the data package used in this study. Constructive
746 comments by two anonymous reviewers and the associate editor are appreciated.

References

- J. Baker, B. Bradley, and P. Stafford. *Seismic hazard and risk analysis*. Cambridge University Press, 2021.
- T. Bardainne, N. Dubos-Sallée, G. Sénéchal, P. Gaillot, and H. Perroud. Analysis of the induced seismicity of the lacq gas field (southwestern france) and model of deformation. *Geophysical Journal International*, 172(3):1151–1162, 2008.
- J. J. Bommer. Earthquake hazard and risk analysis for natural and induced seismicity: towards objective assessments in the face of uncertainty. *Bulletin of earthquake engineering*, 20(6):2825–3069, 2022.
- J. J. Bommer and J. Van Elk. Comment on “the maximum possible and maximum expected earthquake magnitude for production-induced earthquakes at the gas field in groningen, the netherlands” by gert zöller and matthias holschneider. 2017.
- J. J. Bommer, H. Crowley, and R. Pinho. A risk-mitigation approach to the management of induced seismicity. *Journal of Seismology*, 19(2):623–646, 2015.
- S. Bourne and S. Oates. Development of statistical geomechanical models for forecasting seismicity induced by gas production from the groningen field. *Netherlands Journal of geosciences*, 96(5):s175–s182, 2017a.
- S. Bourne and S. Oates. Extreme threshold failures within a heterogeneous elastic thin sheet and the spatial-temporal development of induced seismicity within the groningen gas field. *Journal of Geophysical Research: Solid Earth*, 122(12):10–299, 2017b.

- 771 S. Bourne and S. Oates. Stress-dependent magnitudes of induced earthquakes
772 in the groningen gas field. *Journal of Geophysical Research: Solid Earth*, 125
773 (11):e2020JB020013, 2020.
- 774 S. Bourne, S. Oates, J. van Elk, and D. Doornhof. A seismological model for
775 earthquakes induced by fluid extraction from a subsurface reservoir. *Journal*
776 *of Geophysical Research: Solid Earth*, 119(12):8991–9015, 2014.
- 777 M. Broccardo, A. Mignan, S. Wiemer, B. Stojadinovic, and D. Giardini. Hier-
778 archical bayesian modeling of fluid-induced seismicity. *Geophysical Research*
779 *Letters*, 44(22):11–357, 2017.
- 780 L. Buijze, P. A. van den Bogert, B. B. Wassing, B. Orlic, and J. ten Veen.
781 Fault reactivation mechanisms and dynamic rupture modelling of depletion-
782 induced seismic events in a rotliegend gas reservoir. *Netherlands Journal of*
783 *Geosciences*, 96(5):s131–s148, 2017.
- 784 T. Candela, S. Osinga, J.-P. Ampuero, B. Wassing, M. Pluymaekers, P. A.
785 Fokker, J.-D. van Wees, H. A. de Waal, and A. G. Muntendam-Bos. Depletion-
786 induced seismicity at the groningen gas field: Coulomb rate-and-state models
787 including differential compaction effect. *Journal of Geophysical Research:*
788 *Solid Earth*, 124(7):7081–7104, 2019.
- 789 K. W. Chang and P. Segall. Injection-induced seismicity on basement faults
790 including poroelastic stressing. *Journal of Geophysical Research: Solid Earth*,
791 121(4):2708–2726, 2016.
- 792 H. Clarke, J. P. Verdon, T. Kettlety, A. F. Baird, and J.-M. Kendall. Real-
793 time imaging, forecasting, and management of human-induced seismicity at
794 preston new road, lancashire, england. *Seismological Research Letters*, 90(5):
795 1902–1915, 2019.

- 796 K. Coppersmith, J. Ake, H. Bungum, T. Dahm, A. McGarr, I. Main,
797 I. Wong, and B. Youngs. Report on mmax expert workshop.
798 *Nederlandse Aardolie Maatschappij. Asse.*, 2016. available at
799 [https://nam-onderzoeksrapporten.data-app.nl/reports/download/](https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/cef44262-323a-4a34-afa8-24a5afa521d5)
800 [groningen/en/cef44262-323a-4a34-afa8-24a5afa521d5](https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/cef44262-323a-4a34-afa8-24a5afa521d5).
- 801 K. Coppersmith, J. Ake, H. Bungum, T. Dahm, A. McGarr, I. Main,
802 I. Wong, and B. Youngs. Report on the second workshop on
803 mmax for seismic hazard and risk analysis in the groningen gas
804 field. *Nederlandse Aardolie Maatschappij. Asse.*, 2022. available
805 at [https://nam-onderzoeksrapporten.data-app.nl/reports/download/](https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/77951661-552a-46bc-9f2e-f1580cd6abc3)
806 [groningen/en/77951661-552a-46bc-9f2e-f1580cd6abc3](https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/77951661-552a-46bc-9f2e-f1580cd6abc3).
- 807 D. Dempsey and J. Riffault. Response of induced seismicity to injection rate
808 reduction: Models of delay, decay, quiescence, recovery, and oklahoma. *Water*
809 *Resources Research*, 55(1):656–681, 2019.
- 810 D. Dempsey and J. Suckale. Physics-based forecasting of induced seismicity at
811 groningen gas field, the netherlands. *Geophysical Research Letters*, 44(15):
812 7773–7782, 2017.
- 813 D. Dempsey, J. Suckale, and Y. Huang. Collective properties of injection-
814 induced earthquake sequences: 2. spatiotemporal evolution and magnitude
815 frequency distributions. *Journal of Geophysical Research: Solid Earth*, 121
816 (5):3638–3665, 2016.
- 817 D. Dempsey, A. Kempa-Liehr, A. Ardid, A. Li, S. Orenia, J. Singh, A. Tyler,
818 and S. Cronin. Evaluation of short-term probabilistic eruption forecasting at
819 whakaari, new zealand. *Bulletin of Volcanology*, 84(10):1–26, 2022.
- 820 J. Dieterich. A constitutive law for rate of earthquake production and its ap-

821 plication to earthquake clustering. *Journal of Geophysical Research: Solid*
822 *Earth*, 99(B2):2601–2618, 1994.

823 B. Dost and D. Kraaijpoel. The august 16, 2012 earthquake near huizinge
824 (groningen). *KNMI Scientific Report. Royal Netherlands Meteorological In-*
825 *stitute (KNMI), Utrecht, The Netherlands*, 2013.

826 B. Dost, E. Ruigrok, and J. Spetzler. Development of seismicity and proba-
827 bilistic hazard assessment for the groningen gas field. *Netherlands Journal of*
828 *Geosciences*, 96(5):s235–s245, 2017.

829 J. D. Eshelby. The determination of the elastic field of an ellipsoidal inclusion,
830 and related problems. *Proceedings of the royal society of London. Series A.*
831 *Mathematical and physical sciences*, 241(1226):376–396, 1957.

832 D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: the mcmc
833 hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):
834 306, 2013.

835 M. Galis, J. P. Ampuero, P. M. Mai, and F. Cappa. Induced seismicity provides
836 insight into why earthquake ruptures stop. *Science advances*, 3(12):eaap7528,
837 2017.

838 T. Goebel, M. Weingarten, X. Chen, J. Haffener, and E. Brodsky. The 2016
839 *mw* 5.1 fairview, oklahoma earthquakes: Evidence for long-range poroelastic
840 triggering at >40 km from fluid disposal wells. *Earth and Planetary Science*
841 *Letters*, 472:50–61, 2017.

842 M. O. Häring, U. Schanz, F. Ladner, and B. C. Dyer. Characterisation of the
843 basel 1 enhanced geothermal system. *Geothermics*, 37(5):469–495, 2008.

844 E. R. Heimisson, J. D. Smith, J.-P. Avouac, and S. J. Bourne. Coulomb thresh-
845 old rate-and-state model for fault reactivation: application to induced seis-

846 micity at groningen. *Geophysical Journal International*, 228(3):2061–2072,
847 2022.

848 P. A. Hsieh and J. D. Bredehoeft. A reservoir analysis of the denver earthquakes:
849 A case of induced seismicity. *Journal of Geophysical Research: Solid Earth*,
850 86(B2):903–920, 1981.

851 J. C. Jaeger, N. G. Cook, and R. Zimmerman. *Fundamentals of rock mechanics*.
852 John Wiley & Sons, 2009.

853 K. M. Keranen and M. Weingarten. Induced seismicity. *Annual Review of Earth*
854 *and Planetary Sciences*, 46(1):149–174, 2018.

855 K. M. Keranen, M. Weingarten, G. A. Abers, B. A. Bekins, and S. Ge. Sharp
856 increase in central oklahoma seismicity since 2008 induced by massive wastew-
857 ater injection. *Science*, 345(6195):448–451, 2014.

858 E. Király-Proag, J. D. Zechar, V. Gischig, S. Wiemer, D. Karvounis, and
859 J. Doetsch. Validating induced seismicity forecast models—induced seismicity
860 test bench. *Journal of Geophysical Research: Solid Earth*, 121(8):6009–6029,
861 2016.

862 Z. Li, D. Eaton, and J. Davidsen. Short-term forecasting of mmax during
863 hydraulic fracturing. *Scientific Reports*, 12(1):1–12, 2022.

864 J. Limbeck, K. Bisdom, F. Lanz, T. Park, E. Barbaro, S. Bourne, F. Kiraly,
865 S. Bierman, C. Harris, K. Nevenzeel, et al. Using machine learning for model
866 benchmarking and forecasting of depletion-induced seismicity in the groningen
867 gas field. *Computational Geosciences*, 25(1):529–551, 2021.

868 I. Main. Apparent breaks in scaling in the earthquake cumulative frequency-
869 magnitude distribution: fact or artifact? *Bulletin of the Seismological Society*
870 *of America*, 90(1):86–97, 2000.

- 871 E. Majer, R. Baria, M. Stark, S. Oates, J. Bommer, B. Smith, and H. Asanuma.
872 Induced seismicity associated with enhanced geothermal systems. *Geother-*
873 *mics*, 36(3):185–222, 2007.
- 874 P. Maldonado, A. Peng, D. Ouyang, J. Suckale, and D. E. Ho. Science trans-
875 lation during the covid-19 pandemic: An academic-public health partnership
876 to assess capacity limits in california. *American Journal of Public Health*, 112
877 (2):308–315, 2022.
- 878 W. Marzocchi, J. D. Zechar, and T. H. Jordan. Bayesian forecast evaluation
879 and ensemble earthquake forecasting. *Bulletin of the Seismological Society of*
880 *America*, 102(6):2574–2584, 2012.
- 881 A. Muntendam-Bos and N. Grobbe. Data-driven spatiotemporal assessment
882 of the event-size distribution of the groningen extraction-induced seismicity
883 catalogue. *Scientific Reports*, 12(1):1–11, 2022.
- 884 A. Muntendam-Bos, G. Hoedeman, K. Polychronopoulou, D. Draganov,
885 C. Weemstra, W. van der Zee, R. R. Bakker, and H. Roest. An overview
886 of induced seismicity in the netherlands. *Netherlands Journal of Geosciences*,
887 101, 2022.
- 888 M. D. Petersen, C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. L. Llenos,
889 W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, and K. S.
890 Rukstales. Seismic-hazard forecast for 2016 including induced and natural
891 earthquakes in the central and eastern united states. *Seismological Research*
892 *Letters*, 87(6):1327–1341, 2016.
- 893 M. D. Petersen, C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. M. Shumway,
894 D. E. McNamara, R. A. Williams, A. L. Llenos, W. L. Ellsworth, A. J.
895 Michael, et al. 2017 one-year seismic-hazard forecast for the central and

896 eastern united states from induced and natural earthquakes. *Seismological*
897 *Research Letters*, 88(3):772–783, 2017.

898 M. D. Petersen, C. S. Mueller, M. P. Moschetti, S. M. Hoover, K. S. Rukstales,
899 D. E. McNamara, R. A. Williams, A. M. Shumway, P. M. Powers, P. S. Earle,
900 et al. 2018 one-year seismic hazard forecast for the central and eastern united
901 states from induced and natural earthquakes. *Seismological Research Letters*,
902 89(3):1049–1061, 2018.

903 T. Postma and J.-D. Jansen. The small effect of poroelastic pressure transients
904 on triggering of production-induced earthquakes in the groningen natural gas
905 field. *Journal of Geophysical Research: Solid Earth*, 123(1):401–417, 2018.

906 G. Richter, S. Hainzl, T. Dahm, and G. Zöller. Stress-based, statistical mod-
907 eling of the induced seismicity at the groningen gas field, the netherlands.
908 *Environmental Earth Sciences*, 79(11):1–15, 2020.

909 D. Schorlemmer, M. Gerstenberger, S. Wiemer, D. Jackson, and D. Rhoades.
910 Earthquake likelihood model testing. *Seismological Research Letters*, 78(1):
911 17–29, 2007.

912 R. Schultz, R. J. Skoumal, M. R. Brudzinski, D. Eaton, B. Baptie, and
913 W. Ellsworth. Hydraulic fracturing-induced seismicity. *Reviews of Geo-*
914 *physics*, 58(3):e2019RG000695, 2020.

915 P. Segall. Earthquakes triggered by fluid extraction. *Geology*, 17(10):942–946,
916 1989.

917 P. Segall. Induced stresses due to fluid extraction from axisymmetric reservoirs.
918 *pure and applied geophysics*, 139(3):535–560, 1992.

919 P. Segall and S. D. Fitzgerald. A note on induced stress changes in hydrocarbon
920 and geothermal reservoirs. *Tectonophysics*, 289(1-3):117–128, 1998.

- 921 P. Segall and S. Lu. Injection-induced seismicity: Poroelastic and earthquake
 922 nucleation effects. *Journal of Geophysical Research: Solid Earth*, 120(7):
 923 5082–5103, 2015.
- 924 S. A. Shapiro and C. Dinske. Fluid-induced seismicity: Pressure diffusion and
 925 hydraulic fracturing. *Geophysical Prospecting*, 57(2):301–310, 2009.
- 926 S. A. Shapiro, E. Rothert, V. Rath, and J. Rindschwentner. Characterization
 927 of fluid transport properties of reservoirs using induced microseismicity. *Geo-*
 928 *physics*, 67(1):212–220, 2002.
- 929 S. A. Shapiro, O. S. Krüger, C. Dinske, and C. Langenbruch. Magnitudes of
 930 induced earthquakes and geometric scales of fluid-stimulated rock volumes.
 931 *Geophysics*, 76(6):WC55–WC63, 2011.
- 932 S. A. Shapiro, C. Dinske, and O. Krueger. Maximum magnitudes of earth-
 933 quakes induced by fluid injections and productions: Controlling parameters,
 934 estimations, and case study examples. In *SEG Technical Program Expanded*
 935 *Abstracts 2017*, pages 5356–5361. Society of Exploration Geophysicists, 2017.
- 936 J. D. Smith, R. S. White, J.-P. Avouac, and S. Bourne. Probabilistic earthquake
 937 locations of induced seismicity in the groningen region, the netherlands. *Geo-*
 938 *physical Journal International*, 222(1):507–516, 2020.
- 939 J. D. Smith, E. R. Heimisson, S. J. Bourne, and J.-P. Avouac. Stress-based fore-
 940 casting of induced seismicity with instantaneous earthquake failure functions:
 941 applications to the groningen gas reservoir. *Earth and Planetary Science*
 942 *Letters*, 594:117697, 2022.
- 943 J. Spetzler and B. Dost. Hypocentre estimation of induced earthquakes in
 944 groningen. *Geophysical Journal International*, 209(1):453–465, 2017.

- 945 R. Storn and K. Price. Differential evolution—a simple and efficient heuristic for
946 global optimization over continuous spaces. *Journal of global optimization*,
947 11(4):341–359, 1997.
- 948 A. Strader, M. Schneider, and D. Schorlemmer. Prospective and retrospective
949 evaluation of five-year earthquake forecast models for california. *Geophysical*
950 *Journal International*, 211(1):239–251, 2017.
- 951 M. Taroni, J. Zechar, and W. Marzocchi. Assessing annual global m 6+ seis-
952 micity forecasts. *Geophysical Journal International*, 196(1):422–431, 2014.
- 953 T. Utsu. Representation and analysis of the earthquake size distribution: a
954 historical review and some new approaches. In *Seismicity patterns, their*
955 *statistical significance and physical meaning*, pages 509–535. Springer, 1999.
- 956 K. van Thienen-Visser and J. Breunese. Induced seismicity of the groningen gas
957 field: History and recent developments. *The Leading Edge*, 34(6):664–671,
958 2015.
- 959 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cour-
960 napeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al. Scipy 1.0:
961 fundamental algorithms for scientific computing in python. *Nature methods*,
962 17(3):261–272, 2020.
- 963 C. Willacy, E. van Dedem, S. Minisini, J. Li, J.-W. Blokland, I. Das, and
964 A. Droujinine. Full-waveform event location and moment tensor inversion
965 for induced seismicity. *Geophysics*, 84(2):KS39–KS57, 2019.
- 966 G. Zhai, M. Shirzaei, M. Manga, and X. Chen. Pore-pressure diffusion, enhanced
967 by poroelastic stresses, controls induced seismicity in oklahoma. *Proceedings*
968 *of the National Academy of Sciences*, 116(33):16228–16233, 2019.
- 969 M. D. Zoback. *Reservoir geomechanics*. Cambridge university press, 2010.

970 G. Zöller and M. Holschneider. The maximum possible and the maximum ex-
971 pected earthquake magnitude for production-induced earthquakes at the gas
972 field in groningen, the netherlands. *Bulletin of the Seismological Society of*
973 *America*, 106(6):2917–2921, 2016.

⁹⁷⁴ **Author addresses**

⁹⁷⁵ David Dempsey: 39 Victoria Avenue, Whanganui, 4500, New Zealand

⁹⁷⁶ Jenny Suckale: Department of Geophysics, 397 Panama Mall, Mitchell Build-

⁹⁷⁷ ing, 3rd Floor, Stanford University, Stanford, CA 94305-2215

Table 1: Gas extraction and seismicity over the evaluation period.

Period	Extraction (bcm)	Seismicity N($M \geq 1.0$)	$M \geq 2.5$ events
2017, 1-Feb to 31-Dec	21.2	53	M 2.6 (27/05/17)
2018, 1-Jan to 31-Dec	18.8	28	M 3.4 (08/01/18) M 2.8 (13/04/18)
2019, 1-Jan to 31-Dec	15.6	35	M 3.4 (22/05/19) M 2.5 (09/06/19)
2020, 1-Jan to 31-Dec	7.9	34	M 2.5 (02/05/20) M 2.7 (14/07/20)
2021, 1-Jan to 31-Dec	6.5	29	M 2.5 (04/10/21) M 3.2 (16/11/21)
2022, 1-Jan to 31-Jun	2.5	4	M 2.7 (01/04/22)

Table 2: Summary of forecast model parameters. DS17=Dempsey and Suckale [2017], DS22=this study, PRO=prospective, PPRO=pseudo-prospective.

Parameter	DS17 PRO	DS22 PPRO	DS22 PRO
h	[0.16,0.26]	0.19	0.19
H	[0.2,0.45]	0.36	0.38
θ	[25°, 34°]	46.0°	46.6°
f	[0.50,0.72]	0.50	0.50
A	[0.90,1.0]	1.0	1.0

Table 3: DS22 forecast and select exceedance probabilities for the period 2021 to 2030. $N_{2.5}$ values are rounded to the nearest integer.

p	$N_{2.5}$		$M_{\max}$	
	tapered	unbounded	tapered	unbounded
0.5	5	5	3.2	3.4
0.1	8	8	3.6	4.3
0.01	11	11	3.8	5.4
0.001	13	14	3.9	6.5

Figure 1: Overview of Groningen field and gas production activities. (a) Map of Groningen with field outline (black), straight line approximations of faults (red), and induced earthquakes. (b) Annual gas extraction (black line) in billions of cubic meters (bcm) and annual induced earthquakes (open circles) with $M \geq 1.0$. (c) Average reservoir pressure (solid line) and compaction (dotted) over time.

979 **Figure captions**

Figure 2: Model summary. (a) Fault of length, L , represented as a (red dashed) 1D line element with strike, φ , and along-strike coordinate, s . Contours of pressure decrease (MPa) at 2021 are shown in blue. Blue arrows in top corner show one possible orientation of the two horizontal principle stresses rotated an angle, θ , from north. (b) 3D view of fault showing dip angle, ϕ , and resolved normal and shear stress components, σ_n and τ , that vary along-strike and with time. In DS17, fault vertical thickness was a constant, whereas in DS22 thickness varies along strike. (c) Poroelastic inclusion model adapted from Segall and Fitzgerald [1998] showing how (i) for a non-elastically coupled reservoir, depressurization induces a contraction, (ii) elastic coupling is restored by stretching the contracted reservoir back to its initial volume, and (iii) subsidence compensates the vertical stress change leaving only an extensional horizontal stress in the reservoir. (d) Example rupture length simulation for calculation of event magnitude in the DS17 model, adapted from Dempsey et al. [2016]. Crack-like propagation of slip along the 1D fault, arresting in sections of negative stress drop.

Figure 3: Evaluation of DS17 forecast. (a) DS17 seismicity rate model ensemble showing best fits (black lines) through data (open circles) and three future scenarios: 21 (green), 27 (orange) and 33 bcm/yr (purple). Modelled and observed MFD is shown in the inset. Horizontal dashed lines show approximate scenario median rates between 2017 and 2020. (b) Exceedance probability for number of $M \geq 2.5$ events in the period 1 February 2017 to 1 January 2025 (solid lines) and volume-scaled scenario for 21 bcm/yr (dashed line) for the evaluation period 1 February 2017 to 1 July 2022. Actual number of $M \geq 2.5$ events is shown by the vertical dotted line. (c) Exceedance probability for the largest event in the forecast period for the three scenarios. Actual largest event in the evaluation period shown by the vertical dotted line.

Figure 4: Pseudo-prospective performance of the updated model over the evaluation period. (a) Measured seismicity rate (circles) compared to best fit model, $\lambda(t)$ (solid line) with 90% aleatoric variability (grey envelope). The data are divided into train and test sets at the year 2017 (closed and open circles, respectively), with the former used to calibrate the model shown in Fig. 5. The trained model is extrapolated for the period February 2017 to July 2022 and compared to the withheld test data (open circles). The greyed marker is the extrapolated event count for 2022 based on events in the first six months of the year. (b) Exceedance curve for likelihoods scoring $\lambda(t)$ against synthetic catalogs derived from $\lambda(t)$. Likelihood of $\lambda(t)$ scored against the data is shown by the vertical dotted line (LLK=460, $p=0.04$). The (c) Exceedance probability curve for the number of $M \geq 2.5$ events, and (d) the largest magnitude event, for tapered (solid) and unbounded (dashed) Gutenberg-Richter distributions. Actual outcomes during the evaluation period are shown as vertical dotted lines ($N_{2.5}=10$ with $p_{tp}=0.09$ and $p_{ub}=0.11$; $M_{\max}=3.4$ with $p_{tp}=0.34$ and $p_{ub}=0.60$; tp =tapered and ub =unbounded).

Figure 5: Model fit to spatial seismicity distribution at Groningen. (a) Model match (solid line) to event density (circles) in 2.1 km bins along the x direction. The grey shaded envelope shows an interval of 90% aleatoric variability on the modelled density. (b) Heat map showing area density of cumulative model seismicity as at February 2017 in 2.1×2.5 km bins. Also shown are modelled faults (straight lines) and $M \geq 1.0$ events (open circles). (c) Model match to event density in 2.5 km bins along the y direction.

Figure 6: DS22 prospective forecast over the period 2021 to 2030. (a) Measured seismicity rate (circles) compared to best fit model (solid line) with 90% aleatoric variability (grey envelope). The data are divided into train and test sets at the year 2021 (closed and open circles, respectively), with the former used to calibrate the model. The trained model is extrapolated for the period 2021 to 2030. The greyed marker is the extrapolated event count for 2022 based on events in the first six months of the year. (b) Exceedance probability curve for the number of $M \geq 2.5$ events, and (c) the largest magnitude event, for tapered (solid) and unbounded (dashed) Gutenberg-Richter distributions.

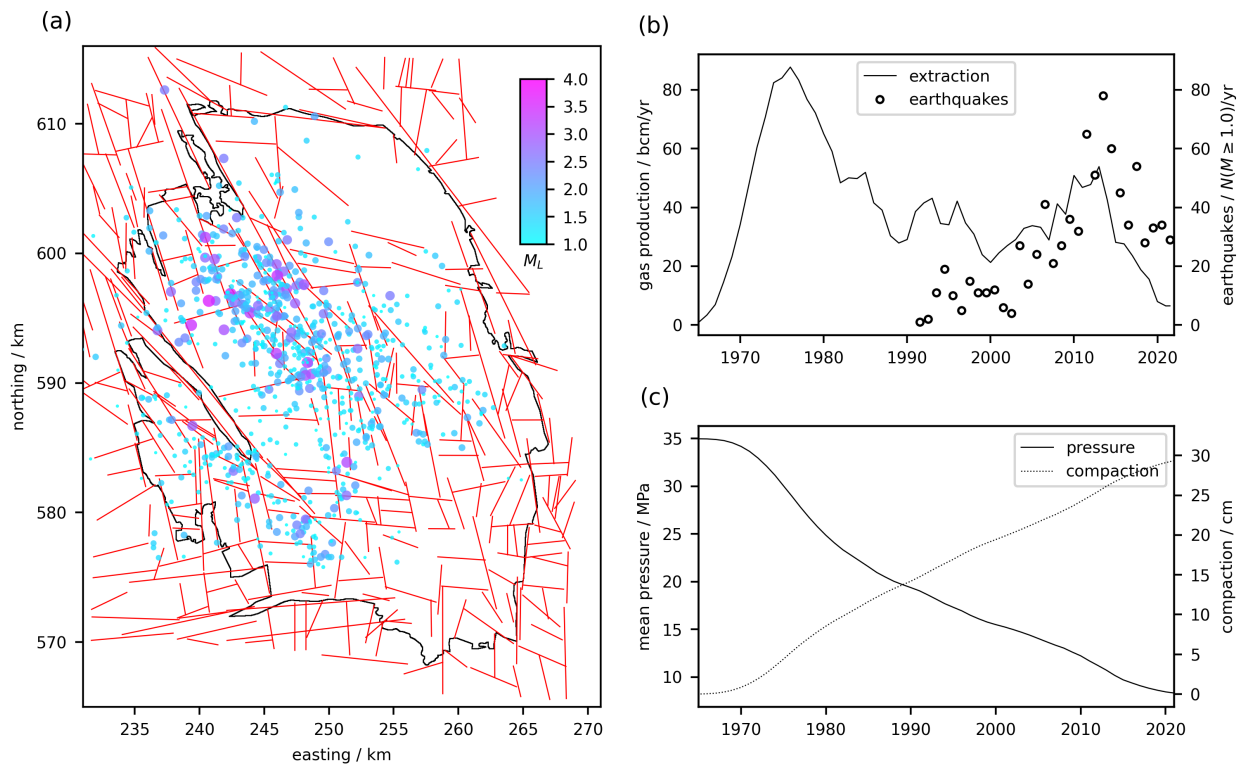


Figure 1: Overview of Groningen field and gas production activities. (a) Map of Groningen with field outline (black), straight line approximations of faults (red), and induced earthquakes. (b) Annual gas extraction (black line) in billions of cubic meters (bcm) and annual induced earthquakes (open circles) with $M \geq 1.0$. (c) Average reservoir pressure (solid line) and compaction (dotted) over time.

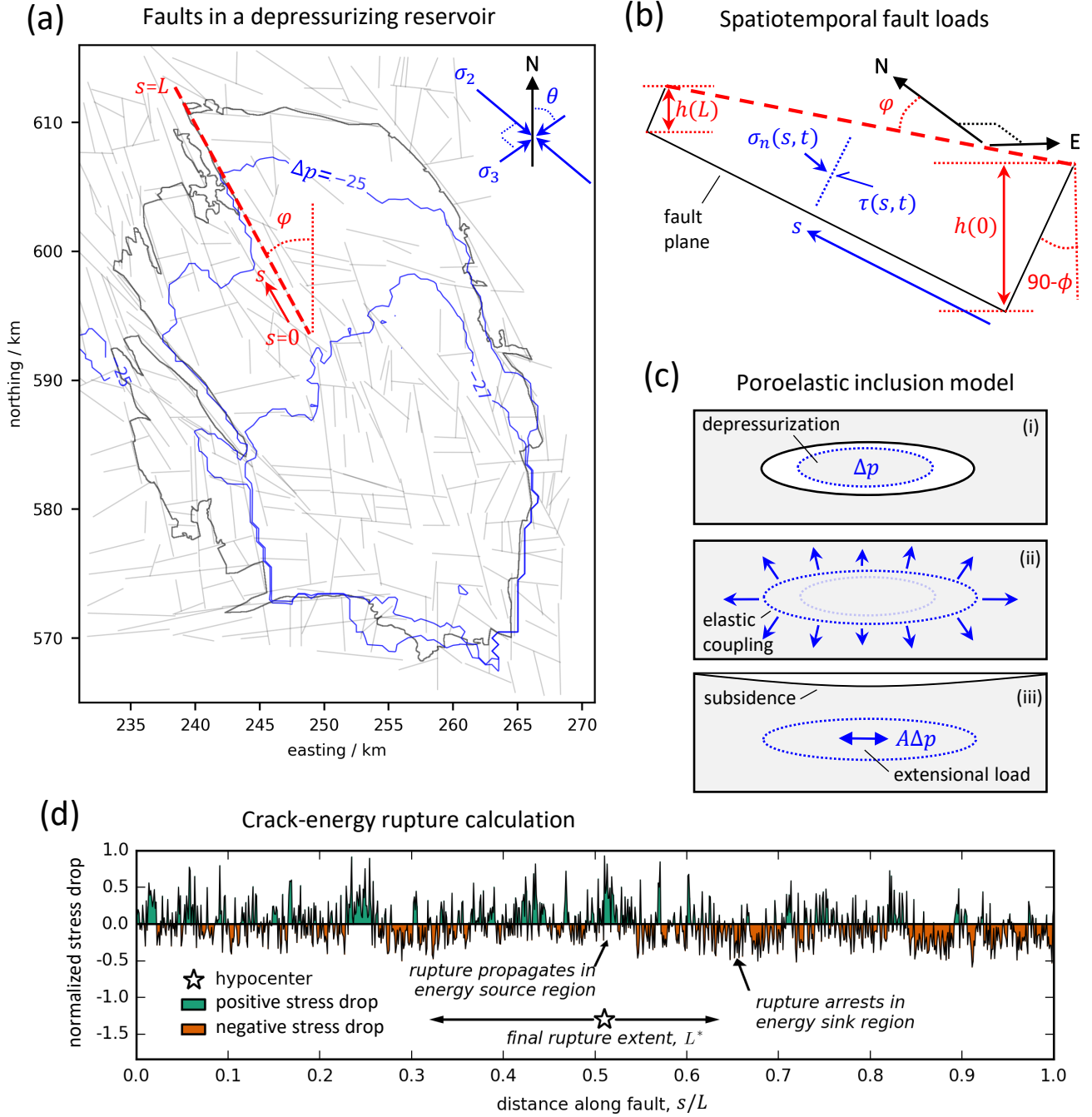


Figure 2: Model summary. (a) Fault of length, L , represented as a (red dashed) 1D line element with strike, ϕ , and along-strike coordinate, s . Contours of pressure decrease (MPa) at 2021 are shown in blue. Blue arrows in top corner show one possible orientation of the two horizontal principle stresses rotated an angle, θ , from north. (b) 3D view of fault showing dip angle, ϕ , and resolved normal and shear stress components, σ_n and τ , that vary along-strike and with time. In DS17, fault vertical thickness was a constant, whereas in DS22 thickness varies along strike. (c) Poroelastic inclusion model adapted from Segall and Fitzgerald [1998] showing how (i) for a non-elastically coupled reservoir, depressurization induces a contraction, (ii) elastic coupling is restored by stretching the contracted reservoir back to its initial volume, and (iii) subsidence compensates the vertical stress change leaving only an extensional horizontal stress in the reservoir. (d) Example rupture length simulation for calculation of event magnitude in the DS17 model, adapted from Dempsey et al. [2016]. Crack-like propagation of slip along the 1D fault, arresting in sections of negative stress drop.

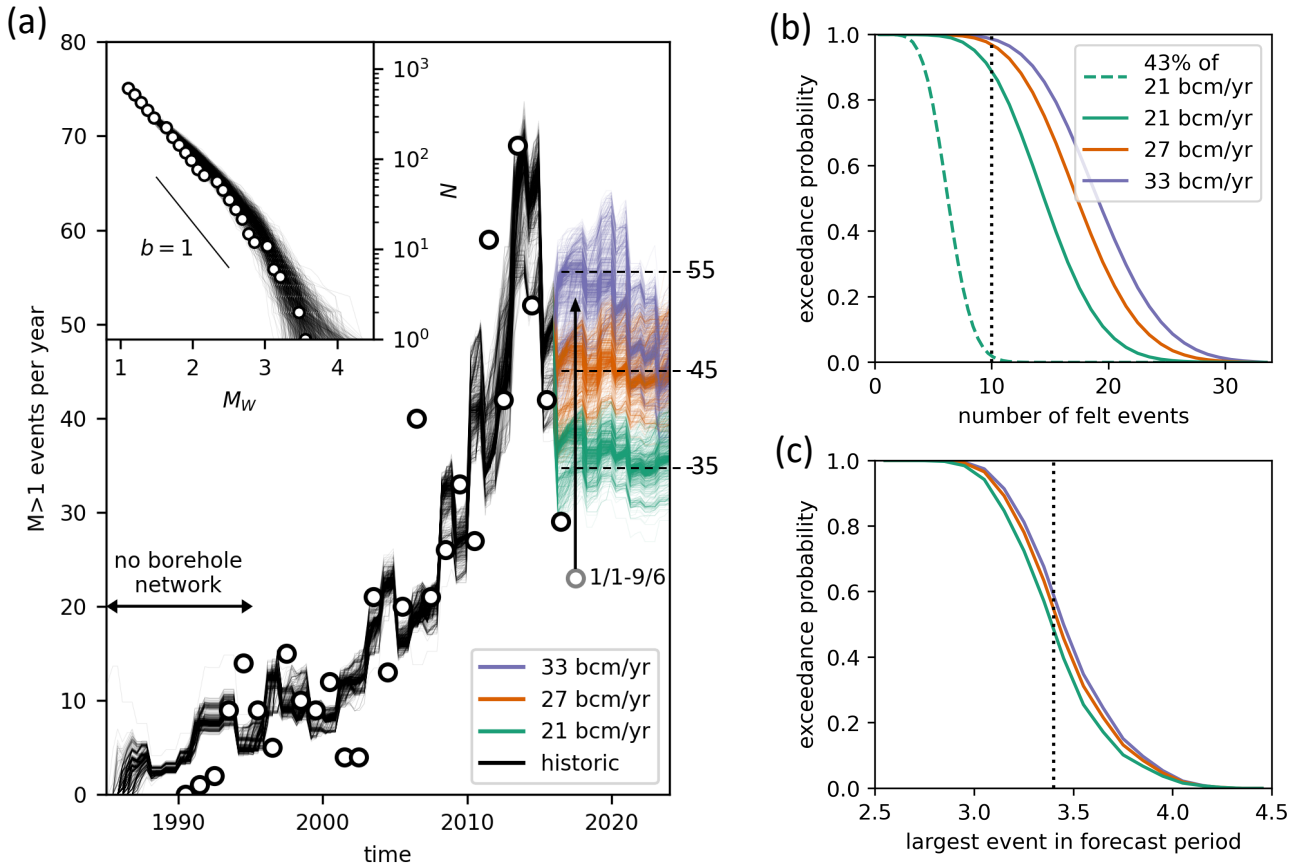


Figure 3: Evaluation of DS17 forecast. (a) DS17 seismicity rate model ensemble showing best fits (black lines) through data (open circles) and three future scenarios: 21 (green), 27 (orange) and 33 bcm/yr (purple). Modelled and observed MFD is shown in the inset. Horizontal dashed lines show approximate scenario median rates between 2017 and 2020. (b) Exceedance probability for number of $M \geq 2.5$ events in the period 1 February 2017 to 1 January 2025 (solid lines) and volume-scaled scenario for 21 bcm/yr (dashed line) for the evaluation period 1 February 2017 to 1 July 2022. Actual number of $M \geq 2.5$ events is shown by the vertical dotted line. (c) Exceedance probability for the largest event in the forecast period for the three scenarios. Actual largest event in the evaluation period shown by the vertical dotted line.

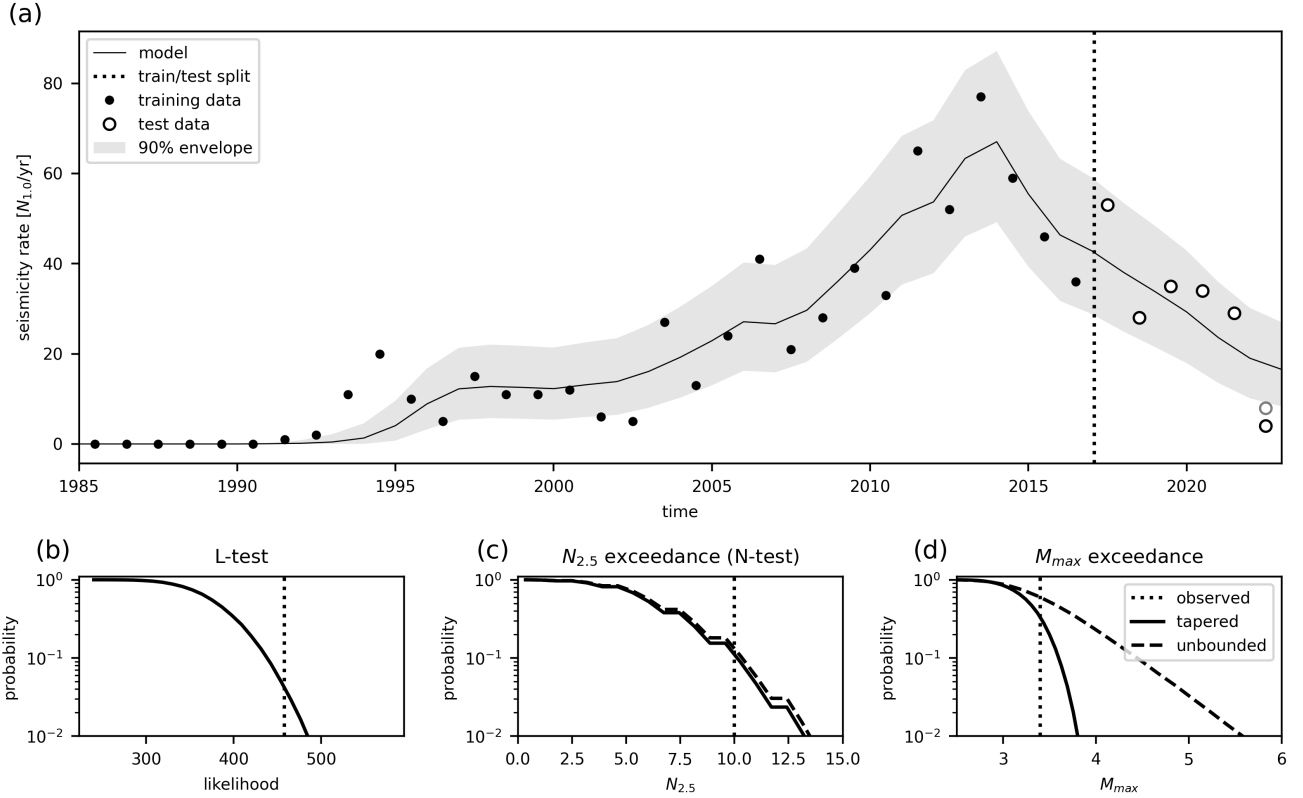


Figure 4: Pseudo-prospective performance of the updated model over the evaluation period. (a) Measured seismicity rate (circles) compared to best fit model, $\lambda(t)$ (solid line) with 90% aleatoric variability (grey envelope). The data are divided into train and test sets at the year 2017 (closed and open circles, respectively), with the former used to calibrate the model shown in Fig. 5. The trained model is extrapolated for the period February 2017 to July 2022 and compared to the withheld test data (open circles). The greyed marker is the extrapolated event count for 2022 based on events in the first six months of the year. (b) Exceedance curve for likelihoods scoring $\lambda(t)$ against synthetic catalogs derived from $\lambda(t)$. Likelihood of $\lambda(t)$ scored against the data is shown by the vertical dotted line (LLK=460, $p=0.04$). The (c) Exceedance probability curve for the number of $M \geq 2.5$ events, and (d) the largest magnitude event, for tapered (solid) and unbounded (dashed) Gutenberg-Richter distributions. Actual outcomes during the evaluation period are shown as vertical dotted lines ($N_{2.5}=10$ with $\text{ptp}=0.09$ and $\text{pub}=0.11$; $M_{\max}=3.4$ with $\text{ptp}=0.34$ and $\text{pub}=0.60$; $\text{tp}=\text{tapered}$ and $\text{ub}=\text{unbounded}$).

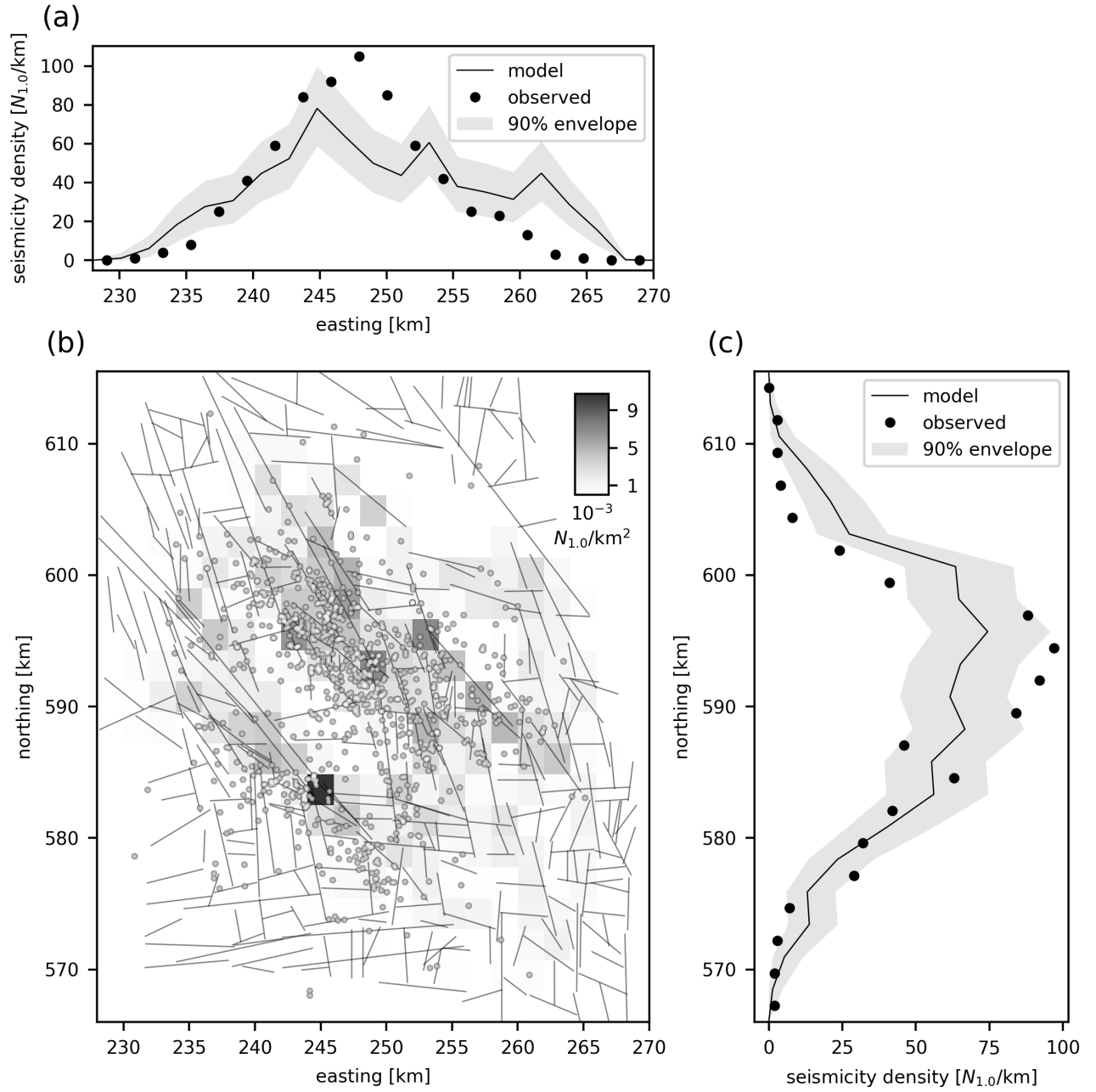


Figure 5: Model fit to spatial seismicity distribution at Groningen. (a) Model match (solid line) to event density (circles) in 2.1 km bins along the x direction. The grey shaded envelope shows an interval of 90% aleatoric variability on the modelled density. (b) Heat map showing area density of cumulative model seismicity as at February 2017 in 2.1×2.5 km bins. Also shown are modelled faults (straight lines) and $M \geq 1.0$ events (open circles). (c) Model match to event density in 2.5 km bins along the y direction.

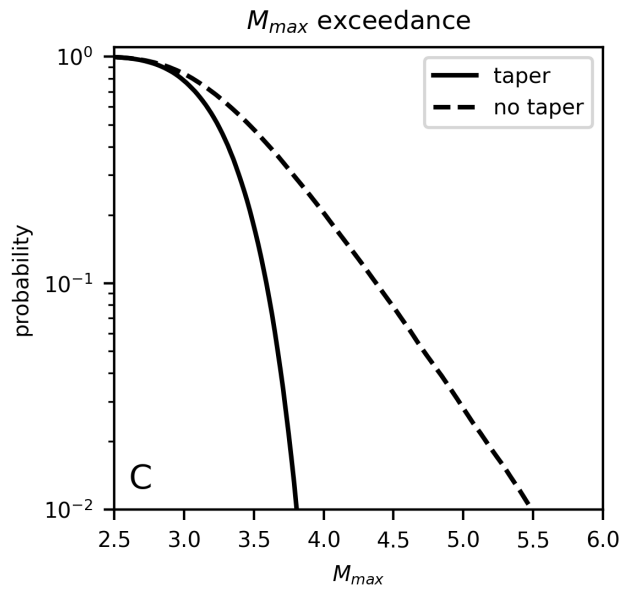
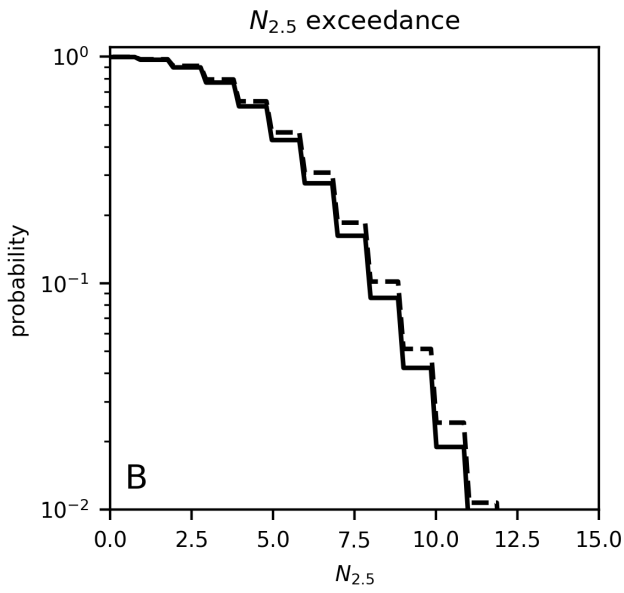
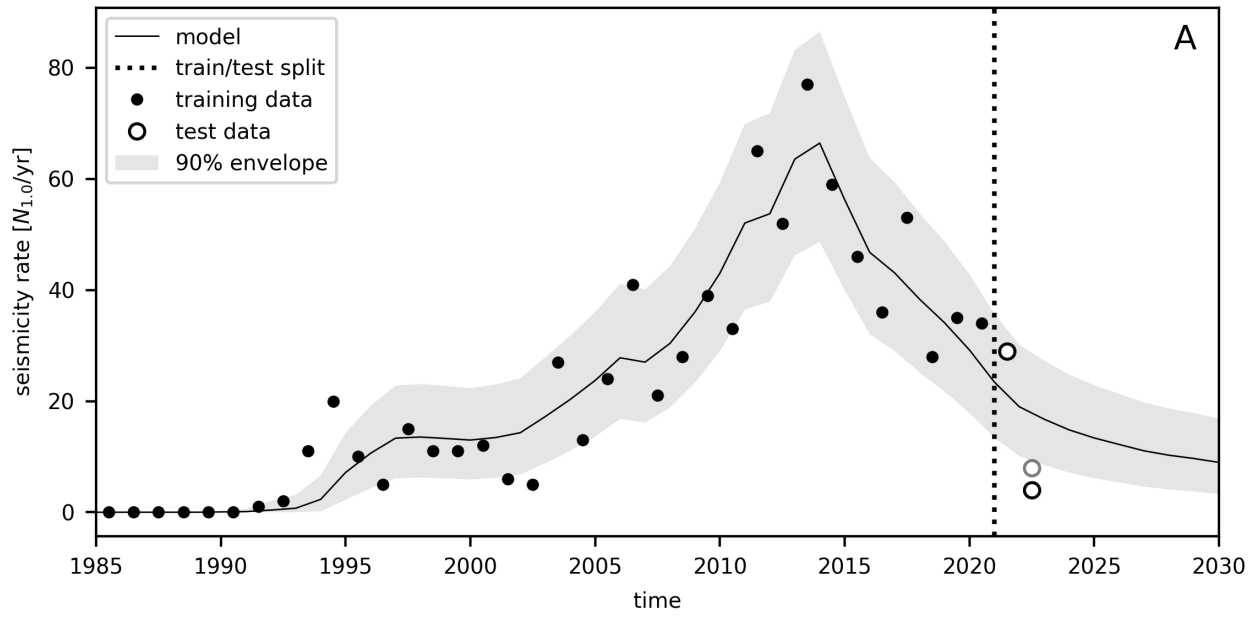


Figure 6: DS22 prospective forecast over the period 2021 to 2030. (a) Measured seismicity rate (circles) compared to best fit model (solid line) with 90% aleatoric variability (grey envelope). The data are divided into train and test sets at the year 2021 (closed and open circles, respectively), with the former used to calibrate the model. The trained model is extrapolated for the period 2021 to 2030. The greyed marker is the extrapolated event count for 2022 based on events in the first six months of the year. (b) Exceedance probability curve for the number of $M \geq 2.5$ events, and (c) the largest magnitude event, for tapered (solid) and unbounded (dashed) Gutenberg-Richter distributions.

Supporting Information for “Physics-based forecasting of induced seismicity at Groningen gas field, the Netherlands: post hoc evaluation and forecast update”

D. E. Dempsey¹, J. Suckale²

¹University of Canterbury, Christchurch, New Zealand

²Stanford University, Stanford, California, USA

Detailed description of *Dempsey and Suckale (2017)* seismological model

The Dempsey & Suckale (2017) forecast model, hereafter referred to as DS17, was used to estimate the possible maximum magnitude at Groningen, $M_{\max}$. This quantity was defined probabilistically as the largest earthquake resulting from hypothetical extraction scenarios at Groningen between 2017 and 2025. The model contained parameterised descriptions of (i) the initial tectonic stress field at Groningen, (ii) a poroelastic model that links gas production to fault loading, and (iii) simulations of earthquake source triggering and propagation.

Geomechanical model

The initial tectonic stress within the Groningen reservoir is poorly constrained. In the DS17 model, the tectonic regime was assumed to be extensional and was thus parameterized by a vertical maximum principle stress, σ_1 , and two horizontal principle stresses, σ_2 and σ_3 . The magnitude of σ_1 was calculated through depth integration of a constant overburden density (2500 kg/s) to the average reservoir depth (3 km).

The magnitude of the minimum horizontal stress, σ_3 , is unknown. For an extensional system, it must be less than σ_1 . However, it is presumed to not be smaller than a theoretical lower limit, σ_{crit} , imposed by frictional equilibrium for well-oriented shear cracks (Jaeger et al., 2009), given:

$$\sigma_{\text{crit}} = (\sigma_1 - p_0) \left[f + \sqrt{1 + f^2} \right]^{-2} + p_0, \quad (1)$$

where p_0 is the initial pressure in the reservoir (~ 35 MPa in 1958.)

Given that σ_3 is in the range σ_{crit} to σ_1 , its value can be parameterized by $h = (\sigma_3 - \sigma_{\text{crit}})/(\sigma_1 - \sigma_{\text{crit}})$. Thus h varies between 0 and 1 with $h = 0$ denoting a critical stress state, and $h = 1$ an isotropic stress state.

The magnitude of the maximum horizontal stress, σ_2 , is assumed to lie between σ_1 and σ_3 , and is parameterized by $H = (\sigma_2 - \sigma_3)/(\sigma_1 - \sigma_3)$.

Finally, the angle between σ_3 and north denotes the rotation of horizontal stresses (main text, Fig. 2a). Mathematically, it is applied as a rotation around a vertical axis

$$S = R(\theta)^T S_D R(\theta), \quad (2)$$

where S_D is a diagonal matrix containing the principle stresses, S is the stress tensor in x - y coordinates, and R is the rotation matrix.

Rotation of the stresses determines which faults are most likely to be well-oriented in the stress field for early reactivation (main text, Fig. 2a). Thus, the regional stress field was determined by four parameters: $[f, h, H, \theta]$. It was assumed to be homogeneous throughout the field.

Poroelastic fault loading

In the DS17 model, the depth-averaged spatio-temporal pressure distribution (main text, Fig. 2a) in the reservoir, $\Delta p(\mathbf{x}, t)$, was derived from a two-component (water and gas) reservoir simulation. The simulations accounted for historic gas extraction and were history matched to pressure and subsidence data. Three future depressurization scenarios were calculated for gas extraction scenarios averaging 21, 27 and 33 bcm/yr between February 2017 and 2025.

A poroelastic model was used to link depressurization to stress changes that triggered earthquakes. The model invokes the embedded ellipse conceptualization of Segall & Fitzgerald (1998), wherein depressurization induces a contraction of an elastically decoupled reservoir. The reservoir later recovers its original shape and true elastic coupling through the application of an extensional strain within the reservoir and subsidence above (main text, Fig. 2c).

This poroelastic model was applied locally, using the pressure change to calculate horizontal stress change by $\Delta\sigma_h(\mathbf{x}, t) = A\Delta p(\mathbf{x}, t)$ where $h=2$ or 3 . The parameter, A , is a poroelastic stress path coefficient (Segall & Fitzgerald, 1998; Zoback, 2010). As a result, negative pressure changes (depressurization) cause the horizontal stresses to decrease, pushing the system closer to a critical extensional stress state. The term $\Delta\sigma_h(\mathbf{x}, t)$ was applied as a modification to the two principle horizontal stresses in a homogeneous background tectonic stress

$$S_D(\mathbf{x}, t) = \underbrace{\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}}_{\text{initial tectonic stress}} + A \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & \Delta p(\mathbf{x}, t) & 0 \\ 0 & 0 & \Delta p(\mathbf{x}, t) \end{bmatrix}}_{\text{poroelastic modification}}. \quad (3)$$

Thus, the regional stress field evolves in space and time as a function of depressurization.

Faults embedded in the stress field experience a time and space varying load (main text, Fig. 2b). In DS17, faults were treated as a straight line elements with the i^{th} fault having normal vector, $\hat{\mathbf{n}}_i$, that depends on fault strike, φ_i , and dip, ϕ_i , as well as a length, L_i . The stress field imposes a traction force on the fault, $\mathbf{t}$, which in turn can be subdividing into normal, σ_n , and shear stress, τ , components.

$$\mathbf{t}(s, t) = S(\mathbf{x}(s), t)\hat{\mathbf{n}}_i, \quad \sigma_n(s, t) = \hat{\mathbf{n}}_i \cdot \mathbf{t}(s, t), \quad \tau(s, t) = |\mathbf{t}(s, t) - \sigma_n(s, t)\hat{\mathbf{n}}_i|, \quad (4)$$

where s is a coordinate along the strike of the fault.

DS17 earthquake physics simulator

The DS17 model was built using the induced earthquake physics simulator described in Dempsey & Suckale (2016). In that simulator, a fault was abstracted as a 1D line segment. The initial shear stress was heterogeneous along the fault ($\tau_0(s)$) with a mean value that depended on fault orientation in the regional stress field. The spatial correlation properties of $\tau_0(s)$ were assigned values consistent with seismological observations. Under tectonic loading, the simulator would output earthquakes with a Gutenberg-Richter MFD (Dempsey et al., 2016) and a Poissonian distribution of event times.

Model earthquakes were triggered by applied stress and strength perturbations. The spatio-temporal change of shear and normal stress on the fault ($\Delta\tau(s, t)$, $\Delta\sigma_n(s, t)$) were derived from the previously described poroelastic depressurization model. Mohr-Coulomb strength on the fault was the product of effective normal stress, $\sigma_n(s, t) - p(s, t)$, and the same friction coefficient, f , used to parameterize the stress state. An earthquake was triggered on the fault at the first location where stress exceeded strength across some minimal rupture patch size for nucleation. Once the earthquake was triggered, two rupture fronts propagated in either direction along the fault, away from the nucleation site. Continued propagation was dependent on strength drop and stress transfer, approximated by a simplified 1D crack energy model (main text, Fig. 2d). Once the rupture size was known, an earthquake magnitude was calculated using the total rupture length and average stress drop.

An ensemble of possible earthquakes was simulated for the 325 longest faults in Groningen (main text, Fig. 1a), with a minimum fault length of 2 km. Interactions between earthquakes on separate faults was not considered. The total simulated seismicity was used to define a time-varying seismicity rate for the field, $\lambda(t)$.

Empirical MFDs at Groningen

The unbounded GR distribution assumes that earthquake magnitudes, M , follow a power-law distribution with slope parameter, b . The exceedance probability that an event, M_i , will have a magnitude larger than M is given:

$$P(M_i \geq M) = 10^{-b(M-M_m)}, \quad (5)$$

where M_m is the minimum magnitude considered.

Bourne & Oates (2020) introduced the GR distribution modified with the exponential taper model of Kagan (2002). Its exceedance probability is given by:

$$P(M_i \geq M) = 10^{-b(M-M_m)} \exp \left[- \left(10^{1.5(M-M_C)} - 10^{1.5(M_m-M_C)} \right) \right]. \quad (6)$$

The parameter M_C is the corner magnitude where rollover of the MFD starts to occur. It is related to the exponential taper parameter, ζ , from Bourne et al. (2018) by

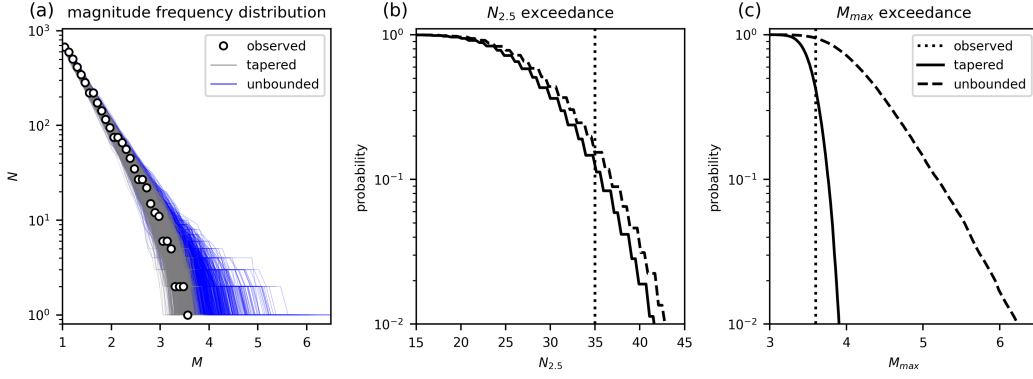


Figure 1. Summary of the Groningen magnitude frequency distribution (MFD) up to February 2017 and credible model fits. (a) Observed MFD (circles) and 1000 synthetic catalogs constructed from tapered (grey) and unbounded (blue) Gutenberg-Richter distributions. For the synthetic catalogs in (a), exceedance probability curves for (b) the number of events $M \geq 2.5$ or larger, and (c) the largest event in the catalog. Both tapered (solid) and unbounded (dashed) models are shown with that actual outcome indicated as a vertical dotted line, i.e., 35 $M \geq 2.5$ events, the largest of which was a 3.6.

$M_C = M_m - \frac{1}{1.5} \log_{10} \zeta$. As $M_C \rightarrow \infty$, the exponential factor approaches one, recovering the unbounded GR distribution, Eq. (5). Bourne & Oates (2020) fit Eq. (6) to Groningen seismicity with $M_m=1.5$ and obtained 95% credible intervals for the parameters: $0.84 < b < 1.07$ and $3.23 < M_C < 4.57$.

Fig. 1(a) shows how both tapered and unbounded GR distributions can fit the MFD of Groningen seismicity prior to February 2017. For this comparison, we have selected $b=0.9$ for both models, $M_C=3.5$ for the tapered model, and generated 1000 synthetic catalogs to compare against the observations. The aleatoric variability in both models approximately bounds the data, although the tapered model has less spread in its distribution.

Exceedance probability curves indicate that both the tapered and unbounded GR models provide a credible fit for the number of $M \geq 2.5$ events (Fig. 1b). The tapered model provides a better fit for the largest observed event with a probability to exceed M 3.6 of 0.43. In comparison, the unbounded model had 0.94 probability to exceed the observed maximum. Although the evidence suggests the tapered model is better supported by the data, in our view, it is not prudent to absolutely exclude the unbounded

model. This is mainly because of the large difference in $M_{\max}$ between the two models at the high exceedance limit, i.e., 1% likelihood to exceed M 3.9 for the tapered model and M 6.3 for the unbounded model.

Sensitivity of 2022 updated seismological model

Parameter sensitivity quantifies how certain model outputs may vary over plausible ranges of the model parameters. The key output whose sensitivity we are interested in is the prospective forecast and its exceedance curves. As examples for illustrating sensitivity, we have selected the stress state orientation, θ , and the poroelastic parameter, A . For each parameter, we constructed an ensemble of minimum misfit models, where each member of the ensemble had the sensitive parameter constrained to a narrow range. Then, we reconstructed the prospective forecast for each ensemble member. For example, θ was constrained in 10° intervals, and nine of these were constructed spanning the range $[0, 90^\circ]$. For A , we constructed nine evenly-spaced intervals in the range $[0.8, 1.0]$. Minimum misfit models, prospective forecasts and exceedance curves are shown in Figure 2.

The forecast model is not particularly sensitive to the parameter θ . Judged by eye, each model constructed over the tested range of θ provides a credible fit to the data, adequately capturing the pace of rise, peak and decrease in seismicity over the last 20 years (Fig. 2a). Furthermore, they all produce very similar prospective forecasts and exceedance curves for both tapered and unbounded GR models (Fig. 2B-C).

Certain elements of the forecast model are sensitive to the parameter A . First, the quality of model fit in the training period is more variable, although all models capture seismicity rise, peak and decline (Fig. 2d). There is more variability in both the rate forecast and exceedance curves for $N_{2.5}$. However, there is less variability in the $M_{\max}$ exceedance curves for both MFDs (Fig. 2E-F).

This exercise is useful to illustrate certain properties of history-matched forecasts. First, a model may achieve a reasonable match to the data when one of its parameters is knowingly assigned an incorrect value, e.g., θ . This happens because, during history matching, other parameters are able to take on values that compensate any misfit induced by the parameter change. This is parameter surrogacy (Doherty & Christensen, 2011) and it was a noted feature of the DS17 model.

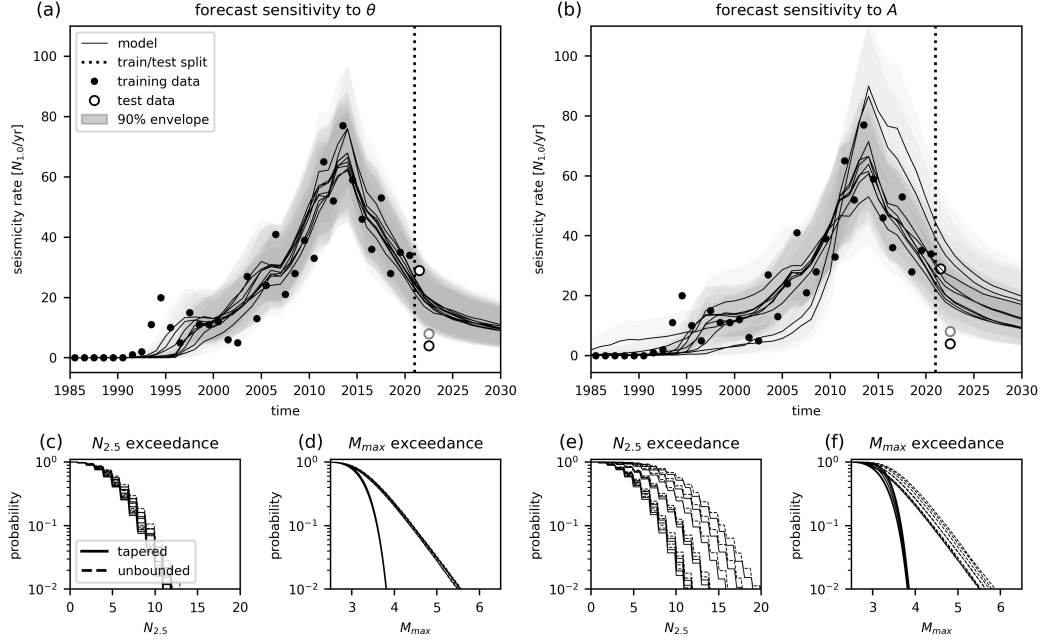


Figure 2. Summary of DS22 forecast sensitivity. (a) Nine best-fit models obtained when constraining θ to 10° intervals in the range $[0,90]$. Measured seismicity rate (circles) compared to best fit models (solid lines) with overlapping 90% aleatoric variability (grey envelopes). The greyed marker is the extrapolated event count for 2022 based on events in the first six months of the year. (b) Same as for (a) but instead nine best-fit models obtained when constraining A to evenly-spaced intervals in the range $[0.8,1.0]$. (c-d) Exceedance curves for prospective forecasts between 2021 and 2030 for the models in (a). (e-f) Same as for (c-d) but for the models in (b).

Second, a set of models with similar accuracy fitting the data have quite limited variability in their forecast outputs, regardless of the inferred correctness of those parameters. This, too, was a feature of the DS17 model and other forecasts that have sought to construct ensembles of well-fitting models (e.g., Richter et al., 2020; Smith et al., 2022).

Third, even seismicity rate models with dissimilar fits (Fig. 2d), some of which are relatively poor, may produce forecast outputs with little variability. In this case, convergence of the forecasts is induced by the upper limit in the tapered GR model or by the history match itself, which enforces seismicity rate decline on all models in the post-2014 peak period.

Description of depressurization scenario file

The file name is XY_PRF_Avg_OS2_CY.csv. The first two columns are x and y coordinates [m] of reservoir model output in the Groningen coordinate frame. Subsequent columns are depth-averaged pressure [bar] at the model output locations at the beginning of each year. The year is given by the column header. The spatiotemporal pressure evolution defined within this file was used to develop the forecast models described in the associated manuscript.

References

- Bourne, S., & Oates, S. (2020). Stress-dependent magnitudes of induced earthquakes in the groningen gas field. *Journal of Geophysical Research: Solid Earth*, 125(11), e2020JB020013.
- Bourne, S., Oates, S., & 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. *Geophysical Journal International*, 213(3), 1693–1700.
- Dempsey, D., & Suckale, J. (2016). Collective properties of injection-induced earthquake sequences: 1. model description and directivity bias. *Journal of Geophysical Research: Solid Earth*, 121(5), 3609–3637.
- Dempsey, D., & Suckale, J. (2017). Physics-based forecasting of induced seismicity at groningen gas field, the netherlands. *Geophysical Research Letters*, 44(15), 7773–7782.
- Dempsey, D., Suckale, J., & Huang, Y. (2016). Collective properties of injection-

- induced earthquake sequences: 2. spatiotemporal evolution and magnitude frequency distributions. *Journal of Geophysical Research: Solid Earth*, 121(5), 3638–3665.
- Doherty, J., & Christensen, S. (2011). Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. *Water Resources Research*, 47(12).
- Jaeger, J. C., Cook, N. G., & Zimmerman, R. (2009). *Fundamentals of rock mechanics*. John Wiley & Sons.
- Kagan, Y. Y. (2002). Seismic moment distribution revisited: I. statistical results. *Geophysical Journal International*, 148(3), 520–541.
- Richter, G., Hainzl, S., Dahm, T., & Zöller, G. (2020). Stress-based, statistical modeling of the induced seismicity at the groningen gas field, the netherlands. *Environmental Earth Sciences*, 79(11), 1–15.
- Segall, P., & Fitzgerald, S. D. (1998). A note on induced stress changes in hydrocarbon and geothermal reservoirs. *Tectonophysics*, 289(1-3), 117–128.
- 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 and Planetary Science Letters*, 594, 117697.
- Zoback, M. D. (2010). *Reservoir geomechanics*. Cambridge university press.