Stress-Dependent Magnitudes of Induced Earthquakes in the Groningen Gas Field

Bourne Stephen James[1] and Oates Steven[2]

[1]Shell Global Solutions International B.V.
[2]Shell

November 16, 2022

Abstract

Geological faults may produce earthquakes under the increased stresses associated with hydrocarbon recovery, geothermal extraction, CO2 storage. The associated risks depend on the frequency and magnitude of these earthquakes. Within seismic risk analysis, the exceedance probability of seismic moments, M, is treated as a pure power-law distribution, Mˆ{-β}, where the power-law exponent, β, may vary in time or space or with stress. Insights from statistical mechanics theories of brittle failure, statistical seismology, and acoustic emissions experiments all indicate this pure power-law may contain an exponential taper, Mˆ{-β}eˆ{-ζ M}, where the taper strength, ζ, decreases with increasing stress. The role of this taper is to significantly reduce the probability of earthquakes larger than ζˆ{-1} relative to the pure power-law. We review the existing theoretical and observational evidence for a stress-dependent exponential taper to motivate a range of magnitude models suitable for induced seismicity risk analysis. These include stress-invariant models with and without a taper, stress-dependent β models without a taper, and stress-dependent ζ models. For each of these models, we evaluated their forecast performance within the Groningen gas field in the Netherlands using a combination of Bayesian inference, and simulations. Our results show that the stress-dependent ζ-model with constant β likely offer (75–85%) higher performance forecasts than the stress-dependent β-models with $\zeta = 0$. This model also lowers the magnitudes with a 10% and 1% chance of exceedance over the next 5 years of gas production from 4.3 to 3.7 and from 5.5 to 4.3 respectively.

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

Stress-Dependent Magnitudes of Induced Earthquakes in the Groningen Gas Field

S. J. Bourne,[1] S. J. Oates,[1]

Corresponding author: S.J. Bourne, Shell Global Solutions International B.V., Kessler Park 1, 2288 GS Rijswijk, The Netherlands. (S.J.Bourne@shell.com)

[1]Shell Global Solutions International B.V., Grasweg 31, 1031HW Amsterdam, The Netherlands.

Abstract. Geological faults may produce earthquakes under the increased stresses associated with hydrocarbon recovery, geothermal extraction, CO_2 storage. The associated risks depend on the frequency and magnitude of these earthquakes. Within seismic risk analysis, the exceedance probability of seismic moments, $\mathcal{M}$, is treated as a pure power-law distribution, $\sim \mathcal{M}^{\beta}$, where the power-law exponent, β, may vary in time or space or with stress. Insights from statistical mechanics theories of brittle failure, statistical seismology, and acoustic emissions experiments all indicate this pure power-law may contain an exponential taper, $\sim \mathcal{M}^{\beta} e^{-\zeta \mathcal{M}}$, where the taper strength, ζ, decreases with increasing stress. The role of this taper is to significantly reduce the probability of earthquakes larger than ζ^{-1} relative to the pure power-law.

We review the existing theoretical and observational evidence for a stress-dependent exponential taper to motivate a range of magnitude models suitable for induced seismicity risk analysis. These include stress-invariant models with and without a taper, stress-dependent β models without a taper, and stress-dependent ζ models. For each of these models, we evaluated their forecast performance within the Groningen gas field in the Netherlands using a combination of Bayesian inference, and simulations. Our results show that the stress-dependent ζ-model with constant β likely offer (75–85%) higher performance forecasts than the stress-dependent β-models with $\zeta = 0$. This model also lowers the magnitudes with a 10% and 1% chance of exceedance

over the next 5 years of gas production from 4.3 to 3.7 and from 5.5 to 4.3 respectively.

1. Introduction

Induced seismicity may arise due to mining, geothermal energy production, artificial lakes, and fluid injection or production, including hydrocarbon production, water disposal or CO_2 storage. Most of these activities occur without inducing any noticeable earthquakes. Nonetheless, due to the quantity and scale of these activities, there is a growing number of notable occurrences of induced earthquakes. Several recent reviews comprehensively summarize the world-wide evidence for seismicity induced by human activities [Majer et al., 2007; Suckale, 2009; Evans et al., 2012; Davies et al., 2013; Ellsworth, 2013; Klose, 2013; NAS, 2013; IEAGHG, 2013; Foulger et al., 2018].

In such cases of induced seismicity, any exposure to the associated hazards of seismic ground motions or the risks of building damage must be assessed using probabilistic seismic hazard and risk analysis [*e.g. Elk et al., 2019]*, and if necessary mitigated. Induced seismicity is a transient non-stationary process in response to time-varying and significant increases in stress that are sufficient to destabilize previously inactive faults. Forecasting such failures within a geological material critically depends on its heterogeneity [Vasseur et al., 2015].

Heterogeneity falls into two classes. First, resolvable heterogeneity that may be mapped and accounted for explicitly with deterministic models such as the large-scale geometries of geological faults and reservoirs that may be mapped by reflection seismic imaging. Second, unresolvable heterogeneity, such as small-scale, spatial variations in geometric, elastic, frictional or prestress properties, will influence fault failures in ways that may only be fully-characterized with statistical models. The evolution of induced seismicity within a region exposed to increasing stress loads will depend on the relative amounts of variability present within these two classes of heterogeneity: resolvable (ordered) and unresolvable (disordered) variability.

In the limit that disordered variability in pre-stress significantly exceeds both the induced stress loads and earthquake stress transfers, the occurrence of any induced seismicity will be governed by the probability distribution of extreme pre-stress values. This approach led to the development of statistical models of induced seismicity occurrence based on Extreme Threshold Theory where the resolvable heterogeneities are included in a deterministic poro-elastic thin-sheet stress model and the unresolvable heterogeneities are represented by the upper tail of a pre-stress probability distribution given by the universal form of a Generalized Pareto distribution [Bourne and Oates, 2017b]. This simple model explains the observed, non-stationary, space-time statistics of induced seismicity within the Groningen gas field and provides a physical explanation for the exponential-like increase in seismicity rates relative to induced stress rates [Bourne et al., 2018]. In this limit of strong pre-stress disorder, the probability distribution of pre-stress explains the initiation of earthquakes will also influence the arrest of seismic slip, and therefore also the probability distribution of induced earthquake magnitudes.

Current methods of forecasting induced earthquake magnitudes are empirical and lack a clear physical basis. Natural and induced seismicity hazard analysis for the United States assumes a stationary process with a stress-invariant pure power-law distribution of seismic moments [Petersen et al., 2018]. Shapiro et al. [2010a] proposed a non-stationary model for fluid injection induced seismicity that includes a pre-stress disorder with a uniform distribution to model event occurrence but assumes a stress-invariant pure power-law distribution of seismic moments [Shapiro et al., 2010b; Langenbruch and Zoback, 2016; Shapiro, 2018]. Hazard analyses for Groningen induced seismicity included a stress-dependent, pure power-law distribution of seismic moments where the power-law exponent varied with reservoir compaction- [Bourne et al., 2014] or induced Coulomb stress [Bourne et al., 2018]. If this pure power-law assumption is not valid then all these models may be incomplete and biased in their earthquake magnitude forecasts especially under the significantly increasing stress loads often associated with induced seismicity.

This study seeks to extend the method of treating unresolvable heterogeneity as stochastic disorder to improve the seismological models used for forecasting induced earthquakes magnitudes for the purpose seismic hazard and risk analysis. We will build on previous work to incorporate the failure mechanics of disordered media into a statistical mechanics theory of natural earthquakes [*e.g. Bak and Tang, 1989; Alava et al., 2006; de Arcangelis et al., 2016]*, and their seismic hazard analysis Main [1996]. These statistical mechanic theories will be used to motivate the choice of models to evaluate, but not to rank or select them. Under many different theories the probability distribution of failure event sizes follows a power-law subject to an exponential taper where the power-law exponent is stress-invariant whilst the characteristic taper scale increases as a critical-point power-law with stress. However, under some other circumstances the power-law exponent may exhibit variation with stress. We reflect these possibilities by specifying 5 different classes of frequency-moment models for induced earthquakes:

1. Stress-invariant power-law with no taper
2. Stress-invariant power-law with a stress-invariant taper
3. Stress-dependent power-law with no taper

4. Stress-invariant power-law with a stress-dependent taper

5. Stress-dependent power-law with a stress-dependent taper

Using Bayesian inference we sample the full posterior distribution of possible models given the observed history of induced seismicity and induced stress within the Groningen gas field for a range of different parametrization choices within each of the 5 model classes. Our evaluation of the Groningen forecast performance for induced earthquake magnitudes reveals the best-performance requires a stress-dependent taper as anticipated by most statistical mechanics theories of brittle fracture.

After briefly stating the standard power-law formulation of seismic moments in statistical seismology (section 2), we will summarize the seismological literature that proposes (section 3) or opposes (section 4) evidence for stress-dependent variations of power-law exponent with stress. We will then describe our model of intra-reservoir induced stress due to pore-pressure changes (section 5) followed by simple statistical analyses of the variations in observed earthquake magnitudes induced by Groningen gas production (section 6). Then after reviewing existing statistical mechanics theories of earthquakes (section 7) we specify our models for the stress-dependence of induced earthquake magnitude distributions (section 8), infer their parameter values (section 9), assess their behavioural characteristics (section 10), and evaluate their performance (section 11), before assessing their implications for seismic hazard and risk (sections 12 and 13).

2. Power-Law Distribution of Seismic Moments

The exceedance probability distribution of earthquake magnitudes typically takes the form:

$$P(>M|>M_{\mathrm{min}}) = 10^{-b(M-M_{\mathrm{min}})}, \tag{1}$$

where M is the earthquake moment magnitude conditional on $M \geq M_{\mathrm{min}}$ and b defines the negative slope of the exponential distribution [Gutenberg and Richter, 1954]. Alternatively, this may be expressed according to the seismic moment, s, which scales with moment magnitudes as,

$$\log_{10} \mathcal{M} = (c + dM), \tag{2}$$

with, $c = 9.1$, and $d = 1.5$. Combining (1) and (2) leads to the equivalent power-law distribution,

$$P(> \mathcal{M} | > \mathcal{M}_{\mathrm{min}}) = \left(\frac{\mathcal{M}}{\mathcal{M}_{\mathrm{min}}} \right)^{-\beta}, \tag{3}$$

and $b = \beta d$.

Seismic hazard and risk analysis is highly influenced by the estimation of β-values. Lower β-values mean larger expected magnitudes and a larger expected maximum magnitude for a given population of earthquakes. In the next two sections we outline the existing evidence for two alternate hypotheses about the influence of stress on β-values.

3. β-Values Vary With Stress

A number of observations and modelling results might suggest that earthquakes b-value depends on the stress level. Measured earthquake b-values decrease systematically from 1.2 to 0.8 with increasing depth in the brittle crust from 5 to 15 km [Mori and Abercrombie, 1997; Spada et al., 2013]. Similar measurements indicate earthquake b-values vary systematically with focal mechanism rake angle as a proxy for stress [Schorlemmer et al., 2005; Gulia and Wiemer, 2010]. Lower stress, normal faulting b-values are typically 1.0–1.2. Whereas higher stress, thrust faulting b-values are typically 0.7–0.9. Intermediate stress, strike-slip b-values are in the range 0.9–1.0 [Wiemer and Wyss, 1997, 2002; Huang et al., 2018].

b-values also appear to be a proxy for shear stress and pore pressure [Scholz, 1968; Schorlemmer et al., 2005; Bachmann et al., 2012]. Bachmann et al. [2012] observed b-values decrease with a decrease in pore-pressure for induced earthquakes of an Enhanced Geothermal System. Whilst systematically smaller b-values were measured for earthquakes induced by larger reservoir compaction [Bourne et al., 2014] or higher Coulomb stress [Bourne et al., 2018] associated with natural gas production. Variations in measured b-values are also used to indicate material heterogeneity [Mogi, 1962; Main et al., 1992; Mori and Abercrombie, 1997] or for fault asperity mapping [Tormann et al., 2014]. The scale of fault heterogeneity appears to follow a power-law where its fractal dimension governs the b-value of seismic slip events within this fault population [Main et al., 1989, 1990, 1992]. Initial heterogeneities in the form of a fractal distribution of fault sizes or fault asperities are one way to explain the Gutenberg-Richter law. Another explanation is that is arises from some distribution of strength.

Variations in observed b-values may also be precursors of future rupture areas and sizes [Schorlemmer et al., 2005]. In this case, b-values decrease monotonically throughout the precursory phase, and then recovers abruptly after peak stress (marked by a sudden stress drop event). Scholz [1968] introduced a statistical model of brittle failure within an heterogeneous elastic medium to explain the apparent decrease in b-values with increasing stress.

4. β-Values Do Not Vary With Stress

In this hypothesis all observed b-values are consistent with a constant value in both space and time and any observed apparent variations are artifacts due to under-sampling (detection threshold and finite sample size effects), magnitude errors, non-homogeneous detection capabilities, and improper statistical tests [Shi and Bolt, 1982; Frohlich and Davis, 1993; Kagan, 1999, 2002b, 2010; Amorèse et al., 2010; Amitrano, 2012; Kamer and Hiemer, 2015]. For example, observed variations in b-values with stress rely on the maximum likelihood estimator Aki [1965], with corrections for the magnitude binning [Utsu, 1965; Bender, 1983; Tinti and Mulargia, 1987]. This method implicitly assumes that the underlying distribution is a pure power-law above some threshold of completeness according to equation (3). If this is not the case, then this estimator will be biased and confounded with any non-power-law stress-dependent variations in the frequency magnitude distribution, as we will show later.

Recent developments in statistical fracture and earthquake mechanics theories indicate that a wide range of physical mechanisms and conditions all lead to the same frequency-magnitude distribution that is a stress-invariant power-law with a stress-dependent exponential taper. We will now review these theories as a physical basis for β-values that do not vary with stress and to introduce an alternative stress-dependence for the frequency distribution of earthquakes induced by Groningen gas production.

5. Poro-Elastic Thin-Sheet Stress Model

The development of external loads on pre-existing weak fault structures within the Groningen gas field depends on the evolution of reversible reservoir deformations induced by pore pressure changes. Within the limit of small strains, these reservoir deformations are well-described by linear poro-elasticity. For thin reservoir geometries where the lateral extent of the reservoir greatly exceeds its thickness, the reservoir deforms predominately as a thin-sheet. Within the poro-elastic, thin-sheet approximation [Bourne and Oates, 2017b], depletion-induced reservoir displacement vector field, $\mathbf{u}(\mathbf{x})$, is constrained by symmetry to vertical displacements, $u(\mathbf{x})\hat{\mathbf{z}}$, where $\hat{\mathbf{z}}$ is the unit vertical vector. From this approximation it follows that the vertically-averaged incremental Coulomb stress states are:

$$\Delta C(\mathbf{x},t) = -\gamma H_p(\mathbf{x})\Gamma(\mathbf{x})\Delta P(\mathbf{x},t) \tag{4}$$

where $\Delta P(\mathbf{x},t)$ is the change in reservoir pore fluid pressure, $\Gamma(\mathbf{x})$ is the magnitude of lateral gradients in the elevation of the top surface bounding the thin-sheet, $\gamma = \nu/(1-2\nu)$ and ν is Poisson's ratio taken to be 0.25. $H_p(\mathbf{x})$ is a poro-elastic material property defined as:

$$H_p(\mathbf{x}) = \frac{H_s}{H_s + H_r(\mathbf{x})} \tag{5}$$

where H_s is a constant related to the shear modulus of the skeleton material comprising the poro-elastic medium and estimated as a model parameter. $H_r(\mathbf{x})$ is the time-invariant ratio of the observed reservoir depletion to the observed reservoir compaction strain, $\Delta P(\mathbf{x},t)/\epsilon_{zz}(\mathbf{x},t)$. Reservoir compaction strain is inferred from geodetic monitoring of surface displacements, and reservoir depletion is measured by in-well pressure gauges. For depletion, *i.e.* $\Delta P(\mathbf{x},t) < 0$, incremental Coulomb stresses increase towards frictional fault failure in locations where $\gamma H_p(\mathbf{x}) > 0$, otherwise depletion acts to increase frictional fault stability. In the presence of pre-existing faults that partially offset the thin-sheet, $\Gamma(\mathbf{x})$, is locally increased and acts to increase the sensitivity of Coulomb stress to pore-pressure changes. The deterministic reservoir map $-\gamma H_p(\mathbf{x})\Gamma(\mathbf{x})$ describes the time-invariant, local sensitivity of Coulomb stress to reservoir pore pressure changes, $\Delta C/\Delta P$, or stress susceptibility. This map was estimated by multiplying $\Gamma(\mathbf{x})$ and $-\gamma H_p(\mathbf{x})$ maps independently inferred from field observed quantities. $\Gamma(\mathbf{x})$ is computed from the top reservoir surface mapped by reflection seismic imaging.

Two modifications help to improve the performance of this thin-sheet model. First, we filter the contribution of individual faults to the topographic gradient field, $\Gamma(\mathbf{x})$, according to their juxtaposition geometry with the reservoir, by including fault segments according to the criterion:

$$r \leq r_{\max}, \tag{6}$$

where r is the local ratio of fault throw to reservoir thickness, and $r_{\max}$ is a model parameter. This represents the consequences of juxtaposition, where faults offset the reservoir against the overlying and ductile Zechstein salt formation. Increased juxtaposition of the reservoir interval against the Zechstein formation may limit induced seismicity by favoring ductile fault creep instead of a stick-slip behavior. Second, we use a smoothed incremental Coulomb stress model, $\Delta\tilde{C}(\mathbf{x}, t)$, evaluated as:

$$\Delta\tilde{C}(\mathbf{x}, t) = \int_S \Delta C(\mathbf{x}, t) G(\mathbf{x}, \mathbf{x}') dS' \tag{7}$$

a surface integral over the entire model domain, S, where $G(\mathbf{x}, \mathbf{x}')$ is the isotropic Gaussian kernel:

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mathbf{x}-\mathbf{x}')^2}{2\sigma^2}} \tag{8}$$

defined by the characteristic smoothing length-scale, σ.

This poro-elastic, thin-sheet stress model has three degrees of freedom $\{\beta_2, \beta_3, \beta_4\}$; the smoothing length-scale, $\beta_2 = \sigma$, the maximum juxtaposition ratio, $\beta_3 = r_{\max}$, and the poro-elasticity constant $\beta_4 = H_s$. These three parameters are optimized jointly with a given seismological models that defines the conditional probabilities of earthquake occurrence and size given the smoothed incremental Coulomb stress field, $\Delta\tilde{C}(\mathbf{x},t)$, and the observed catalog of induced earthquakes, $\mathbf{D}$.

This model is applicable to any reservoir subject to pore-pressure changes that is thin relative to it's lateral extent and smoothing length-scale such that uni-axial displacements dominate. In the particular case of the Groningen reservoir the posterior distribution of thin-sheet models inferred given the observed history of pore-pressure depletion, reservoir compaction, and induced seismicity [Bourne and Oates, 2017b]. For $M \geq 1.5$ event occurrences observed from 1/1/1995 to 1/6/2019, the maximum posterior probability thin-sheet parameter values are $\beta_2 = 3$ km, $\beta_3 = 0.41$, $\beta_4 = 10^{5.3}$ MPa (Appendix A). For $M \geq 1.5$ event magnitudes observed from 1/1/1995 to 1/6/2019, the maximum posterior probability thin-sheet parameter values are obtained using the event locations and origin times from are $\beta_2 = 3.5$ km, $\beta_3 = 1.1$, $\beta_4 = 10^7$ MPa (Appendix A). The apparent difference between the optimal smoothing length-scales between these two models is not significant as both posterior distributions include both values within their 95% credible intervals. The larger apparent different in the juxtaposition parameter, β_3, nonetheless yields very similar coulomb stress models and reflects the previously observed bi-modal distribution with modes at both $\beta_3 = 0.4$ and $\beta_3 = 1.1$ [Bourne and Oates, 2017b, Figure 12]. The apparent difference in skeleton modulus β_4 may also reflect inference uncertainty. Inference of a single thin-sheet model given may allow better forecast performance by utilizing both the observed event occurrences and magnitudes to constrain a single stress model but this was outside the scope of our current study.

6. Observed Seismicity

The Groningen field is located within the north-east of the Netherlands. The gas-bearing reservoir interval comprises the Upper Rotliegend Group (Permian) and the Limburg Group (Carboniferous) sediments, separated by the Saalian unconformity [Stauble and Milius, 1970]. The depth of the Rotliegend reservoir is 2600-3200 m. The field extent is controlled primarily by fault closures with occasional local dip closures. The top seal is the Zechstein salt. Production of Groningen gas started in 1963 and pressure depletion rates increased rapidly until 1973 before reducing significantly to conserve Groningen gas reserves. From 2000 to 2014, depletion rates rose moderately in response to increased market demand and decreased capacity of other smaller gas fields. Starting in 2014, depletion rates were significantly reduced in response to induced seismicity.

The Royal Netherlands Meteorological Institute (KNMI) has monitored seismicity in the Netherlands since at least 1986 [Dost et al., 2012]. For the Groningen Field earthquake catalog, the magnitude of completeness for located events is taken to be $M_L = 1.5$, starting in April 1995, with an event detection threshold of $M_L = 1.0$ [see **?**]. Here we restrict our analyses to the 279 events with $M_L \geq 1.5$ recorded within the Groningen Field between 1^{st} January 1995 and 1^{st} June 2019. Epicenters of events in the catalog are determined to within about 500-1000 m but, because of the sparseness of the monitoring array, depths were routinely estimated. For these events a depth of 3000 m–approximate reservoir depth–has been assumed. This is consistent with a limited number of reliable depth estimates from a reservoir-level borehole geophone array. Event magnitudes are reported as local magnitudes with a typical error of 0.1.

The spatial distribution of epicenters is localized within regions of the reservoir associated with larger incremental Coulomb stresses as represented by the poro-elastic thin-sheet model (Figure 1). Event origin times also appear to favor larger incremental Coulomb stress states (Figure 2) as most events occur at later times when incremental stresses are larger albeit subject to considerable variability. Likewise, larger magnitude events, *e.g.* $M \geq 2.5$, appear mostly localized in the times and places associated with the largest 20% of the exposed reservoir stress states. The stress-dependence of event occurrence probability appears to follow an exponential-like trend consistent with an Extreme Threshold theory of initial frictional reactivations within a heterogeneous and disordered fault system [Bourne and Oates, 2017b].

The observed frequency-magnitude distribution of events (Figure 3) shows clear evidence for under-reporting of $M < 1.5$ events and an apparent increase in variability with increasing magnitude due to finite sample effects. The apparent b-values of these events also appear to decrease systematically with increasing Coulomb stress [Bourne et al., 2018] or compaction-induced strain [Bourne et al., 2014]. However, the available surface displacements and seismicity observations cannot reliably distinguish between a stress or a strain driven process. Harris and Bourne [2017] demonstrated the observed frequency-magnitude distribution of 1995 to 2015 $M \geq 1.5$ events with epicentres inside a central elliptical region of the Groningen field is significantly different from those located outside this region with a statistical confidence exceeding 95% under the Kolmogorov-Smirnov test statistic. This elliptical region was centred close to the centroid of seismicity and oriented and sized to divide these events into approximately two equally-sized populations. Maximum likelihood estimates for the b-values were $b = 0.7$ and $b = 1.2$ for the inside and outside events respectively.Poro-elastic thin-sheet reservoir stress models indicate the region inside this ellipse experienced systematically higher maximum Coulomb stress states throughout this time period indicating these significantly lower b-value estimates are associated with a history of higher Coulomb stress states.

All these observations may however be an artifact of assuming a pure power-law frequency distribution of seismic moments without allowing for the possibility of other distributions such as a power-law with an exponential taper. All these previous observations also relied on catalogs of Groningen earthquake magnitudes reported to one decimal place. In the following sections we will assess the observable relationship between the distribution of earthquake magnitudes, now reported to 2 decimal places, and the reservoir stress history due to pore pressure depletion according to poro-elastic thin-sheet reservoir deformation model calibrated to the observed history of pore pressure depletion, surface displacements, and the space-time distribution of earthquake occurrences [Bourne and Oates, 2017b; Bourne et al., 2018].

6.1. Frequency-Magnitude Stress Dependence

To investigate stress-dependence of the frequency-magnitude distribution without making any assumptions about the particular form of this distribution we will use the Kolmogorov-Smirnov test statistic. First, we compute the incremental Coulomb stress, ΔC_i, at the origin time, t_i and epicentral location, $\mathbf{X}_i$ of each observed $M \geq 1.5$ event from 1995 to 2019, according to the poro-elastic thin-sheet reservoir model. Based on these values, we divide the events into two disjoint samples: a low stress sample, $\Delta C_i < \Delta C$, and a high stress sample, $\Delta C_i \geq \Delta C$. By increasing the stress threshold, ΔC, we compute the Kolmogorov-Smirnov test statistic p-value for all possible divisions of the events (Figure 4a), and repeat this procedure for alternative minimum magnitudes of completeness in the range $1 \leq M_{\min} \leq 2$ (Figure 4b). This p-value measures the probability that these two independent samples were drawn from the same underlying probability distribution.

The smallest p-values found are about 10^{-3} and correspond to $M_{\min} = 1.5$, a $\Delta C =$ 0.7 MPa stress threshold, with about 100 and 200 events in the low- and high-stress samples respectively. This result appears robust to alternative values of $M_{\min}$ such that the 95% confidence threshold is exceeded also for $1.0 \leq M_{\min} \leq 1.7$. For $M_{\min} > 1.7$, the loss of statistical power due to the smaller number of these larger events likely predominates. Consequently we conclude there is a statistically significant stress dependence in the frequency-magnitude distribution of Groningen induced earthquakes. Figure 5 shows the empirical exceedance distribution functions and epicentral map locations for this optimal stress-based division of the observed events.

By simple visual inspection, the different distributions appear consistent with β-values decreasing with stress or ζ-values increasing with stress. Ergodicity is implicit within this stress covariate hypothesis. That is to say a temporal stress change is indistinguishable from a spatial stress change of the same amount. The separation of high and low-stress events in space (Figure 5) more than in time (Figure 2) might indicate the influence of some initial spatial heterogeneity (quenched disorder). However, closer inspection of the map shows spatial mixing with many low- and high-stress events occurring in similar locations. This means there are three distinct spatial domains. A low-stress domain that has never experienced incremental stress above the 0.7 MPa threshold over the period of observation. A high-stress domain that has never experienced incremental stress below the 0.7 MPa threshold over the period of observation. Finally, an intermediate stress domain that has experienced stress states that have crossed the 0.7 MPa threshold at some time during the period of observation.

Any continuous stress-dependence of the frequency-magnitude distribution implies both samples still represent a mixture of different distributions reflecting the range of stress states within each sample. In this case sub-division of the events into more than 2 disjoint samples fails to reveal any reliable evidence for this which we attribute to the reduction of statistical power which limits our resolution of this stress dependency under the Kolmogorov-Smirnov non-parametric test statistic.

6.2. Apparent Stress Dependence of β- and ζ-Values

The significant stress-dependent differences found in the observed frequency-magnitude distribution may reflect a decrease in the power-law exponent, β, and the exponential taper exponent, ζ, with increased Coulomb stress. To measure any apparent variations of β- or ζ-values with Coulomb stress, we first ordered the $M \geq 1.5$ observed events from 1/4/1995 to 1/6/2019 according to the incremental maximum Coulomb stress at their time and place of occurrence within the poro-elastic thin-sheet reservoir deformation model. This yields a sequence of N incremental Coulomb stress values $\{\Delta C_1, \ldots, \Delta C_N\}$, and a paired sequence of event magnitudes $\{M_1, \ldots, M_N\}$. For the first k events in this paired sequence, we computed the posterior distribution of β-values for a constant β-value model with no exponential cut-off ($\zeta = 0$), and repeated this for every set of k consecutive events. Figure 6a shows the resulting β-value estimates and their uncertainties for $k = 20$ which tend to decrease with increasing Coulomb stress. A clear step-like decrease is evident at $\Delta C = 0.7$ MPa which is consistent with the previous Kolmogorov-Smirnoff test (Figure 4a). Such gradual evolution due to a mixing of different states has recently been demonstrated in lab data [Jiang et al., 2017] and also seen in volcanic seismicity [Roberts et al., 2016].

Repeating this procedure for a constant ζ-value model with fixed at its presumed universal value ($\beta = 2/3$), yields a similar trend of decreasing values with increasing Coulomb stress (Figure 6b). These piece-wise constant estimates for the variation of β- or ζ-values with Coulomb stress depend in detail on the choice of sample size, k. Larger k-values allow reduce uncertainties in the estimated β- and ζ-values but lower their resolution of any stress dependency. Likewise, smaller k-values increases stress resolution at the expense of precision. Nonetheless, similar results were obtained over a wide range of k-values indicating an apparent general tendency for β- and ζ-values to decrease with increasing Coulomb stress under the poro-elastic thin-sheet model. Once more, there is evidence for mode switching or mixing under increased stress.

7. Statistical Mechanics of Earthquakes

We will now briefly review the statistical mechanics aspects of earthquakes that motivate our choice of possible models that are included in the evaluation (Figure 7). We will only use these theories for hypothesis identification and *not* for hypothesis testing, which we will do instead using the available observations of Groningen induced seismicity.

Heterogeneity is the key to forecasting failure events within geological materials as consistently demonstrated in the laboratory experiments [Vasseur et al., 2015, 2017]. Statistical models distinguish themselves from deterministic models of fractures by incorporating the influence of unresolvable heterogeneities as stochastic disorder. Statistical theories of brittle rock strength originate with Weibull [1939] and now fall within a broad class of statistical models of fractures [*e.g. Alava et al., 2006]* and earthquakes [de Arcangelis et al., 2016].

Figure 8 illustrates this abstraction process of replacing the unknown distribution of fault heterogeneities (disorder) that influence the initiation and termination of frictional fault slip under an external stress with stochastic variables representing the probabilities of failure given the local stress states. These local stress states depend on the external stress and the redistribution of stresses due to previous failures.

Within these theories, the frequency-moment power-law may be derived in one of at least four different ways.

1. The geometric constraints associated with the number of permutations available for tiling rupture areas over a fault surface [Main and Burton, 1984].

2. Within the normalization group model for a wide-variety near-critical physical systems [*e.g.* chapter 15 Turcotte, 1997].

3. Within percolation theory near the percolation threshold [*e.g. Stauffer and Aharony, 1994]*.

4. Within self-organized criticality theory [Bak and Tang, 1989; Main, 1996].

Likewise, the frequency-moment distribution as a power-law with an exponential taper also has a physical basis in at least four different statistical mechanics theories.

1. Within fiber bundle models of brittle failure with equal-load sharing [*e.g. Pradhan, 2010]*.

2. Within percolation theory below the percolation threshold [*e.g. Stauffer and Aharony, 1994]*.

3. Within Ising models of brittle failure with local-load sharing.

4. Within interface theories of crack depinning in the presence of heterogeneity [*e.g. Daguier et al., 1997*].

5. Within information theory, using the concept of maximum entropy to find the least-informative probability distribution subject to observational constraints on the mean magnitude and mean total seismic moment rate [Main and Burton, 1984].

For earthquakes, we are concerned with the limit that these redistributed stress perturbations are small relative to the external stress known as damage mechanics. Damage mechanics models exist in two distinct classes (Figure 7). First, network models that address the evolution of failure across a distributed collection of interacting elements. Second, interface models that focus on the advance of a fracture tip line within a heterogeneous medium.

Network damage models take three key forms with respect to failures. Random fuse networks [Roux et al., 1988; de Arcangelis et al., 2007; Hansen, 2011], provide a model of brittle failure within a scalar central force network [Gilabert et al., 2007]. Each fuse within the network has a randomly assigned and invariant failure threshold (quenched disorder). Increasing external voltage leads to failure of individual fuses and re-distribution of current across the network that potentially triggers additional failures at constant applied voltage. Mean field theory [Toussaint and Hansen, 2006] shows this is a percolation process in the limit of infinite disorder [Roux et al., 1988] where re-distributed loads are equally shared. Random spring networks [Nukala et al., 2005] provide a model of brittle failure within a tensor central force network. Here, springs failure under a quenched random strain threshold and forces are re-distributed across the remaining spring network. Under simple shear loads, failure within this network is equivalent to random fuse networks.

Random block-spring networks [Burridge and Knopoff, 1967] represent frictional failures within a tensor central force network. A network of slider-blocks in frictional contact with a rigid basal surface and are connected to each other and to a driver plate by a network of springs. Displacement of the driver plate loads the blocks which slide when the basal shear exceeds the frictional threshold. Basal shear stresses are initiated as a random quenched disorder. Within mean field theory [Sornette and Physique, 1992], the first cycle of failures is equivalent to the fiber bundle model [Hansen and Hemmer, 1994; Hemmer and Hansen, 1992; Kloster et al., 1997; Pradhan, 2010]. Toussaint and Pride [2005] demonstrates an isomorphism of weak lattice damage models with fiber bundle model which in turn is isomorphic with percolation theory for equal load sharing or the Ising model for local load sharing. Using renormalization group theory, Shekhawat et al. [2013] unified the theories of fracturing within a disorder brittle material for infinite disorder (percolation) and zero disorder (nucleation) to show a power-law failure avalanche size distribution with an exponential-like taper for finite disorder. Also using renormalization group theory, [Coniglio and Klein, 1980] demonstrate a correspondence between percolation and Ising models.

An alternative theoretical approach is to represent an existing crack front as a deformable line that advances under an external stress through a random toughness medium [*e.g. Daguier et al., 1997]*. This crack front advances episodically between equilibrium states in which heterogeneities temporarily resist crack propagation. The resulting size of crack growth events depends on the competition between distortions of the crack front due to the material's inhomogeneities and the elastic self-stress field that acts to straighten this front [Bonamy and Bouchaud, 2011]. Within the theory elastic fracture mechanics and in the limit of quasi-static deformations, this crack depinning process leads to failure sizes distributed as a universal power-law with a stress-dependent exponential taper [Ponson et al., 2006]. This observation that some many diverse models all collapse to the same failure-size distribution is remarkable and motivates the application of statistical mechanics to seismic hazard analysis Main [1996]. In the limit that random pre-stress variability significantly exceeds induced stress loads and earthquake stress transfers then the frequency distribution of induced earthquake magnitudes may be described by mean-field theories within statistical fracture mechanics.

This phenomena is not limited to geological materials. A wide variety of physical systems exhibit crackling noise when driven towards failure slowly [J.P. Sethna et al., 2001] and the event-size distributions are power-laws with exponential-like tapers. Also with regard to fitting observed global natural seismicity, Kagan [2002b] strongly favors a power-law with an exponential taper and a universal value for β. He also finds no statistically significant evidence for any variations in β [Kagan, 2002a].

7.1. A Generalized Frequency-Moment Distribution

Following the common form of failure-size distributions found within a wide range statistical mechanics models of brittle failure, we follow Kagan [2002b] and write a generalized distribution for earthquakes according to the seismic moment, $\mathcal{M}$, exceedence probability (survival) function:

$$P(\geq M_o | M_o \geq M_{o,m}) = \left(\frac{M_o}{M_{o,m}} \right)^{-\beta} e^{-\zeta(\frac{M_o}{M_{o,m}} - 1)}, \tag{9}$$

where $\mathcal{M}_m$ is the lower threshold for completeness in the observed catalogue and the corner moment, $\mathcal{M}_c$, characterizing the exponential taper is $\mathcal{M}_c = \mathcal{M}_m/\zeta$. As expected, for $\mathcal{M} = \mathcal{M}_m$ the exceedance probability is 1.

Within these statistical mechanics models of a fault or fracture system being driven from stability towards critical instability β is a universal constant and $\mathcal{M}_c$ evolves as a power-law relative to the system's critical point, such that

$$\mathcal{M}_c = \frac{\mathcal{M}_m}{\zeta} \sim (\epsilon_c - \epsilon)^{-\gamma}. \tag{10}$$

Figure 10 illustrates how this survival function evolves with increasing ζ. The maximum likelihood estimator Aki [1965], with corrections for the magnitude binning [Utsu, 1965; Bender, 1983; Tinti and Mulargia, 1987] assumes $\zeta = 0$. If this is not true, the estimator becomes biased upwards. Figure 11 illustrates this bias using magnitudes simulated according to (9). When ζ scales as a critical-point function of external strain then this bias appears as a systematic and non-linear decrease in b-values. To evaluate the observed stress-dependency of earthquake magnitudes within the Groningen field we now require a suitable model for the development of stress due to depletion of reservoir pore-pressures associated with gas production.

8. Model Specifications

8.1. Power-Law Distribution With an Exponential Taper

We start by representing the seismic moment, $\mathcal{M}$, as an independent random variable distributed according to a power-law distribution with an exponential taper according to:

$$P(\geq \mathcal{M} | \mathcal{M} \geq \mathcal{M}_m) = \left(\frac{\mathcal{M}}{\mathcal{M}_m}\right)^{-\beta} e^{-\zeta(\frac{\mathcal{M}}{\mathcal{M}_m}-1)}, \tag{11}$$

where β is the power-law exponent, and s_m/ζ is the corner moment of the exponential taper. For ζ=0 this distribution reduces to the power-law previously given by (3).

The associated probability density of the tapered power-law model is

$$p(\mathcal{M}|\mathcal{M} \geq \mathcal{M}_m)\, d\mathcal{M} = \frac{1}{\mathcal{M}_m}\left(\beta + \zeta\frac{\mathcal{M}}{\mathcal{M}_m}\right)\left(\frac{\mathcal{M}}{\mathcal{M}_m}\right)^{-\beta-1} e^{-\zeta(\frac{\mathcal{M}}{\mathcal{M}_m}-1)}\, d\mathcal{M}, \qquad (12)$$

and the log-likelihood of this model given the set of seismic moment observations, $\mathcal{M}_i = \{\mathcal{M}_1, \ldots, \mathcal{M}_n\}$, follows as

$$\ell = \sum_{i=1}^{n}\left(\log(\beta_i + \zeta\frac{\mathcal{M}_i}{\mathcal{M}_m}) - (1+\beta_i)\log\frac{\mathcal{M}_i}{\mathcal{M}_m} - \zeta_i\left(\frac{\mathcal{M}_i}{\mathcal{M}_m} - 1\right) - \log\mathcal{M}_m + \log d\mathcal{M}\right), \qquad (13)$$

as previously given by Kagan [2002b]. If the observed seismic moments, $\mathcal{M}_i$, are computed from moment magnitudes according to (2) and these magnitudes are binned within intervals of size, ΔM, then the minimum seismic moment, $\mathcal{M}_m$, must be computed as

$$\log\mathcal{M}_m = \left(c + d(M_c - \frac{1}{2}\Delta M)\right)\log 10, \qquad (14)$$

where M_c is the magnitude of completeness above which all events within the region of interest are reliably detected and located. We will use this one general form of the log-likelihood function for the inference and evaluation of all the different possible earthquake magnitude models considered in this study.

A complete seismological model also requires a model for event occurrence, which we shall model according to the Extreme Threshold Failure model [Bourne and Oates, 2017b]. Within the Extreme Threshold Failure model, the occurrence rate of $M \geq 1.5$ events induced inside the Groningen reservoir are well-described by the Poisson intensity function

$$\lambda = h\Delta\dot{C}\theta_0 e^{\theta_1\Delta C}. \qquad (15)$$

Here, λ corresponds to the expected number of events per unit surface area and unit time. The apparent dependence on the local reservoir thickness, h, and stress rate $\Delta\dot{C}$ is not fundamental to this stress-dependent process. To clarify this, the corresponding expected event rate per unit volume and per unit of incremental Coulomb stress, χ_m, may be written as

$$\chi_m = \theta_0 e^{\theta_1 \Delta C}. \tag{16}$$

Here, χ_m characterizes the stress susceptibility of the system for inducing events of at least seismic moment $\mathcal{M}_m$. Multiplying (16) and (11), yields the generalised stress susceptibility, χ, for events of at least seismic moment $\mathcal{M}$ given $\mathcal{M} \geq \mathcal{M}_m$, such that:

$$\chi = \chi_m \left(\frac{\mathcal{M}}{\mathcal{M}_m} \right)^{-\beta} e^{-\zeta(\frac{\mathcal{M}}{\mathcal{M}_m} - 1)}. \tag{17}$$

Equation (17) defines a family of seismological models for induced seismicity conditioned maximum incremental Coulomb stress field, $\Delta C(\mathbf{x}, t)$, according to the poro-elastic thin-sheet equation (4). All that remains now is to specify the functional form of any magnitude stress dependence according $\beta = \beta(\Delta C)$ and $\zeta = \zeta(\Delta C)$. We will do this by specify four distinct and physically plausible model classes: stress-invariant magnitudes, stress-dependent β-values, stress-dependent ζ-values, and stress-dependent β and ζ values.

8.2. Stress-Invariant Distributions

This class of stress-invariant models has up to 2 degrees of freedom, $\{\beta, \zeta\}$ where the log-likelihood function (13) takes the special case where:

$$\begin{aligned} \beta_i &= \beta, \\ \zeta_i &= \zeta. \end{aligned} \tag{18}$$

Also of interest are two special cases with single degrees of freedom. The first case is an unknown invariant power-law with zero taper, specified as

$$\begin{aligned} \beta_i &= \beta, \\ \zeta_i &= 0. \end{aligned} \tag{19}$$

The second case is an unknown invariant taper with a known universal power-law, such that

$$\begin{aligned} \beta_i &= \frac{2}{3}, \\ \zeta_i &= \zeta. \end{aligned} \tag{20}$$

These basic invariant magnitude-frequency models are all unable to explain the significant difference observed between the low- and high-stress partitions of the Groningen earthquake catalog (Figures 4 and 5). Nonetheless, they provide useful performance references for the following two alternative classes of stress-dependent models.

8.3. Stress-Dependent β-Values

Within this class of models we represent the stress-dependence of the frequency-magnitude distribution according to (11) given $\zeta_i = 0$ and $\beta_i = f(\Delta C_i)$, where ΔC_i, is the maximum incremental Coulomb stress state at the occurrence time, t_i, and epicentral location, $\mathbf{x}_i$ of each event such that $\Delta C_i = \Delta C(t_i, \mathbf{x}_i)$.

As a first possible parameterization of $f(\Delta C_i)$, we will consider an inverse power-law of the form:

$$\begin{aligned} \beta_i &= \theta_0 + \left(\frac{\Delta C_i - \theta_1}{\theta_2} \right)^{-\theta_3}, \\ \zeta_i &= 0. \end{aligned} \tag{21}$$

To avoid implausibly large β-values we include the constraint $\beta_i = \min(\beta_i, 1)$. This model has 4 degrees of freedom $\{\theta_0, \theta_1, \theta_2, \theta_3\}$ where $\theta_1, \theta_2, \theta_3$ are non-negative. In general, β-values decrease with increasing Coulomb stress to the lower bound θ_0. This model has an asymptote at $\Delta C_i = \theta_1$ and so its range of physical validity is restricted to $\Delta C_i > \theta_1$.

The scale and shape of the stress dependence are given by θ_2 and θ_3 respectively which are both restricted to be non-negative. For $\theta_3 = 0$ the model is stress invariant, and for $\Delta C_i \ll \theta_1 \forall i$ the model reduces to a linear function of ΔC_i equivalent to the theoretical model proposed by Scholz [1968], whilst also limiting the extent of this linear region to avoid the non-physical possibility of negative β-values.

We will also consider an alternative parameterisation of $f(\Delta C_i)$ to represent a smooth step-like transition from an upper to a lower bound with increasing stress without increasing the degrees of freedom. This is motivated by Figure 6a and previous observations of mode switching in volcanic seismicity [Roberts et al., 2016].

$$\begin{aligned} \beta_i &= \theta_0 + \theta_1 \left(1 - \tanh(\theta_2 \Delta C_i - \theta_3)\right), \\ \zeta_i &= 0. \end{aligned} \tag{22}$$

In this case, the smallest and largest possible β-values are bounded such that, $\beta_{\min} = \theta_0$, and the largest possible decrease in the β-value with increasing stress is $\beta_{\max} - \beta_{\min} = \frac{1}{2}\theta_1$. The shape and location of this smooth step down in β-values are governed by θ_2 and θ_3 respectively. The observable performance of these two stress-dependent β-value models is not greatly sensitive to the these alternative parameterization choices as they both represent a smooth non-linear approach to a lower bound. They will differ in extrapolation to earlier times with lower stress as only the second model has an upper bound. However, under extrapolation to later times with higher stress the two models become equivalent as they approach a common lower bound. For seismic hazard and risk analysis we only require this second type of extrapolation.

8.4. Stress-Dependent ζ-Values

Within this alternative class of stress dependent models we represent the stress-dependence of the frequency-magnitude distribution according to (11) given $\beta_i = \beta$ and $\zeta_i = f(\Delta C_i)$. As a first parameterization, we model the stress dependence of ζ according to a critical-point power-law scaling motivated by statistical fracture mechanics [Alava et al., 2006, *e.g.*], such that

$$\begin{aligned} \beta_i &= \theta_0, \\ \zeta_i &= \begin{cases} \theta_1(\theta_3 - \Delta C_i)^{\theta_2} & \text{if } \Delta C_i \leq \theta_3, \\ 0 & \text{otherwise ,} \end{cases} \end{aligned} \tag{23}$$

where θ_3 is the critical stress of the system corresponding to the divergence of failure correlation length-scales and the onset of global failure. θ_2 is the non-negative critical exponent of this power-law, and θ_1 is a proportionality constant. So, as $\Delta C \to \theta_3$, then $\zeta \to 0$. This means seismic moments initiated under critical stress states are distributed as a power-law, whereas sub-critical stress states involve power-law with an exponential taper. Within this model, the power-law exponent, β, is a constant whilst the strength of the exponential taper decreases as stress states approach the critical point, as previously argued by Main [1995, 1996].

Given this parameterization choice, $\theta_1 = 0$ corresponds to the power-law distribution without any exponential taper, and $\theta_2 = 0$ corresponds to an exponential taper independent of the stress state. This model has 4 degrees of freedom $\{\theta_0, \theta_1, \theta_2, \theta_3\}$. The joint posterior distribution of these parameters given the Groningen events and stress model exhibit a trade off between parameters. This may be avoided by fixing θ_1 to its maximum a posterior probability (MAP) value, but doing so may also inadvertently bias the model.

Motivated by these findings, we consider an alternative positive definite parameterization of the ζ stress function with just 2 degrees of freedom which still allows for rapid decrease of ζ-values with increasing stress towards the critical point ($\zeta = 0$) in the form of an exponential trend:

$$\begin{aligned} \beta_i &= \theta_0, \\ \zeta_i &= \theta_1 e^{-\theta_2 \Delta C_i}. \end{aligned} \tag{24}$$

This alternative model has 3 degrees of freedom $\{\theta_0, \theta_1, \theta_2\}$. With this parameterization choice, $\theta_1 = 0$ corresponds to a pure critical-state power-law with no exponential taper for all stress states as also postulated in Main [1995, 1996].

Then for $\theta_1 > 0$, and $\theta_2 = 0$, then exponential taper is present but independent of the stress state. If both parameters are non-zero, then the exponential taper depends on the stress state, and for $\theta_2 > 0$ is follows that $\zeta \to 0$ as $\Delta C \to \infty$. So we see that this reduced parameterization if equivalent to the previous power-law choice in the limit that the critical stress point is much larger than the presently observed stress states. Although Taylor expansion of the power-law (23) under these conditions leads to a linear trend, *i.e.* $\zeta_i = \theta_1 + \theta_2 \Delta C_i$, this is not guaranteed to be positive definite without an additional constraint that creates a discontinuity in the first derivative leading to increased instability during inference. This linear form also lacks the requirement for non-linear growth in ζ with increasing sub-critical stress states. For these reasons we do not include an explicit linear parameterization for stress-dependent of ζ-values.

8.5. Stress-Dependent β-ζ-Values

Within this hybrid class of models we consider a 5-parameter combination of the hyperbolic-tangent stress-dependent β-model and the exponential stress-dependent ζ-

model defined here as

$$\beta_i = \theta_0 + \theta_1 \left(1 - \tanh \theta_2 \Delta C_i\right),$$
$$\zeta_i = \theta_3 e^{-\theta_4 \Delta C_i}. \tag{25}$$

Joint inference of the model parameters $\{\theta_0, \theta_1, \theta_2, \theta_3, \theta_4\}$ in-principle allows for competition between the two paradigms of stress-dependent β with $\zeta = 0$ and stress-dependent ζ with some universal fixed β. In practice, the limited number of observed events, the uncertainties in their magnitudes and reservoir stress states, and biased sampling of higher stress states may critically limit the statistical power of this most-complex model.

9. Bayesian Inference

Adopting the established methods of Bayesian inference we will estimate the set of parameters, $\mathbf{\Theta}_i$, for each of the specified models, M_i. Although models and magnitudes are both denoted by the same symbol M, they may be distinguished as models are always associated with an integer subscript, M_i, whereas any magnitude subscripts are restricted to M_c and M_t representing the completeness and threshold magnitudes respectively. (Table 1), given the observed earthquake data set, $\mathbf{D}$. According to Bayes' theorem:

$$\Pr(\mathbf{\Theta}_i|\mathbf{D}, M_i) = \frac{\Pr(\mathbf{D}|\mathbf{\Theta}_i, M_i)\Pr(\mathbf{\Theta}_i|M_i)}{\Pr(\mathbf{D}|M_i)}, \tag{26}$$

where $\Pr(\mathbf{\Theta}_i|\mathbf{D}, M_i \equiv P(\mathbf{\Theta}_i)$ is the posterior probability distribution of the model parameters, $\Pr(\mathbf{D}|\mathbf{\Theta}_i, M_i) \equiv L(\mathbf{\Theta}_i)$ is the likelihood distribution, $\Pr(\mathbf{\Theta}_i|M_i) \equiv \pi(\mathbf{\Theta}_i)$ is the prior distribution of parameter values, and $\Pr(\mathbf{D}|M_i) \equiv \mathcal{Z}_i$ is the normalization factor or Bayesian evidence. As $\mathcal{Z}_i$ is independent of $\mathbf{\Theta}_i$ it may be ignored for the purposes of model inference. Using standard MCMC methods provided by the Python library PyMC3 [Salvatier et al., 2015], we sample each model's parameter space distributed according to its un-normalized posterior using equilibrium Markov chains. This sampled posterior constitutes a complete joint inference of all parameter values, and may be marginalized over each parameter to yield individual parameter value estimates.

Relative to earlier studies [Bourne and Oates, 2017b; Bourne et al., 2018], our MCMC sampling methods incorporate three improvements. First, the adaptive Metropolis Hastings sampler was replaced with the No-U-Turn (NUTS) sampler that provides automatic tuning of the Hamiltonian sampler and uses symbolic derivatives of the likelihood function to improve sampling efficiency and reduce correlations between successive samples. Second, single trace sampling was replaced by multiple independent trace sampling in parallel on multiple CPU and, when possible, GPU cores. Third, sample chains are initiated by random draws from the prior distribution, $\pi(\mathbf{\Theta}_i)$, rather than at the parameter values that maximize the posterior distribution, $P(\mathbf{\Theta}_i)$. This last change avoids sampling bias and assists confirmation of sample repeatability between the independent Markov chains.

In addition, the earthquake data set, $\mathbf{D}$, incorporates two improvements relative to Bourne et al. [2018]. First, the seismological survey, KNMI, reduced the rounding of reported earthquake magnitude values from 0.1 to 0.01. Second, the observed time period increased by 18 months from 1/1/1995–1/1/2018 to 1/1/1995–22/5/2019 (6% increase), to incorporate another 20 $M \geq 1.5$ events within the Groningen catalog (7% increase).

For model inference from these data, we aim to use uninformative uniform prior distributions that honor non-negative conditions where applicable. The range of these distributions are sufficiently large such that further increases do not influence the posterior distributions.

9.1. Stress-Invariant Models

We trained the power-law distribution with an exponential taper model with constant β- and ζ-values as specified by (18) with uniform prior distributions: $0.3 \leq \beta \leq 1$, and $0 \leq \zeta \leq 1$. The resulting joint posterior probability density (Figure 13) indicates a β value consistent with its typically observed value, $\beta = 2/3$, and a non-zero ζ-value consistent with the presence of an exponential-taper on the power-law distribution of seismic moments within the Groningen catalogue. The posterior distribution obtained is characterized by the following mean values and 95% credible intervals defined by the highest posterior density:

$$\begin{aligned} \bar{\beta} &= 0.64 & (0.56 < \beta < 0.71) \\ \bar{\zeta} &= 1.2 \times 10^{-3} & (3.5 \times 10^{-5} < \zeta < 2.5 \times 10^{-3}) \end{aligned} \tag{27}$$

This is consistent with the usually-observed value of $\beta = 2/3$ and the presence of an exponential taper ($\zeta > 0$). The joint posterior probability density distribution (Figure 13) indicates no evidence for any strong covariance between the inferred β- and ζ-values that would appear as a clear diagonal trend in the distribution. That is lower than average β-values are equally likely to be paired with lower or higher than average ζ-values and vice-versa.

9.2. Stress-Dependent β-Models

9.2.1. Inverse power-law β-model

The posterior distribution of parameter values for the inverse power-law β-model specified according to (21) was sampled subject to uniform prior distributions, $1/3 \leq \theta_0 \leq 1$, $\theta_1 = 0$, $0 \leq \theta_2 \leq 1$, and $0 \leq \theta_3 \leq 10$. Our choice of this θ_0 lower bound reflects the absence of lower values in prior observations of stress-dependent b-values reported elsewhere [Mori and Abercrombie, 1997; Wiemer and Wyss, 1997, 2002; Schorlemmer et al., 2005; Gulia and Wiemer, 2010; Spada et al., 2013; Huang et al., 2018]. The joint posterior probability density distribution (Figure 14) indicates a bounded distribution with well-sampled and highly correlated uncertainties, and localized MAP values.

The posterior distribution obtained is characterized by the following mean values and 95% credible intervals (HPD).

$$\begin{aligned} \bar{\theta}_0 &= 0.49 \qquad (0.33 < \theta_0 < 0.63) \\ \bar{\theta}_2 &= 0.49 \qquad (0.28 < \theta_2 < 0.63) \\ \bar{\theta}_3 &= 5.77 \qquad (1.96 < \theta_3 < 10.0) \end{aligned} \tag{28}$$

This apparent variation with stress may be a statistical artefact of neglecting stress variations in ζ as illustrated in Figure 11. The posterior ensemble β function of incremental Coulomb stress (Figure 19a) are consistent with the previous finding of a significant difference between the frequency-magnitude distribution of events occurring under stress states below and above $\Delta C = 0.7$. Joint optimization of this magnitude-frequency model and the poro-elastic thin-sheet model with its three degrees of freedom (σ, $r_{\max}$, H_s), yields a similar ensemble function (Figure 19c) albeit with a broader prediction interval reflecting the additional variabilities within this ensemble stress model.

9.2.2. Hyperbolic Tangent $\boldsymbol{\beta}$-Model

The posterior distribution of parameter values for the hyperbolic tangent β-model specified according to (22) was sampled subject to uniform prior distributions, $1/3 \leq \theta_0 \leq 1$, $0 \leq \theta_1 \leq 2.5$, $0 \leq \theta_2 \leq 5$, and $\theta_3 = 0$. The joint posterior probability density distribution (Figure 15) once again indicates a bounded distribution with singular MAP values.

The posterior distribution obtained is characterized by the following mean values and 95% credible intervals (HPD).

$$
\begin{aligned}
\bar{\theta}_0 &= 0.52 \qquad (0.34 < \theta_0 < 0.67) \\
\bar{\theta}_1 &= 1.2 \qquad (0.27 < \theta_2 < 3.2) \\
\bar{\theta}_2 &= 2.2 \qquad (0.7 < \theta_3 < 4.2)
\end{aligned}
\tag{29}
$$

Under this alternative parameterization of the stress-dependent β-model, the correlation structures between the parameters do differ but a lead to similar evidence of apparent stress-dependence. The associated ensemble β-function of stress (Figure 19b) appears broadly similar to the inverse-power law model, with the largest differences limited to the lowest stress states. We attribute this to sampling bias as the observed events are significantly more prevalent under the higher stress states leaving few observations to constrain this low-stress response. Joint optimization of this magnitude-frequency model with the thin-sheet stress model leads to similar results once more (Figure 19d), and again with increased variability associated with counting the uncertainty in our knowledge of the stress states associated with each event.

9.3. Stress-Dependent ζ-Models

9.3.1. Power-Law ζ-Model

For the power-law ζ-model specified by (23), and given the constraint $\theta_1 = 1$, the posterior distribution obtained is characterized by the following mean values and 95% credible intervals (HPD).

$$
\begin{aligned}
\bar{\theta}_0 &= 0.65 \qquad (0.57 < \theta_0 < 0.72) \\
\bar{\theta}_2 &= 3.15 \qquad (0.02 < \theta_2 < 6.48) \\
\bar{\theta}_3 &= 2.24 \qquad (0.36 < \theta_3 < 4.00)
\end{aligned}
\tag{30}
$$

We set the constraint $\theta_1 = 1$ to avoid a trade-off found with θ_2 that is likely due to the limitations of finite sample size and under-sampling of the seismogenic response to lower stress states (see $\Delta C < 0.5$ in Figure 2). This constraint does not affect the subsequent out-of-performance of this model, but simplifies the posterior distribution. The estimated β-value, θ_0, is consistent with a universal value of $\beta = 2/3$. The posterior distribution of θ_3 takes values that are mostly larger than ΔC_i corresponding to $\zeta > 0$ reflecting the presence of an exponential taper to the power-law distribution of seismic moments. Furthermore, as the 95% confidence interval for θ_2 excludes $\theta_2 = 0$, there is significant evidence for ζ decreasing with increasing Coulomb stress in accord with the critical point scaling laws of statistical fracture mechanics.

9.3.2. Exponential ζ-Model

Within the exponential ζ-model defined by (24) the sampled posterior distributions (Figure 17) yield mean values and 95% credible intervals (HPD) as follows.

$$\begin{aligned} \bar{\theta}_0 &= 0.65 \qquad (0.57 < \theta_0 < 0.73) \\ \bar{\theta}_1 &= 0.42 \qquad (0.001 < \theta_1 < 0.93) \\ \bar{\theta}_2 &= 9.33 \qquad (5.9 < \theta_2 < 14.8) \end{aligned} \tag{31}$$

These results are insensitive to our choice of uniform prior distributions. The estimated β-value, θ_0, is once more consistent with the usually observed β-value of 2/3. These results also reveals significant evidence for $\theta_1 > 0$ which again reflects confidence about the presence of an exponential taper of the power-law seismic moment distribution. In addition, $\theta_2 > 0$ is a significant finding consistent with a stress-dependent exponential taper where ζ decreases under increasing Coulomb stress (Figure 20).

9.4. Stress-Dependent β-ζ-Models

Within the hybrid model that combines both β- and ζ stress dependence as defined by (24) the sampled posterior distributions remain stable with unique and localized MAP values (Figure 17). The associated mean values and 95% credible intervals:

$$\begin{aligned}
\bar{\theta}_0 &= 0.53 \qquad (0.34 < \theta_0 < 0.67) \\
\bar{\theta}_1 &= 1.1 \qquad (0.07 < \theta_1 < 2.2) \\
\bar{\theta}_2 &= 2.3 \qquad (0.7 < \theta_2 < 4.5) \\
\bar{\theta}_3 &= 0.33 \qquad (0 < \theta_3 < 0.85) \\
\bar{\theta}_4 &= 9.7 \qquad (4.9 < \theta_4 < 15)
\end{aligned} \tag{32}$$

indicate significant in-sample evidence for stress dependence of both β- and ζ-values.

10. Model Characteristics

Figure 21 illustrates how the stress susceptibility, χ_s defined by equation (17), varies with Coulomb stress, ΔC, according to the different magnitude-frequency models. The particular instance of each model was selected according to MAP parameter values given the observed Groningen events and poro-elastic thin-sheet stress model. The different lines in each plot show how stress susceptibility varies for different magnitude thresholds. All models share the fundamental property of monotonic increases in susceptibility with stress, so in all plots every line moves up to the right. Looking beyond this similarity, there are key and distinguishing differences between each of these magnitude-frequency models.

For the simplest magnitude-frequency model of a constant β and no exponential taper, $\zeta = 0$ (Figure 21a), all susceptibility lines are straight, parallel and equally-spaced on this log-linear plot. These lines remain straight and parallel because the model is invariant under increasing stress, and they remain equally spaced because frequency-moment distribution is a pure power-law. With the introduction of a stress-invariant exponential taper, $\zeta > 0$ (Figure 21b), these lines remain parallel reflecting the stress invariance of the model, but the line spacing increases with the magnitude threshold and rapidly so above the corner magnitude as the exponential tail dominates. In this example the corner magnitude is 3.5. The key difference between these two stress-invariant magnitude-frequency models is the expected rate of larger magnitude events at larger incremental stress states (lower right corner of these plots). For example, the emergence of $M \geq 4.5$ susceptibility rates above 0.5×10^{-10} /m^3/MPa (stress-axis intercept) increases from 0.75 MPa to 0.85 MPa by including the exponential taper. This highlights the importance of any non-zero taper for induced seismicity hazard and risk analysis that are typically driven by larger than previously seen magnitudes under larger than previously experienced stress states.

Under both stress-dependent models (Figure 21c, d), these lines are neither straight, nor parallel nor equally spaced. The only common feature is the top line corresponding to the rate of $M \geq 1.5$ events which follows the same Extreme Threshold exponential trend given by (16) in all models. The stress-dependent β-model with no taper (Figure 21c) has a constant line space for any given incremental stress. This is most easily recognized on the right side of the plot but is true everywhere. This line spacing decreases with stress, reflecting the smaller β-values at larger stresses. This means the largest line spacings occur for the smallest magnitudes at the smallest stress states (lower left corner). Consequently, look along the stress axis, $\chi_s = 0.5 \times 10^{-10}$ /m^3/MPa, the line spacing decreases with stress. This means the additional stress required to exceed the next magnitude threshold becomes progressively *smaller* as the system evolves to higher stress states.

The opposite is true for the stress-dependent ζ-model with a constant β-value (Figure 21d). Along the stress-axis, line spacing increases with stress. This indicates the additional stress required to exceed the next magnitude threshold becomes progressively *larger* as the system evolves to higher stress states. Consequently the largest line spacing occurs for the largest magnitude thresholds at the largest stress states (lower right corner). The key distinguishing feature of these two stress-dependent models is the intercept of each line on the incremental Coulomb stress axis.

11. Model Evaluation

We take two complimentary approaches to evaluating model performance for forecasting event magnitudes. The first compares out-of-sample posterior model likelihood distributions for the most recent subset of observed events (2012–2019). The second compares the observed and model-based simulations of maximum magnitude and total seismic moment time series over the entire history of gas production (1965–2019).

11.1. Out-of-Sample Likelihoods

We favour out-of-sample over in-sample likelihoods as a better measure of the forecast performance required by seismic hazard and risk analysis. Typical hazard and risk analysis periods for Groningen induced seismicity are 5 to 10 years and are always beyond the current observation period [Elk et al., 2019]. This means seismicity forecasts rely on near-term extrapolations of the seismological models conditioned on a given gas production scenario. We therefore choose to exclude all in-sample model evaluation methods, such as the Bayesian Information Criterion, as these do not properly reflect this out-of-sample forecast requirement.

The most reliable measure of forecast performance is a blind prediction made prior to the observations required to evaluate forecast performance. In our case, this means waiting for at least 5 years. To avoid such a delay, we evaluate out-of-sample model performance using the existing observations. To do this, we divide the observed earthquake data set, $\mathbf{D}$, into two disjoint, time-contiguous parts, $\mathbf{D}_1$ and $\mathbf{D}_2$, corresponding to a training period, $\mathcal{T}_1$, and an evaluation period, $\mathcal{T}_2$. In this study, when not specified otherwise, these periods are $\mathcal{T}_1 = 1/1/1995$ to $31/12/2012$, and $\mathcal{T}_2 = 1/1/2013$ to $1/6/2019$. This choice splits the data into approximately two equal parts, and also ensures the evaluation period covers at least 5 years to represent the typical forecast demand for these seismological models. This is a form of cross-validation where the choice of out-of-sample data is restricted to reflect the forecast requirement. This is a retrospectively 'blind' test where the choice of the start time for the out of sample 'future' events was made prior to, and independent from, the later analysis. Nonetheless there remains a residual possibility of unconscious researchers' bias influencing our analysis. Indeed, true forecast performance typically lags behind hindcast performance within meteorological models.

To evaluate out-of-sample model performance we first sample the posterior joint model parameter distribution, $P(\mathbf{\Theta}_i)$, given $\mathbf{D}_1$ according to (26). Then we sample the out-of-sample posterior predictive distribution of likelihood values, $L(\mathbf{D}_2|\mathbf{\Theta}_i)$, for the $\mathbf{D}_2$ data set given the sampled posterior distribution $P(\mathbf{\Theta}_i)$ obtained in the first step. These results are summarized by the distribution of log-likelihood values evaluated as

$$\ell_i = \log L(\mathbf{D}_2|\mathbf{\Theta}_i). \tag{33}$$

By this measure, every model has zero degrees of freedom to explain the out-of-sample observations, $\mathbf{D}_2$, as it is not fitted to these data. Models with too many degrees of freedom will tend to yield posterior distributions that over-fit the in-sample observations, $\mathbf{D}_1$, with highly variable parameter values. This likely increases bias and reduces precision in model-based forecasts for the out-of-sample observations, $\mathbf{D}_2$, which systematically reduces the out-of-sample log-likelihood values obtained according to equation (33). Likewise, models with too few degrees of freedom, will likely fail to fit enough of the observed variations within $\mathbf{D}_1$ and carry-over this deficiency. Limitations associated with small sample sizes may confound this evaluation due to chance effects that increase performance variability and broaden the measured out-of-sample log-likelihood distribution. This limits our ability to reliably rank the model when their log-likelihood distributions overlap.

Instead, we use these distributions to measure the probability, P_{ij}, of one model, M_i, out-performing another model, M_j, according to the probability of ℓ_i exceeding ℓ_j:

$$P_{ij} = \Pr(\ell_i > \ell_j). \tag{34}$$

This probability P_{ij} is estimated by the fraction of randomly sampled pairs from their respective distributions that satisfy this criterion. Posterior distribution sample sizes are made large enough to ensure sampling errors, ΔP_{ij} are insignificant when comparing models (*e.g.* $\Delta P_{ij} < 0.01$). This was verified but increasing the sample size to demonstrate the results at this level of precision are reproduced. Accordingly, self-comparison of any model yields $P_{ii} = 0.5$.

Figure 22a shows the out-of-sample log-likelihood distributions obtained for the three stress invariant models. Better performance appears as larger log-likelihood values so the best and worst versions of a model are found in the upper and lower tails of these distributions respectively. As the distributions all overlap the ranking of model performance is somewhat ambiguous. So although the best performances are associated with the upper tail of the M_3, the upper tails of M_1 and M_2 still exceed the performance of the lower tail of M_3. Nonetheless, the two models that allow for the presence of an exponential taper, $\zeta \neq 0$ (M_2, M_3), are both capable of better performance than the baseline model without any exponential taper ζ=0 (M_1) as shown by the locations of their upper tails. Likewise, all stress-dependent models also exhibit better performance than the baseline model M_1 (Figure 22b) as their upper tails all exceed the upper tail of M_1. However, within these models the performance gain of an exponential taper appears much less clear as the three models with the best performing upper tails include two with $\zeta = 0$ (M_5, M_7), and one that combines stress-dependent β and ζ effects (M_{13}).

The complete $\mathbf{D}_2$ data are dominated by the smallest magnitude events, so for instance half of the observed events are in the range $1.5 \leq M \leq 1.8$ compared to the largest observed magnitude at $M = 3.6$. Since these models are intended for probabilistic seismic hazard and risk assessment their performance in forecasting larger magnitude events must be considered. We start to do this by increasing the magnitude threshold, M_t, for the events admitted into the $\mathbf{D}_2$ data set to obtain the subset $\mathbf{D}_{2t}$. Then the out-of-sample likelihood analysis is repeated using the same posterior distributions of parameter values as before, $P(\mathbf{\Theta}_i)$, to evaluate the out-of-sample likelihood values $L_t(\mathbf{D}_{2t}|\mathbf{\Theta}_i)$. The modified likelihood function, L_t, is given by equations (13) and (14) where $M_c = M_t$. In this manner the models are still trained by all $M \geq 1.5$ events within the training data but then evaluated only on the larger $M \geq M_t$ events within the out-of-sample evaluation data.

Figure 23 shows the likelihood distributions obtained for magnitude thresholds $M_t = \{1.75, 2.0, 2.5\}$. Once more, the better performing models are located within the upper tails of each distribution. Table 2 summarise the performance of all models relative to the baseline model, M_1, according to the P_{i1} metric as specified by (34). As the magnitude threshold increases, it is clear that the performance of ζ=0 models significantly decreases from a top-ranked performance for $M \geq 1.5$ to a bottom-ranked performance for $M \geq 2.5$. Furthermore, the only models that fail to exceed the baseline model performance ($P_{i1} \leq 0.5$) are those with a stress-dependent β-values (M_5, M_7, M_{13}). This indicates that β-values which decrease with increasing Coulomb stress do not describe the tail of the observed magnitude distribution as well as any of the other models which all possess stress-invariant β-values.

In contrast, the performance of $\zeta \neq 0$ models with constant β-values either improve (M_2, M_3) or remain stable (M_{10}, M_{11}) under increasing magnitude thresholds. As expected, the presence of an exponential taper measurably improves the out-of-sample forecast performance for $M_t \geq 2$ events. However, within this analysis, there is no evidence for stress-dependent ζ-values as stress-invariant $\zeta \neq 0$ models perform marginally better against the baseline model for $M_t \geq 2$.

11.2. Simulated Seismic Moments and Magnitudes

Simulation of event catalogs using the different magnitude-frequency models allows their performance to be evaluated regarding the time series of maximum magnitudes and total seismic moment time series. Such an evaluation differs from the previous consideration of out-of-sample likelihood given the observed magnitudes by testing the simulation results and placing greater emphasis on forecasting the larger magnitudes that most-influence seismic hazard and risk analysis. The time series of total seismic moments represents the cumulative sum of seismic moments for all prior $M \geq 1.5$ events. Likewise, the time series of maximum magnitudes represents the largest magnitude observed so far. These simulated time series depend on both the simulated number and magnitude of events. Over the typical range of β-values associated with this seismicity most of the total seismic moment is contributed by the maximum magnitude event. As such, both time series are closely related, but with one key distinction. The total seismic moment in 1995 is unknown whereas the maximum magnitude is known to be $M_{\max} = 2.4$ from a regional monitoring network reporting all $M \geq 2$ events.

Figures 24 and 25 compare the observed and simulated time series of maximum magnitudes and total seismic moments for 2 stress invariant and 3 stress-dependent magnitude-frequency models. All these results share the same event occurrence simulations based on the posterior distribution of Extreme Threshold Failure models [Bourne and Oates, 2017b]. The posterior distribution of all models were obtained using the just $\mathbf{D}_1$ so the out-of-sample observations in this case occur prior to 1/1/1995 and from 1/1/2013. The simulations were run from 1965 to 2019 using the reservoir pore pressure model.

The model of stress-invariant β-values given ζ=0 (M_1, uni) systematically over-predicts both time series for all observed events and exceeds the 95% prediction interval for maximum magnitudes. Maximum magnitude time series residuals (Figure 26) indicate the absolute mean residual ($\Delta M_{\max} = -0.5$) is significantly larger than the expected magnitude measurement error ($\pm$ 0.1–0.2). The upper bound of the 95% prediction interval is always about 2 magnitude units above the observed maximum magnitude. A similar over-prediction bias is seen in the total seismic moments time series where the median time series always exceeds the observed total seismic moment after 1996. Including a stress-invariant taper of the frequency-magnitude distribution (M_3) significantly reduces the simulation bias whilst also significantly increasing its precision. This is shown by the reduced width of the 95% prediction interval that still contains all the variability in the observed total seismic moment time series although the first half of the time series (1995-2007) is systematically over-predicted. This early over-prediction bias is also seen in the maximum magnitudes time series and even exceeds the 95% prediction interval. The appearance of increasing precision with time does not reflect the increase observational constraints but instead appears due to the influence of the stress-invariant exponential taper that starts to significantly lower the probability of $M \geq 3.5$ events.

The inverse power-law model for stress-dependent β-values with ζ=0 (M_5, ets0.ipc3), exhibits the same systematic tendency for over-prediction of both maximum magnitudes and total seismic moments albeit to a lesser extent and without exceeding the 95% confidence interval. The absolute mean maximum magnitude bias ($\Delta M_{\max} = 0.4$) is still significant relative to the magnitude measurement errors. In contrast, the exponential stress-dependent ζ-model with $\beta = 2/3$ (M_{11}, ets0.ltc3) exhibits no bias in either maximum magnitudes or total seismic moments and does not exceed the 95% prediction interval despite this interval being significantly smaller than the previous two models. Moreover, the observed variability approaches both the upper and lower bounds of the simulated variability. As such this model demonstrates zero bias and a simulated variability consistent with the observed variability.

Figure 27 shows the distributions of out-of-sample likelihood, $L_s(\mathbf{D}_{2t}|\mathbf{\Theta}_i)$, for observed maximum magnitudes, $\mathbf{D}_{\max,2}$, given the simulated maximum magnitude time series. We obtained the maximum magnitudes data set, $\mathbf{D}_{\max}$ by selecting the subset of events within the complete data set, $\mathbf{D}$, that are larger than all previous events. This data set is then partitioned according to event origin times to yield the out-of-sample maximum magnitudes data set $\mathbf{D}_{\max,2}$.

We estimate the out-of-sample likelihoods, L_s, using the distribution of simulated maximum magnitude time series according to the posterior model distribution, $P(\mathbf{\Theta}_i|\mathbf{D}_1)$. We performed these simulations in a nested manner to yield simulated maximum magnitude time series, $\mathcal{S}_{ijk}$, where i denotes the model, j denotes a single random sample from the posterior distribution, $P(\mathbf{\Theta}_i|\mathbf{D}_1)$, and k denotes the simulation index. In this manner, time series $\mathcal{S}_{ijk}$ represents the k^{th} simulation of the j^{th} posterior sample from the i^{th} model.

For each i and j, and observed event within $\mathbf{D}_{\max,2}$, we select the set of simulated maximum magnitudes at the observed origin times. Using a Gaussian kernel density estimate for the probability density function of these simulated maximum magnitudes, we compute the likelihood of this observed maximum magnitude. Repeating this for all $\mathbf{D}_{\max,2}$ events L_s is estimated as the product of these single-event likelihood values. Repeating this for all values of j yields the posterior distribution of likelihood value for model M_i. Repeating all of these steps for all values of i results in the collection of likelihood distributions shown by Figure 27. Better performing models appear with likelihood distributions located to the right of poorer performing models. However, as all these distributions overlap with different tail shapes the relative performance ranking is not completely clear. Nonetheless it is clear the top two models are both include stress-dependent ζ-values.

As these distributions substantially overlap, we summarize the overall relative performance according to the probability of one model yielding a better out-of-sample likelihood than another model, P_{ij}, according to (34). Table 3 shows the pairwise probabilities P_{ij}

that model M_i out-performs model M_j given their respective out-of-sample maximum magnitude likelihood distributions. If the models are listed in rank order of performance then this matrix, P_{ij} would show monotonically increasing values as i increases (moving top to bottom in each column of Table 3) and as j decreases (moving right to left in each row of Table 3). In this case, the results are less clear in the sense that no such complete and unambiguous ranking exists. This is because the relative performance of several middle-ranking models are so similar, likely due to the small out-of-sample size available. Nonetheless, we may still confidently identify the best-performing models.

The baseline model (M_1) is out-performed by all other models, although none exceed 95% probability, although two models (M_{11}, M_{13}) are close with a 94% chance of exceeding the baseline performance. The only common feature of these two models is a stress-dependent ζ variation. One stress-dependent β model (M_7) does indicate a 91% chance of exceeding the baseline performance, but its chances of out-performing the leading two models are just 25% and 19% respectively. Ranking all models by increasing performance based on this metric yields $\{M_1, M_2, M_3, M_5, M_{10}, M_7, M_{11}, M_{13}\}$. This is essentially the numerical model sequence shown in Table 3 except the best-performing stress-dependent β-model, M_7, exchanges places with the worst performing stress-dependent ζ-model, M_{10}. We attribute this to a poor parametrization choice for M_{10} resulting in a large trade-off between the θ_2 and θ_3 parameters shown in Figure 16 associate with insufficient information to constrain the location of the critical point, θ_3.

As previously discussed in section 10, a key diagnostic is the time series of counts for events that exceed a given magnitude threshold (see Figure 21), especially the early time evolution. To revisit this, we compute the mean simulated cumulative event count time series over a range of magnitude exceedance thresholds. Figure 28 compares these simulated time series with the observed events from 1995 to 2019. The simulated $M \geq 1.5$ cumulative counts are identical for all models as this depends on the activity rate model alone which is shared by all magnitude-frequency models. For larger magnitude exceedance thresholds, the simulated time series differ between the different magnitude-frequency models.

For the stress-invariant β-model with ζ=0 (M_1, Figure 28a), the simulated count time series is systematically over-predicted for magnitudes $M \geq 2.5$ during their early time evolution. This error increases with magnitude threshold, such that the simulated $M \geq 4$ counts are comparable to the observed $M \geq 3.5$ counts. A similar bias is also apparent in the stress-dependent β model (Figure 29a), although in this case over-prediction is only apparent for magnitudes $M \geq 3.0$. The stress-invariant beta-zeta model (M_3, Figure 28b) improves the fit for larger magnitudes ($M > 3$) at the end of the time period but also over-predicts for magnitudes $M \geq 2.5$ during their early time evolution.

The stress-dependent β model (M_{11}, Figure 29a) slightly improves the early time prediction of $M \leq 2.5$ events relative to the stress-invariant models, but systematically over-predicts events counts for $M \geq 3.5$ events. The stress-dependent ζ model (M_{11}, Figure 29b) shows the best match to the observed rates with no apparent bias for any of the observed magnitude thresholds especially at later times when the fractional count errors are smallest. The higher dimensional, hybrid model with stress-dependent β- and ζ-values (M_{13}, Figure 29c) clearly under-predicts the numbers of $M \geq 2.5$ and $M \geq 3$ events. This result would be counter-intuitive if model performance was evaluated on the same data used for model inference. In this case, the extra degrees of freedom should give a better fit. However, we evaluate the model on data not previously used for model inference. This means any over-fitting associated with too many degrees of freedom will likely appear as bias in this out-of-sample evaluation.

12. Discussion

12.1. Probabilities of Larger Magnitudes

Central to the analysis of seismic hazard and risk induced by Groningen gas production is the reliable forecasting of larger than previously experienced magnitudes. Probabilistic analyses for Groningen show magnitudes in the range 4.5–5.5 provide the largest contribution to seismic hazard and risk metrics associated with public safety within the built environment [Bourne et al., 2015; Elk et al., 2019]. Smaller magnitudes are always too small to influence the hazard and risk metrics, whilst larger magnitudes are too infrequent to influence the hazard and risk metrics.

Using the gas production history and a single future gas production scenario (2019 GTS Raming) we simulated earthquakes catalogs for the entire history of gas production (1965–2019), and the next 5 years of future gas production (2019–2024). Earthquake occurrence was simulated used the posterior distribution of Extreme Threshold Failure models [Bourne and Oates, 2017b] inferred using the 1995–2019 $M \geq 1.5$ events and a poro-elastic thin-sheet Coulomb stress model (section 5). Earthquake magnitudes were simulated using the $M_1, M_3, M_7, M_{11}, M_{13}$ magnitude-frequency models in turn.

Figure 30 shows the distribution of maximum simulated magnitudes associated with the first 54 years (1965–2019) and the next 5 years (2019–2024) according to the different magnitude-frequency models. Over the observed period (1965–2019) the exponentially-tapered ζ-models (M_3, M_{11}, M_{13}) most closely match the observed maximum magnitude,

$M = 3.6$. Models that lack such a taper (M_1, M_7) over-predicted this magnitude 80–90% of the time with a mean over-prediction bias of 0.5–0.7 in magnitude. Maximum magnitudes are an important model performance metric as they typically form part of traffic light systems used to trigger interventions as is the case for Groningen induced seismicity.

Simulation-based forecasts for the next 5 years yield different distributions for the maximum magnitude event all with the same mode $M = 3.2$ and similar medians (Table 4, 50%, M=3.2–3.5). However, hazard and risk are driven by larger, less-likely magnitudes in the upper tail of these distributions and differences between these upper tails are considerable. Table 4 shows models with an exponential taper ($M_2, M_3, M_{10}, M_{11}, M_{13}$) exhibit much lower magnitudes with a 1% chance of exceeding (3.9–4.5) than the other models (M_1, M_5, M_7, M_{11}) that lack an exponential taper (5.3–5.5). This is a difference of 0.8–1.6 in magnitude.

Forecasts over longer periods necessarily face increasing uncertainties associated with larger extrapolations of the pore-pressure depletion model given the observed depletion history, and also larger extrapolations of the seismological model given the observed seismicity history. Limiting forecast periods to 5 years or less limits our exposure to extrapolation related uncertainties.

12.2. Including an Upper Bound

The magnitudes models, as formulated so far, do not include an upper bound, corresponding to a maximum possible magnitude, $M_{\max}$. For any finite system there must be a finite limit on the magnitude of earthquakes within that system. This quantity is not directly observable within the Groningen earthquake catalog, but we are still able to incorporate a prior distribution for $M_{\max}$ within the magnitude-frequency models. Following Cornell and Van Marke [1969], the survival function of a magnitude distribution may be truncated to reflect some prior belief in a maximum possible seismic moment, $\mathcal{M}_{\max}$, providing an upper bound to the distribution, according to

$$P(\geq \mathcal{M}|\mathcal{M}_m \leq \mathcal{M} \leq \mathcal{M}_{\max}) = \frac{P(\mathcal{M}) - P(\mathcal{M}_{\max})}{1 - P(\mathcal{M}_{\max})} \tag{35}$$

where the un-truncated survival function $P(\mathcal{M})$ is given by (11). In the case of Groningen induced seismicity, van Elk et al. [2017] reported a collective expert-judgment based prior discrete distribution of maximum possible magnitudes with a 3.75–7.25 range and a 4.8 median and 5.0 mean. For the most-part, the influence of the posterior distribution of exponential tapers occurs at significantly lower magnitudes than this prior distribution of $M_{\max}$.

For the data analyzed in this study, incorporating stress-dependent exponential tapering of the power-law seismic moment distribution alongside an upper bound in earthquake magnitude-frequency models used for probabilistic hazard and risk analysis of induced seismicity within the Groningen gas field reduces bias that may otherwise in this case over-state the hazard and risk. Utilizing data-driven, stress-dependent ζ-models also reduces the impact of imposing a maximum magnitude based on an expert judgment. This procedure is therefore more robust to possibility of expert bias.

12.3. Seismic Hazard Implications

Figure 32 illustrates the influence of including the possibility of a stress-dependent ζ model for induced earthquake magnitudes on probabilistic seismic hazard analysis. Even in the presence of a distribution of upper bounds to the magnitude distribution that starts at $M_{\max} = 4.3$, the implications of including an exponential taper that introduces deviations from the pure power-law for $M \geq 3$ events are still significant. The maximum seismic hazard is reduced by 31% from $0.174g$ to $0.130g$.

13. Conclusions

In summary, the stress-dependent ζ-model with constant β (M_{10}, M_{11}) offer higher performance magnitude-frequency forecasts than the stress-dependent β-models with $\zeta = 0$ (M_5, M_7) 75–85% of the time (Table 3) and lower the magnitude with a 10% and 1% chance of exceedance from 4.3 to 3.8 and from 5.5 to 4.5 respectively. Likewise, stress-dependent ζ-models outperform stress-invariant ζ-models (M_2, M_3) about 90% of the time, although in this case the stress-dependence of ζ increases the magnitude with 1% chance of exceedance from 3.9 to 4.4–4.5. The hybrid model with stress-dependent β and ζ values, M_{13}, includes all these possibilities in one joint posterior distribution, resulting in a 1% magnitude of 4.6, which is much closer to the stress-dependent ζ models than any of the other frequency-magnitude distributions (Figure 31).

There are two possibilities for incorporating these alternative magnitude-frequency models into a probabilistic seismic hazard and risk analysis.

1. Treat these model selection uncertainties as aleatory and rely on the M_{13} model alone to represent all possible models within the Monte Carlo simulations of induced seismicity, hazard and risk.

2. Treat model selection as an epistemic uncertainty and include each independent model class as different branches on a logic tree of alternative Monte Carlo simulations. To do this, we may use the evidence-based weight factors given by Table 3. So for the mutually independent and collectively exhaustive model set $\{M_1, M_2, M_7, M_{11}\}$ would

be represented by the logic tree branch weights $\{0.04, 0.08, 0.18, 0.7\}$. Given, the small weights attached to the first two models, and the need to limit the computational cost of probabilistic seismic hazard and risk assessments we recommend truncating this to the top two models $\{M_7, M_{11}\}$ with weights $\{0.2, 0.8\}$.

We recommend the second option, as it explicitly includes both the current seismological model used for hazard and risk analysis (M_7) and the new model (M_{11}) with its likely improved performance. Including both on the logic tree would allow the influence of each model on hazard and risk to be independently assessed. The new stress-dependent exponential-taper power-law model introduced here likely offers better forecast performance and better represents the physical processes of failure size distributions within a heterogeneous material under increasing stress. The limited sample size of Groningen earthquakes means we cannot be definitive in our preference for a single frequency-magnitude model. Instead, we represent our currently limited knowledge using a range of different models weighted by their measured performance evidence rather than expert judgment. Over time, further earthquake observations within Groningen, other analogue fields, or laboratory experiments may be decisive.

Appendix A: Poro-Elastic Thin-Sheet Model Inference

Figure 33 shows the marginal posterior distributions of the poro-elastic thin-sheet intra-reservoir stress model parameters defined in section 5.

Acknowledgments. We gratefully acknowledge our colleagues from Nederlandse Aardolie Maatschappij, Rob van Eijs, Onno van der Wal, Jan van Elk, Dirk Doornhof, Assaf Mar-Or, and from Shell Global Solutions International, Stijn Bierman, Phil Jonathan for their ongoing support during this study. The earthquake catalogue was provided by Koninklijk Nederlands Meteorologisch Instituut (KNMI). These data may be obtained from KNMI at www.knmi.nl. All other data were provided by Nederlandse Aardolie Maatschappij BV and are made available in an open access data repository [Bourne and Oates, 2017a]. The implementation of these seismological models made use of SciPy [Jones et al., 2001] and PyMC3 [Salvatier et al., 2015] and most of the figures were created using Matplotlib [Hunter, 2007]. Finally, the authors wish to thank Nederlandse Aardolie Maatschappij and Shell Global Solutions International for granting permission to publish this paper.

References

Aki, K., 1965. Maximum likelihood estimation of b in the formula logN=a-bM and its confidence limits. Bulletin of the Earthquake Research Institute of Tokyo University 43, 237–239.

Alava, M.J., Nukala, P.K.V.V., Zapperi, S., 2006. Statistical models of fracture. Advances in Physics 55, 349–476.

Amitrano, D., 2012. Variability in the power-law distributions of rupture events. The European Physical Journal Special Topics 205, 199–215.

Amorèse, D., Grasso, J.R., Rydelek, P.A., 2010. On varying b-values with depth: Results from computer-intensive tests for Southern California. Geophysical Journal International 180, 347–360.

de Arcangelis, L., Godano, C., Grasso, J.R., Lippiello, E., 2016. Statistical physics approach to earthquake occurrence and forecasting. Physics Reports 628, 1–91.

de Arcangelis, L., Redner, S., Herrmann, H., 2007. A random fuse model for breaking processes. Journal de Physique Lettres 46, 585–590.

Bachmann, C.E., Wiemer, S., Goertz-Allmann, B.P., Woessner, J., 2012. Influence of pore-pressure on the event-size distribution of induced earthquakes. Geophysical Research Letters 39.

Bak, P., Tang, C., 1989. Earthquakes as a Self-Organized Critical Phenomenon. Journal of Geophysical Research 94, 15635–15637.

Bender, B., 1983. Maximum likelihood estimation of b values grouped data. Bulletin of the Selsmologmal Socmty of America 73, 831–851.

Bonamy, D., Bouchaud, E., 2011. Failure of heterogeneous materials: A dynamic phase transition? Physics Reports 498, 1–44.

Bourne, S., Oates, S., 2017a. Induced Seismicity Within The Groningen Gas Field [Data set]. https://doi.org/10.5281/zenodo.1035226 .

Bourne, S.J., Oates, S.J., 2017b. 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, 10,299–10,320.

Bourne, S.J., Oates, S.J., Bommer, J.J., Dost, B., Elk, J.V., Doornhof, D., 2015. A Monte Carlo Method for Probabilistic Hazard Assessment of Induced Seismicity due to Conventional Natural Gas Production. Bull. Seis. Soc. Am. 105.

Bourne, S.J., Oates, S.J., van Elk, J., Doornhof, D., 2014. A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir. Journal of Geophysical Research: Solid Earth 119, 8991–9015.

Bourne, S.J., Oates, S.J., Elk, J.V., 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, 1693–1700.

Burridge, R., Knopoff, L., 1967. Model and theoretical seismicity. Bulletin of the Seismological Society of America 57, 341–371.

Coniglio, A., Klein, W., 1980. Clusters and Ising critical droplets: A renormalisation group approach. Journal of Physics A: Mathematical and General 13, 2775–2780.

Cornell, C.A., Van Marke, E.H., 1969. The major influences on seismic risk, in: Proc. 4th World Conf. on Earthquake Engineering, Santiago, Chile, pp. A–1, 69–93.

Daguier, P., Nghiem, B., Bouchaud, E., Creuzet, F., 1997. Pinning and Depinning of Crack Fronts in Heterogeneous Materials. Phys. Rev. Lett. 78.

Davies, R., Foulger, G., Binley, A., Styles, P., 2013. Induced seismicity and hydraulic fracturing for the recovery of hydrocarbons. Marine and Petroleum Geology 45, 171–185.

Dost, B., Goutbeek, F., Eck, T.V., Kraaijpoel, D., 2012. Monitoring induced seismicity in the North of the Netherlands: status report 2010. Technical Report. Koninklijk Nederlands Meteorologisch. de Bilt, The Netherlands.

van Elk, J., Mar-Or, A., Doornhof, D., 2017. Assessment of Building Damage based on Production Scenario “Basispad Kabinet” for the Groningen field: Addendum to ”Induced Seismicity in Groningen Assessment of Hazard, Building Damage and Risk”. Technical Report November. Nederlandse Aardolie Maatschappij B.V.. Assen, The Netherlands.

Elk, J.V., Bourne, S.J., Oates, S.J., Bommer, J.J., 2019. A Probabilistic Seismic Risk Model to Inform Decision-Making in Response to Induced Earthquakes in the Groningen Gas Field. Earthquake Spectra 35, 537–564.

Ellsworth, W., 2013. Injection-induced earthquakes. Science 341, 250–260.

Evans, K., Zappone, A., Kraft, T., Deichmann, N., Moia, F., 2012. A survey of the induced seismic responses to fluid injection in geothermal and CO2 reservoirs in Europe. Geothermics 41, 30–54.

Foulger, G.R., Wilson, M.P., Gluyas, J.G., Julian, B.R., Davies, R.J., 2018. Global review of human-induced earthquakes. Earth-Science Reviews 178, 438–514.

Frohlich, C., Davis, S., 1993. Teleseismic b Values; Or, Much Ado About 1.0. Journal of Geophysical Research 98, 631–644.

Gilabert, A., Vanneste, C., Sornette, D., Guyon, E., 2007. The random fuse network as a model of rupture in a disordered medium. Journal de Physique 48, 763–770.

Gulia, L., Wiemer, S., 2010. The influence of tectonic regimes on the earthquake size distribution: A case study for Italy. Geophysical Research Letters 37, n/a–n/a.

Gutenberg, B., Richter, C., 1954. Seismicity of the Earth and Associated Phenomena. Princeton University Press, Princeton, New Jersey. 2nd edition.

Hansen, A., 2011. Random Fuse Networks: A Review. Physical Aspects of Fracture , 59–72.

Hansen, A., Hemmer, P.C., 1994. Burst avalanches in bundles of fibers: Local versus global load-sharing. Physics Letters A 184, 394–396.

Harris, C.K., Bourne, S.J., 2017. Computing the Distribution of Pareto Sums using Laplace Transformation and Stehfest Inversion. Pure and Applied Geophysics .

Hemmer, P.C., Hansen, A., 1992. The Distribution of Simultaneous Fiber Failures in Fiber Bundles. Journal of Applied Mechanics 59, 909.

Huang, H.H., Huang, T.C., Wu, Y.M., Koulakov, I., Chen, S.K., Chao, W.A., 2018. Relationship Between Earthquake b -Values and Crustal Stresses in a Young Orogenic Belt. Geophysical Research Letters 45, 1832–1837.

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. Computing In Science & Engineering 9, 90–95.

IEAGHG, 2013. Induced seismicity and its implications for CO2 storage risk. Technical Report June. International Energy Agency Environmental Projects Ltd. (IEAGHG).

Cheltenham, UK.

Jiang, X., Liu, H., Main, I.G., Salje, E.K., 2017. Predicting mining collapse: Superjerks and the appearance of record-breaking events in coal as collapse precursors. Physical Review E 96, 1–7.

Jones, E., Oliphant, T., Peterson, P., 2001. SciPy: Open source scientific tools for Python. http://www.scipy.org/.

J.P. Sethna, K.A. Dahmen, C.R. Myers, 2001. Crackling noise: Preprint. Nature 410, 242–250.

Kagan, Y.Y., 1999. Universality of the seismic moment-frequency relation. Pure and Applied Geophysics 155, 537–573.

Kagan, Y.Y., 2002a. Seismic moment distribution revisited : II. Moment conservation principle. Geophys. J. Int 149, 731–754.

Kagan, Y.Y., 2002b. Seismic moment distribution revisited: I. Statistical results. Geophysical Journal International 148, 520–541.

Kagan, Y.Y., 2010. Earthquake size distribution: Power-law with exponent β=1/2? Tectonophysics 490, 103–114.

Kamer, Y., Hiemer, S., 2015. Data-driven spatial b value estimation with applications to California seismicity: To b or not to b. J. Geophys. Res. Solid Earth 120,, 5191–5214,.

Klose, C.D., 2013. Mechanical and statistical evidence of the causality of human-made mass shifts on the Earth's upper crust and the occurrence of earthquakes. Journal of Seismology , 109–135.

Kloster, M., Hansen, A., Hemmer, P.C., 1997. Burst avalanches in solvable models of fibrous materials. Physical Review E - Statistical Physics, Plasmas, Fluids, and Related

Interdisciplinary Topics 56, 2615–2625.

Langenbruch, C., Zoback, M.D., 2016. How will induced seismicity in Oklahoma respond to decreased saltwater injection rates? Science Advances 2, 1–9.

Main, I., Sammonds, P., Meredith P.R., Jones, C., 1990. Influence of fractal flaw distributions on rock deformation in the brittle field, in: Knipe, R. J. & Rutter, E.H. (Ed.), Deformation Mechanisms, Rheology and Tectonics,. Geological Society Special Publication No. 54, pp. 71–79.

Main, I.G., 1995. Earthquakes as Critical Phenomena: Implications for Probabilistic Seismic Hazard Analysis 85, 1299–1308.

Main, I.G., 1996. Statistical Physics, Seismogenesis and Seismic Hazard. Reviews of Geophysics , 433–462.

Main, I.G., Burton, P.W., 1984. Information theory and the earthquake frequency-magnitude distribution. Ekp 74, 1409–1426.

Main, I.G., Meredith, P.G., Jones, C., 1989. A reinterpretation of the precursory seismic b-value anomaly from fracture mechanics. Geophysical Journal International 96, 131–138.

Main, I.G., Meredith, P.G., Sammonds, P.R., 1992. Temporal variations in seismic event rate and b-values from stress corrosion constitutive laws. Tectonophysics 211, 233–246.

Majer, E., Baria, R., Stark, M., Oates, S., Bommer, J., Asanuma, H., 2007. Induced seismicity associated with enhanced geothermal systems. Geothermics 36, 185–222.

Mogi, K., 1962. Magnitude–frequency relation for elastic shocks accompanying fractures of various materials and some related problems in earthquakes. Bull. Earthquake Res. Inst. Tokyo Univ. 40, 831–853.

Mori, J., Abercrombie, R., 1997. Depth dependence of earthquake frequency-magnitude distributions in California: Implications for rupture initiation. Journal of Geophysical Research 102, 15,081–15,090.

NAS, 2013. Induced seismicity potential in energy technologies. National Academy of Sciences, Washington, DC, USA.

Nukala, P.K.V., Zapperi, S., Åimunović, S., 2005. Statistical properties of fracture in a random spring model. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 71, 1–11.

Petersen, M.D., Mueller, C.S., Moschetti, M.P., Hoover, S.M., Rukstales, K.S., McNamara, D.E., Williams, R.A., Shumway, A.M., Powers, P.M., Earle, P.S., Llenos, A.L., Michael, A.J., Rubinstein, J.L., Norbeck, J.H., Cochran, E.S., 2018. 2018 one-year seismic hazard forecast for the central and eastern United States from induced and natural earthquakes. Seismological Research Letters 89, 1049–1061.

Ponson, L., Bonamy, D., Bouchaud, E., 2006. Two-Dimensional Scaling Properties of Experimental Fracture Surfaces. Phys. Rev. Lett. 96, 035506.

Pradhan, S., 2010. Failure Processes in Elastic Fiber Bundles. Rev. Mod. Phys. 82, 449–. arXiv:0808.1375v3.

Roberts, N.S., Bell, A.F., Main, I.G., 2016. Mode switching in volcanic seismicity: El Hierro 2011-2013. Geophysical Research Letters 43, 4288–4296.

Roux, S., Hansen, A., Herrmann, H., Guyon, E., 1988. Rupture of heterogeneous media in the limit of infinite disorder. Journal of Statistical Physics 52, 237–244.

Salvatier, J., Wiecki, T., Fonnesbeck, C., 2015. Probabilistic Programming in Python using PyMC , 1–241507.08050.

Scholz, C.H., 1968. The frequency - magnitude relation of microfracturing in rock and its relation to earthquakes. Bull. Sesmological Soc. Am. 58, 399–415.

Schorlemmer, D., Wiemer, S., Wyss, M., 2005. Variations in earthquake-size distribution across different stress regimes. Nature 437, 539–542.

Shapiro, S., Dinske, C., Langenbruch, C., 2010a. Seismogenic index and magnitude probability of earthquakes. The Leading Edge March.

Shapiro, S., Dinske, C., Langenbruch, C., 2010b. Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations. The Leading Edge .

Shapiro, S.A., 2018. Seismogenic Index of Underground Fluid Injections and Productions. Journal of Geophysical Research: Solid Earth 123, 7983–7997.

Shekhawat, A., Zapperi, S., Sethna, J.P., 2013. From damage percolation to crack nucleation through finite size criticality. Physical Review Letters 110, 1–5.

Shi, Y., Bolt, B.A., 1982. The standard error of the magnitude-frequency b-value. Bulletin of the Seismological Society of America 72, 1677–1687.

Sornette, D., Physique, J.D., 1992. Mean-field solution of a block-spring model of earthquakes. Journal de Physique I, EDP Sciences 11, 2089–2096.

Spada, M., Tormann, T., Wiemer, S., Enescu, B., 2013. Generic dependence of the frequency-size distribution of earthquakes on depth and its relation to the strength profile of the crust. Geophysical Research Letters 40, 709–714.

Stauble, A., Milius, G., 1970. Geology of Groningen Gas Field, Netherlands, in: Halbouty, M. (Ed.), Geology of giant petroleum fields. American Association of Petroleum Geologists, Tulsa, Memoir 14 edition. pp. 359–369.

Stauffer, D., Aharony, A., 1994. Introduction To Percolation Theory. Taylor & Francis, London. second edition.

Suckale, J., 2009. Induced Seismicity in Hydrocarbon Fields. Advances in Geophysics 51, 55–106.

Tinti, S., Mulargia, F., 1987. Confidence intervals of b values for grouped magnitudes. Bull. Seismol. Soc. Am. 77, 2125–2134.

Tormann, T., Wiemer, S., Mignan, A., 2014. Journal of Geophysical Research : Solid Earth Systematic survey of high-resolution b value imaging along Californian faults : Inference on asperities. Journal of Geophysical Research: Solid Earth 119, 2029–2054.

Toussaint, R., Hansen, A., 2006. Mean-field theory of localization in a fuse model. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 73.

Toussaint, R., Pride, S.R., 2005. Interacting damage models mapped onto Ising and percolation models. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 71.

Turcotte, D., 1997. Fractals and Chaos in Geology and Geophysics. Cambridge University Press, Cambridge, United Kingdom. second edition.

Utsu, T., 1965. A method for determining the value of b in a formula log n = a - bM showing the magnitude-frequency relation for earthquakes. Geophys. Bull. Hokkaido Univ. 13, 103.

Vasseur, J., Wadsworth, F.B., Heap, M.J., Main, I.G., Lavallée, Y., Dingwell, D.B., 2017. Does an inter-flaw length control the accuracy of rupture forecasting in geological materials? Earth and Planetary Science Letters 475, 181–189.

Vasseur, J., Wadsworth, F.B., Lavallée, Y., Bell, A.F., Main, I.G., Dingwell, D.B., 2015. Heterogeneity: The key to failure forecasting. Scientific Reports 5, 1–7.

Weibull, W., 1939. A statistical theory of the strength of materials. Ingvetensk. Akad. Handl. 149.

Wiemer, S., Wyss, M., 1997. Mapping the frequency-magnitude distribution in asperities: An improved technique to calculate recurrence times? Journal of Geophysical Research: Solid Earth 102, 15115–15128.

Wiemer, S., Wyss, M., 2002. Mapping spatial variability of the frequency-magnitude distribution of earthquakes. Advances in Geophysics 45.

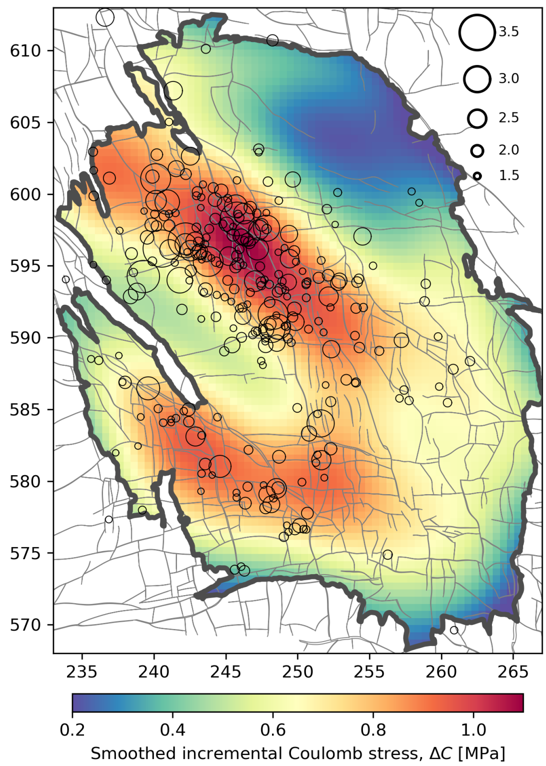


Figure 1. Observed distribution of epicenters and magnitudes of earthquakes induced by Groningen gas production since the monitoring of $M \geq 1.5$ events started in 1995. Colors denote the poro-elastic thin-sheet model of smoothed incremental maximum Coulomb within the reservoir induced by pore-pressure depletion from the start of production in 1965 until 2019. Circle denotes earthquakes and their area scales continuously with earthquake magnitude as indicate by the legend. Thin gray lines denote fault traces at the top of the reservoir. A dark gray polygon denotes the original gas-water contact. Map coordinate units are kilometers. Epicenter location errors are about 500 m and magnitude errors are about 0.1.

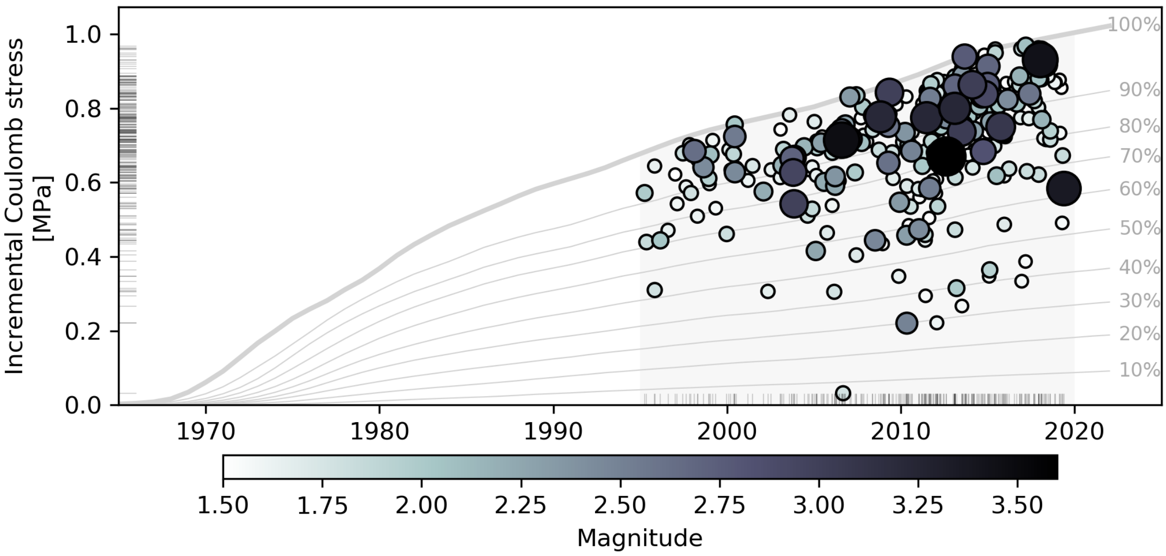


Figure 2. Time series of the incremental Coulomb stress and magnitude associated with the Groningen $M \geq 1.5$ events observed from 1995 to 2019 (grey panel). The area and colour of the circles denotes the magnitude of each event. Grey lines indicate the evolution of stress exposure within the reservoir according to the poro-elastic thin-sheet model and denote the reservoir volume fraction exposed to at most that stress state. Most events occur within the largest 20% of the exposed stress states (80%–100%).

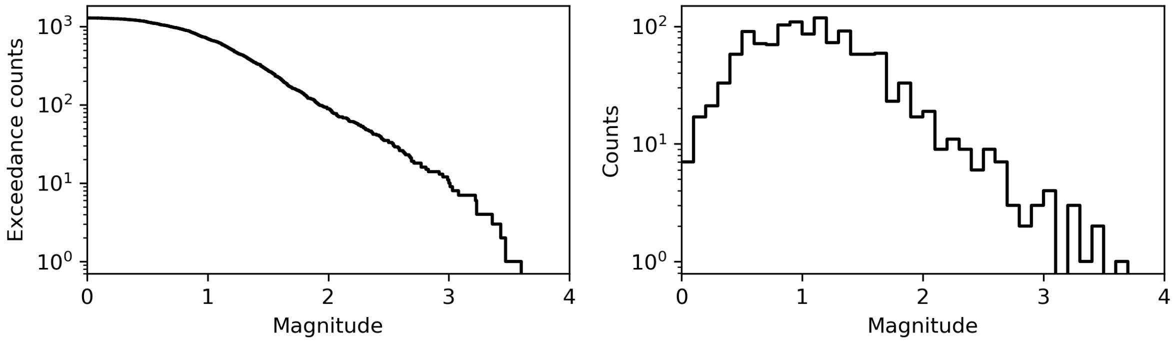


Figure 3. The frequency magnitude distribution of earthquakes from 1/1/1995 to 1/6/2019 associated with Groningen gas production. (left) Exceedance counts denotes the number of events with at least the given magnitude. (right) Counts denotes the number of events within magnitude bins of size 0.1. The magnitude of completeness for this catalog is in the range 1.3–1.5.

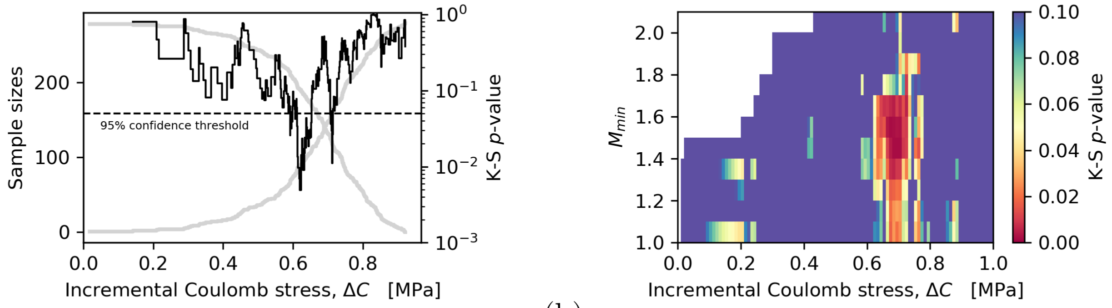


Figure 4. (a) Kolmogorov-Smirnov test statistic p-values for each possible division of the observed $M \geq 1.5$ events since April 1995 into low and high incremental Coulomb stress, ΔC, groups according to the poro-elastic thin-sheet stress model ($\sigma = 3.5$ km, $r_{\max} = 1.12$, $H_s = 10^{13}$ Pa). The i^{th} event belongs to the low stress group if $\Delta C_i < \Delta C$; otherwise it belongs to the high stress group. (b) As (a), except for all $M \geq M_{\min}$ events over the range $1.0 \leq M_{\min} \leq 2.0$.

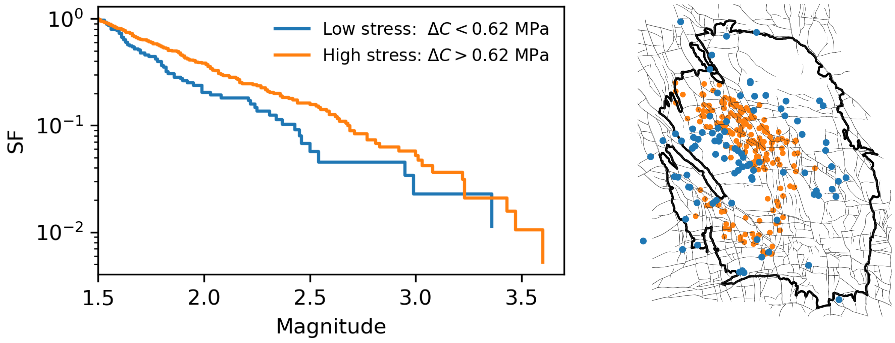


Figure 5. The observed earthquake size distribution and epicentral locations of low- and -high stress groups most likely to originate from different probability distributions.

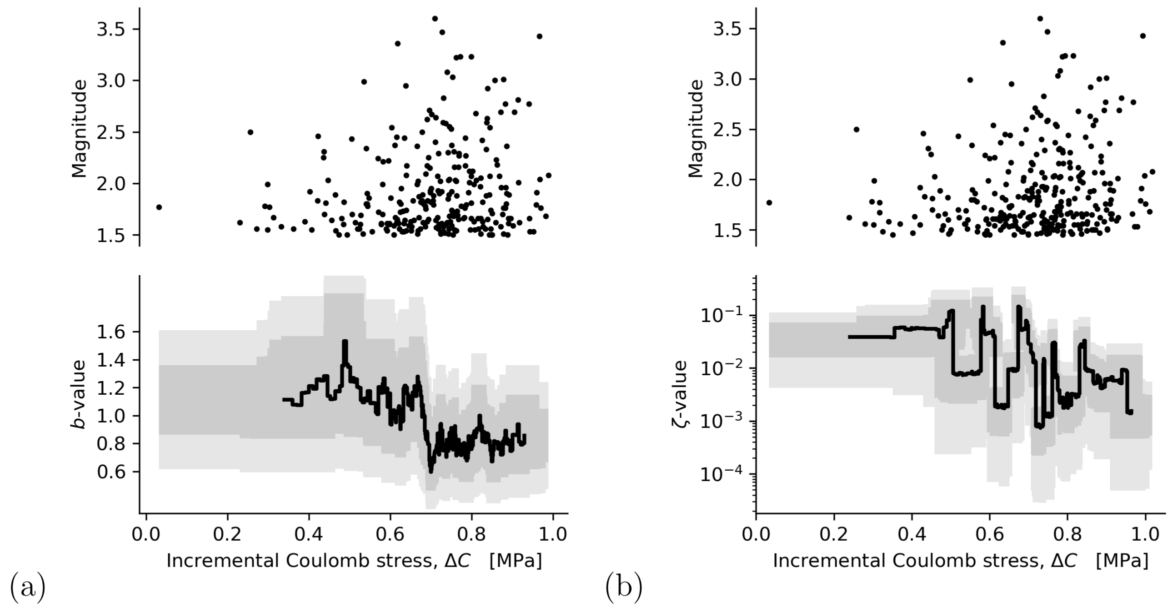


Figure 6. Variation of posterior (a) b-value and (b) ζ-value estimates with incremental Coulomb stress given $M_{\min} = 1.5$ and a constant population sample of 20 events. Light and dark gray bands denote the 67% and 95% confidence intervals, and $\beta = \frac{2}{3}b$.

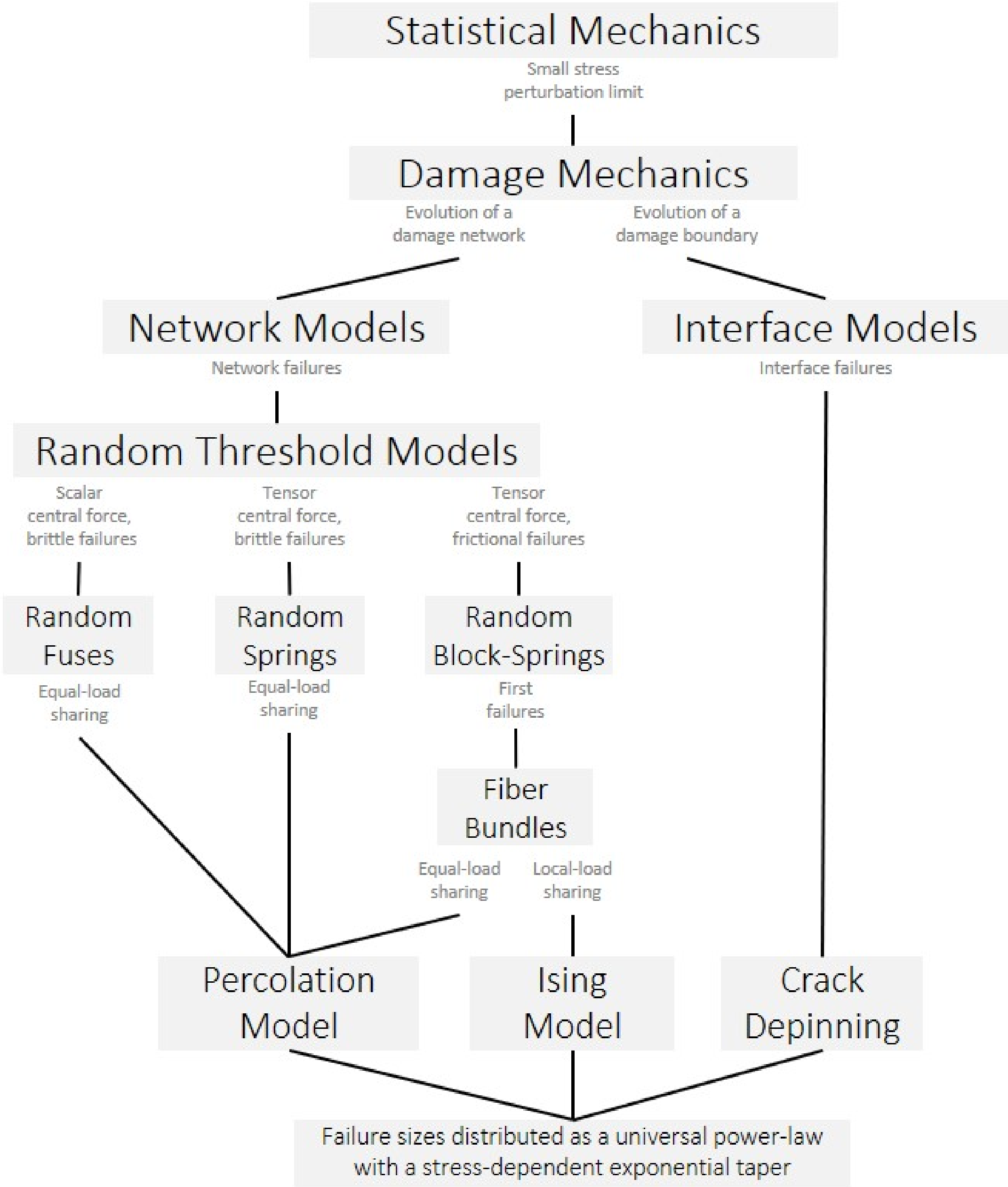


Figure 7. The network of statistical damage mechanics theories that seek to describe mechanical failure as a stochastic process. These different models all lead to failure sizes distributed according to a stress-invariant power-law with a stress-dependent exponential-like taper.

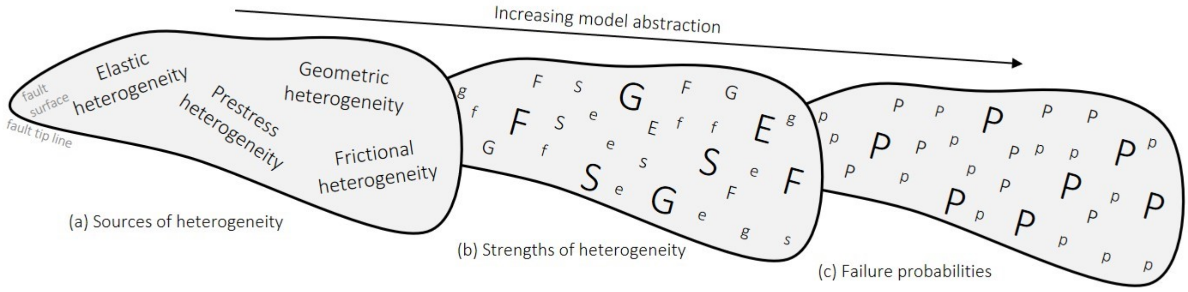


Figure 8. Schematic to illustrate (a) the different sources and (b) the different strengths of unresolved fault heterogeneity and (c) their stochastic representation as local failure probabilities that lead to the emergence of an exponentially tapered power-law distribution of failure sizes.

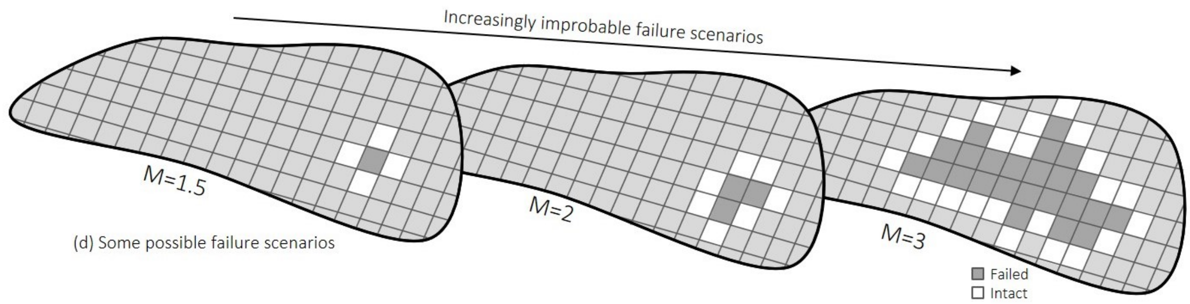


Figure 9. The probability-area distribution of fault failure events arising on a single fault with a network of uniform failure probabilities is an exponentially-tapered power-law. The lower probabilities of larger failure areas are governed by competition between the lower probability of a larger number of connected dark gray failed cells bordered by white intact cells and the larger number of alternative geometric configurations with the same failure area.

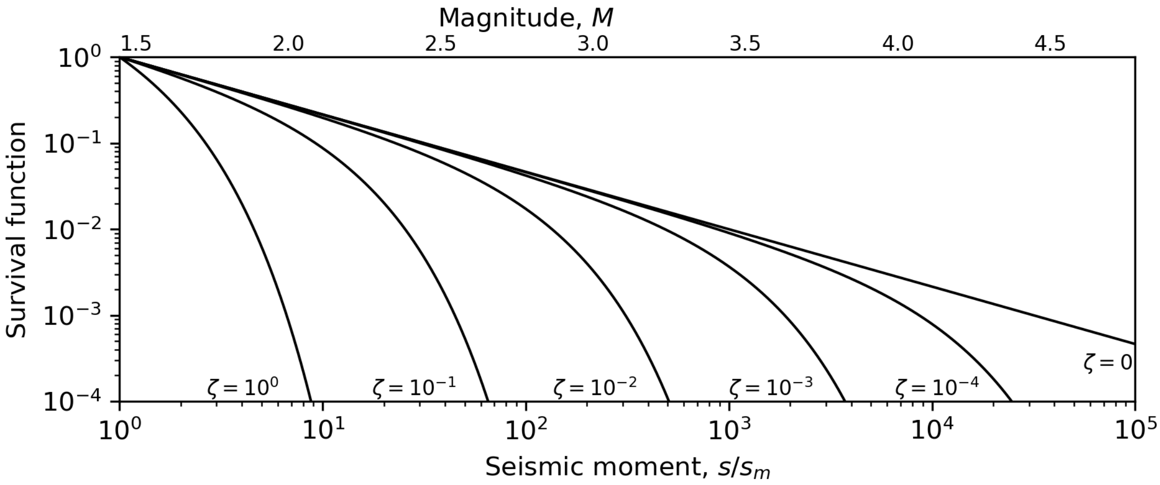
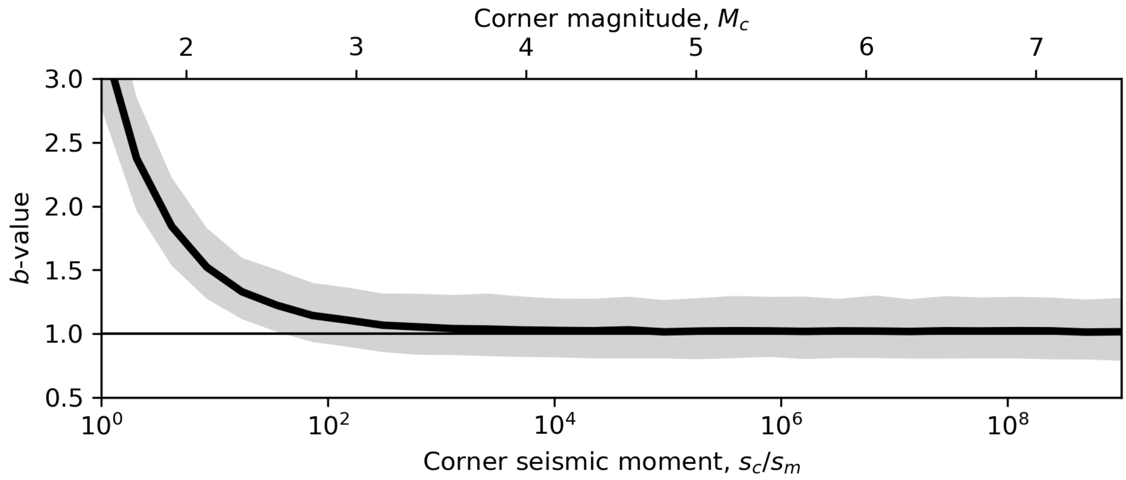
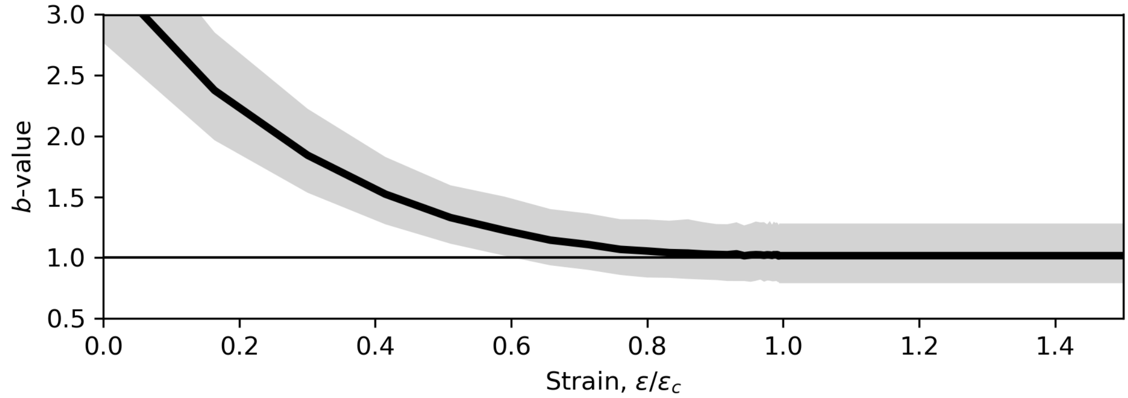


Figure 10. Seismic moment exceedance probability functions (survival functions, SF) as a power-law with an exponential cut-off according to according to equation (9) for a constant power-law exponent $\beta = \frac{2}{3}$, and ζ-values, varying from 1 to 0, where $\mathcal{M}_m = 2.2 \times 10^{11}$ Nm ($M_{\min}$=1.5).



(a)



(b)

Figure 11. (a) Apparent decrease in b-value with increasing corner moment, $\mathcal{M}_c$. Based on 1000 simulated earthquake catalogues each with 50 $M \geq 1.5$ events for $\beta = \frac{2}{3}$ and a given $\mathcal{M}_c$, and then repeated for 30 different values of $\mathcal{M}_c$. The black line and grey band denote the ensemble average and 5% to 95% interval of these simulations. (b) Apparent decrease in b-value with increasing strain, ϵ relative to a critical strain, ϵ_c, according to a critical-point scaling law, $\mathcal{M}_c \sim (\epsilon_c - \epsilon)^{-\gamma}$. In this example, $\gamma = 2$.

$\mathrm{M_i}$	Equations	Parameters	Label
Stress invariant models			
M_1	(19)	$\boldsymbol{\Theta}_1 = \{\beta\}$	uni1
M_2	(20)	$\boldsymbol{\Theta}_2 = \{\zeta\}$	uni1.z
M_3	(18)	$\boldsymbol{\Theta}_3 = \{\beta, \zeta\}$	uni2
Stress-dependent, inverse-power-law β-models			
M_4	(21)	$\boldsymbol{\Theta}_4 = \{\theta_0, \theta_1, \theta_2, \theta_3\}$	ets0.ipc4
M_5	$(21 \vert \theta_1{=}0)$	$\boldsymbol{\Theta}_5 = \{\theta_0, \theta_2, \theta_3\}$	ets0.ipc3
Stress-dependent, hyperbolic tangent β-models			
M_6	(22)	$\boldsymbol{\Theta}_6 = \{\theta_0, \theta_1, \theta_2, \theta_3\}$	ets0.htc4
M_7	$(22 \vert \theta_3{=}0)$	$\boldsymbol{\Theta}_7 = \{\theta_0, \theta_1, \theta_2\}$	ets0.htc3
M_8	(4), $(22 \vert \theta_3{=}0)$	$\boldsymbol{\Theta}_8 = \{\beta_2, \beta_3, \beta_4, \theta_0, \theta_1, \theta_2\}$	ets3.htc3
Stress-dependent, critical-point scaling ζ-models			
M_9	(23)	$\boldsymbol{\Theta}_9 = \{\theta_0, \theta_1, \theta_2, \theta_3\}$	ets0.cps4
M_{10}	$(23 \vert \theta_1{=}10^{-4})$	$\boldsymbol{\Theta}_{10} = \{\theta_0, \theta_2, \theta_3\}$	ets0.cps3
Stress-dependent, exponential trend ζ-models			
M_{11}	(24)	$\boldsymbol{\Theta}_{11} = \{\theta_0, \theta_1, \theta_2\}$	ets0.ltc3
M_{12}	(4), (24)	$\boldsymbol{\Theta}_{12} = \{\beta_2, \beta_3, \beta_4, \theta_0, \theta_1, \theta_2\}$	ets3.ltc3
Stress-dependent β-ζ-models			
M_{13}	(25)	$\boldsymbol{\Theta}_{13} = \{\theta_0, \theta_1, \theta_2, \theta_3, \theta_4\}$	ets0.b3.z2
M_{14}	(4), (25)	$\boldsymbol{\Theta}_{14} = \{\beta_2, \beta_3, \beta_4, \theta_0, \theta_1, \theta_2, \theta_3, \theta_4\}$	ets3.b3.z2

Table 1. Summary of the different seismological magnitude-frequency models, M_i, evaluated according to their posterior, out-of-sample, predictive performance. The labels are composed by string to represent a model type and a following digit to denote the associated degree of freedom. For instance, ets0.htc3 denotes the MAP elastic thin-sheet model with zero degrees of freedom combined with the hyperbolic tangent of incremental Coulomb stress model with 3 degrees of freedom to represent a stress-dependent β-value.

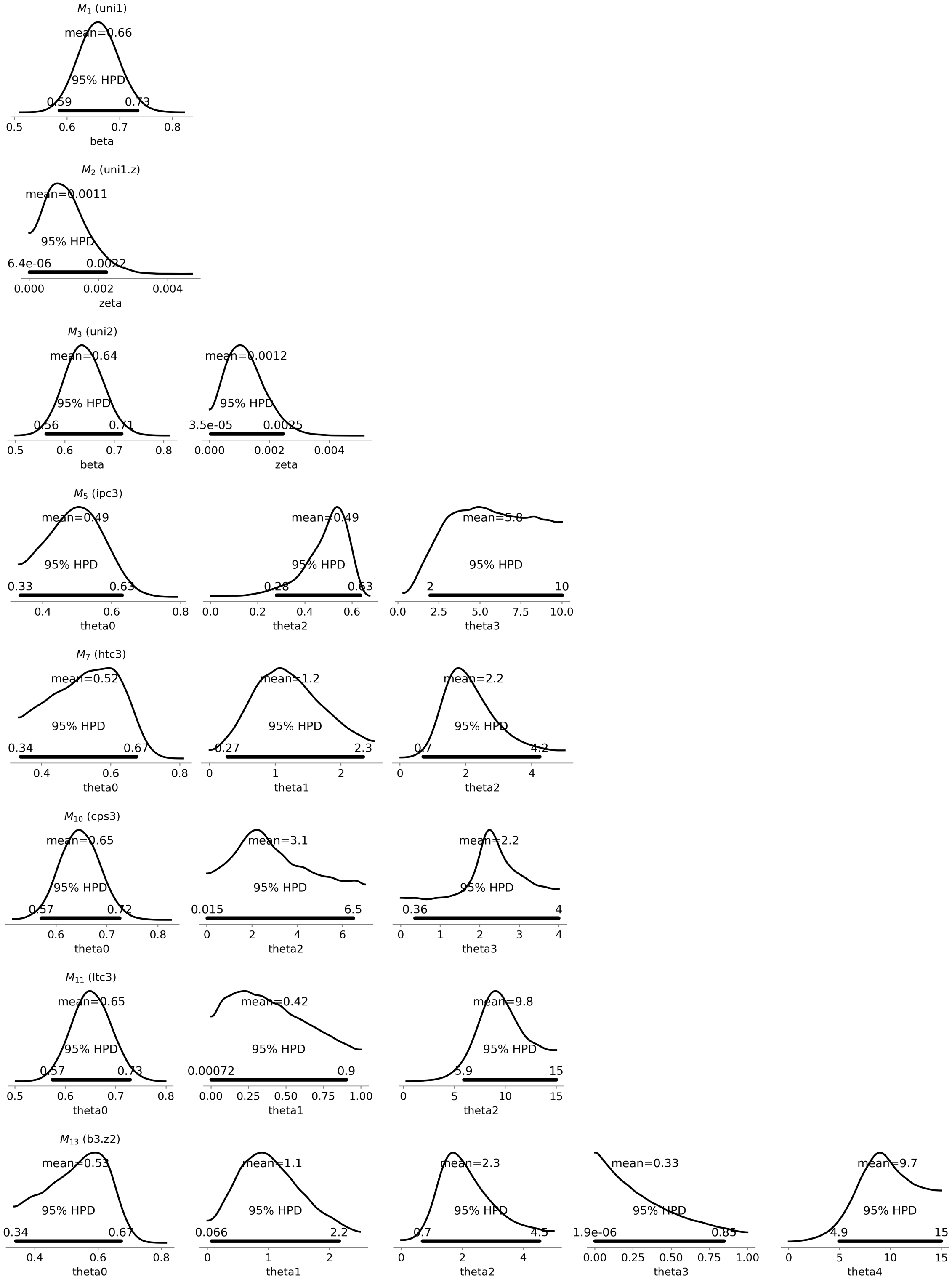


Figure 12. Marginal posterior probability density distributions inferred for each model given the observed magnitudes of $M \geq 1.5$ events from 1995 to 2019. Thick horizontal lines denote the 95% credible interval defined by the highest posterior density (HPD) interval.

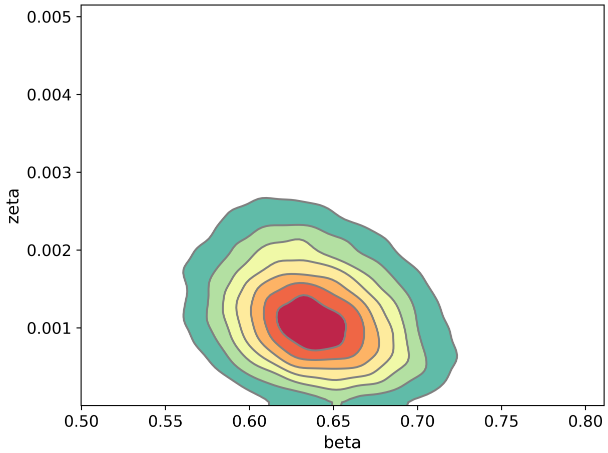


Figure 13. The joint posterior distribution of the stress-invariant β- and ζ-model (M_3, uni2) obtained given the observed magnitudes of $M \geq 1.5$ events from 1995 to 2019. These sampled distributions are represented by Gaussian kernel densities that introduce some data-adaptive smoothing.

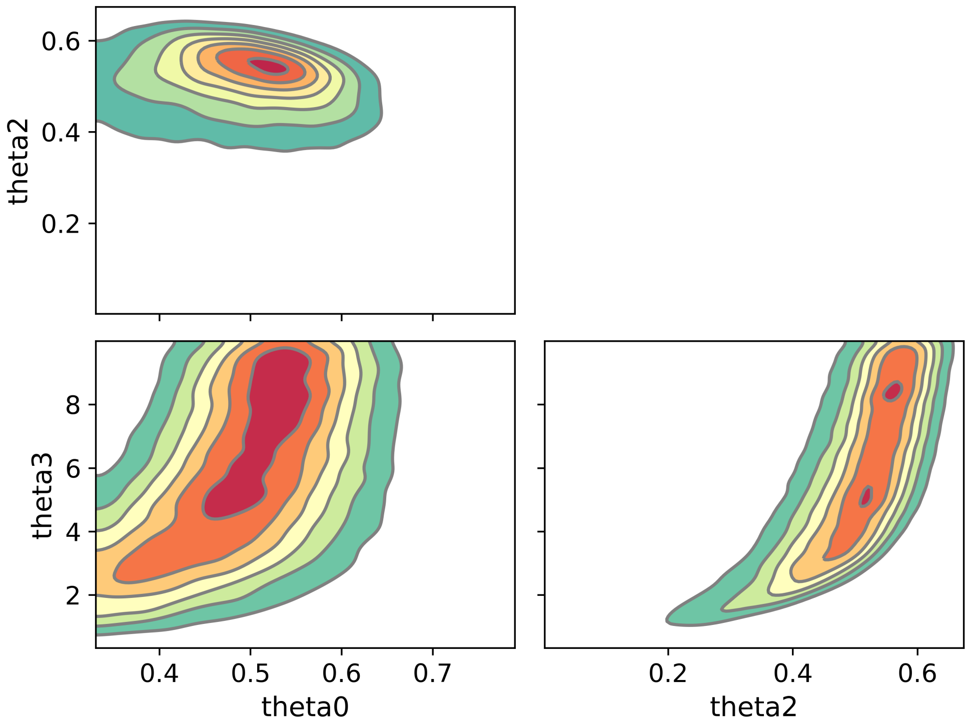


Figure 14. Pairwise joint posterior distributions of the inverse power-law β-model parameters given (M_5, ets0.ipc3) inferred given $\theta_1 = 0$ and the MAP poro-elastic thin-sheet model ($\sigma = 3.5$ km, $r_{\max} = 1.1$, $H_s = 10^7$ MPa) and the observed catalogue of $M \geq 1.5$ earthquakes from 1-Jan-1995 to 1-Jan-2019. There is stronger evidence of covariance given the less-than-circular joint density maps.

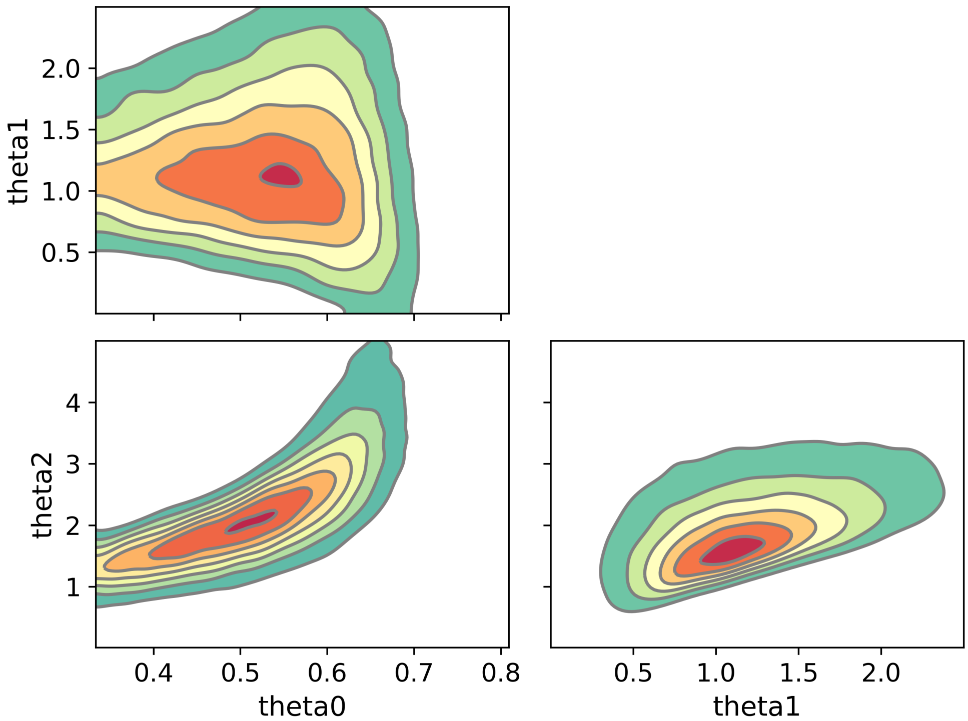


Figure 15. Pairwise joint posterior distributions of the hyperbolic tangent β-model parameters (M_7, ets0.htc3) inferred given $\theta_3 = 0$ and the MAP poro-elastic thin-sheet model ($\sigma = 3.5$ km, $r_{\max} = 1.1$, $H_s = 10^7$ MPa) and the observed catalogue of $M \geq 1.5$ earthquakes from 1-Jan-1995 to 1-Jan-2019.

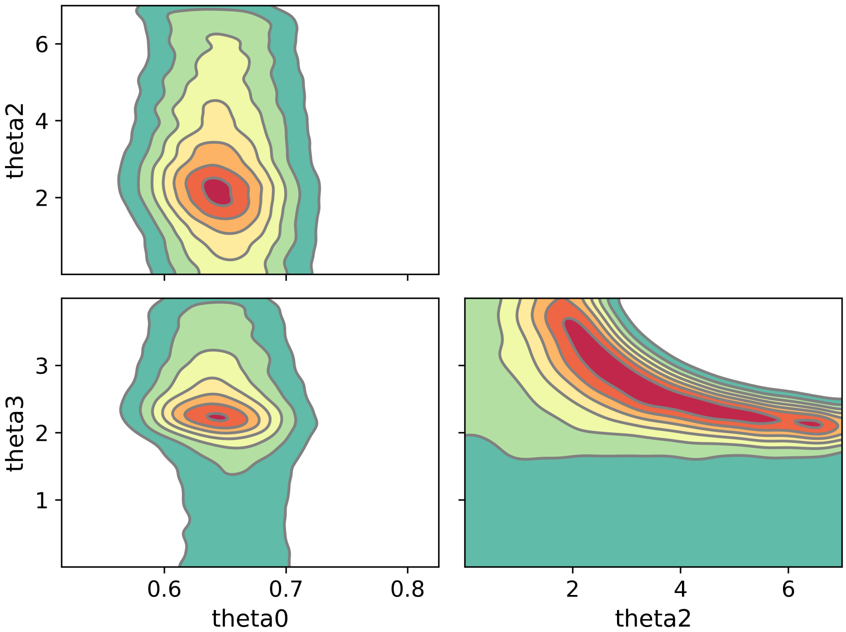


Figure 16. Pairwise joint posterior distributions of inverse power-law stress-dependent ζ-model parameters (M_{10}, ets0.cps3) defined according to (23) given the additional constraint $\theta_1 = 0$.

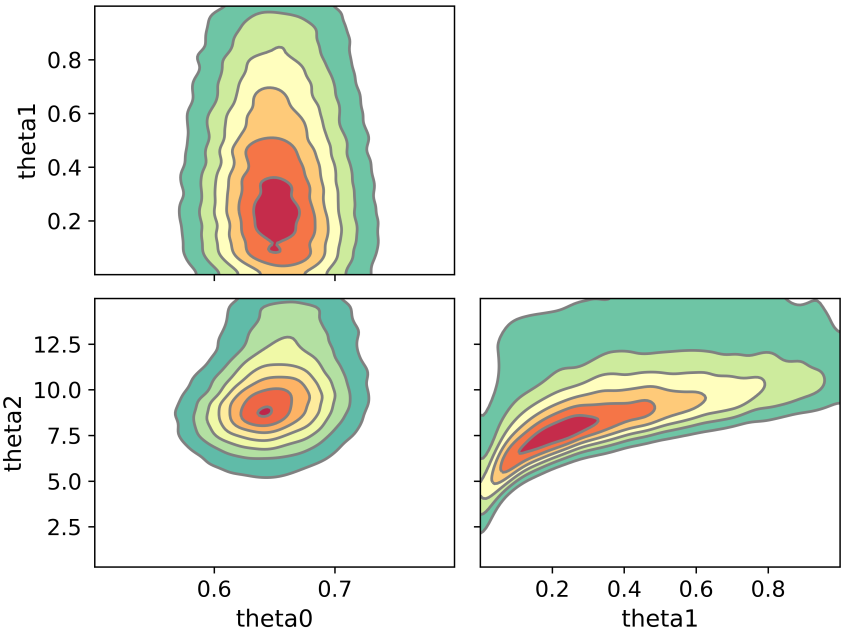


Figure 17. Pairwise joint posterior distribution of the exponential stress dependent ζ-model parameters (M_{11}, ets0.ltc3) defined according to (24).

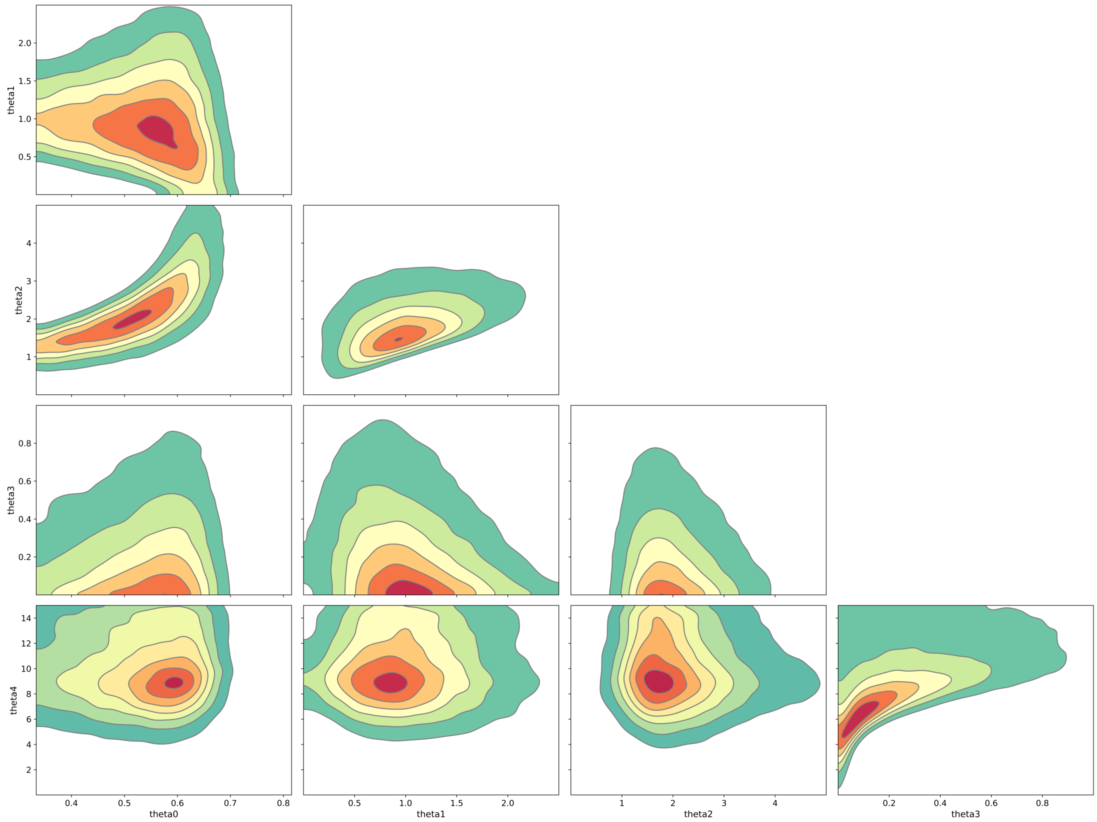


Figure 18. Pairwise joint posterior distributions of the stress-dependent β-ζ-model parameters (M_{13}, ets0.b3.z2) defined according to (25).

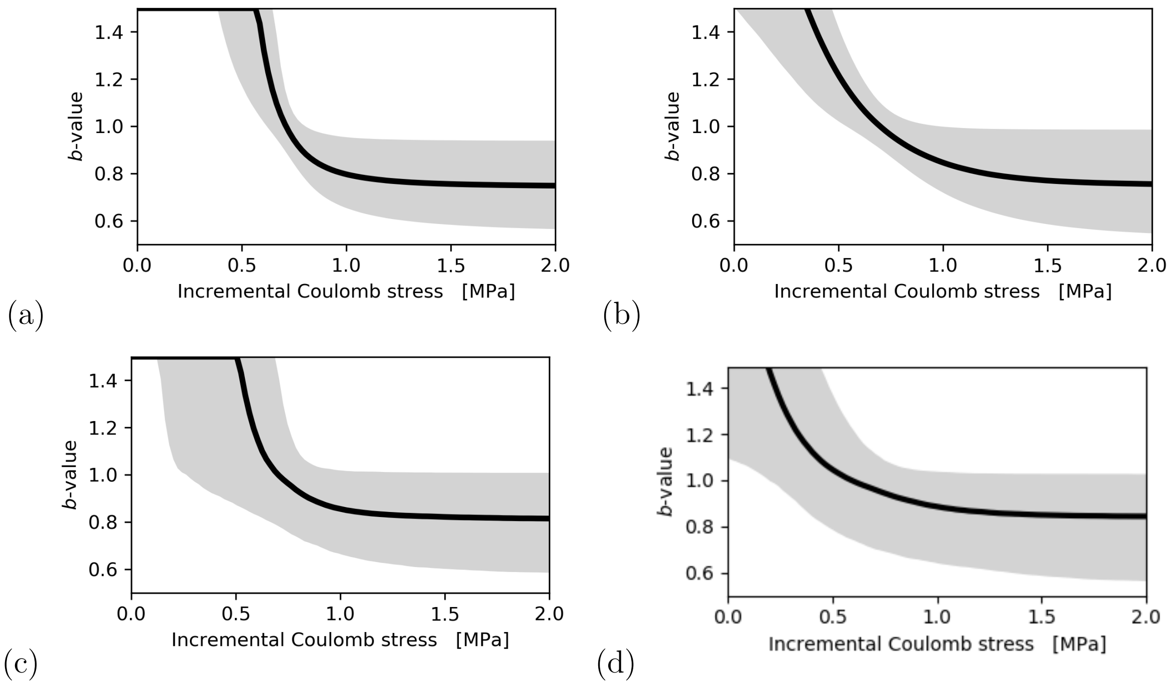


Figure 19. Posterior ensemble b-value functions of incremental Coulomb stress according to (a) the inverse power-law (M_5, ets0.ipc3), and (b) the hyperbolic tangent (M_7, ets0.htc3) model distributions shown by Figures 14 and 15 respectively. (c) As (a), except including the full posterior distribution of pore-elastic thin-sheet parameters, σ, $r_{\max}$, H_s (ets3.ipc3). (d) As (c), except for the hyperbolic tangent model (M_8, ets3.ipc3). Note that b-values are shown instead of β-values, where $b = 1.5\beta$. Black curves and grey shading denote the median and 95% prediction intervals respectively.

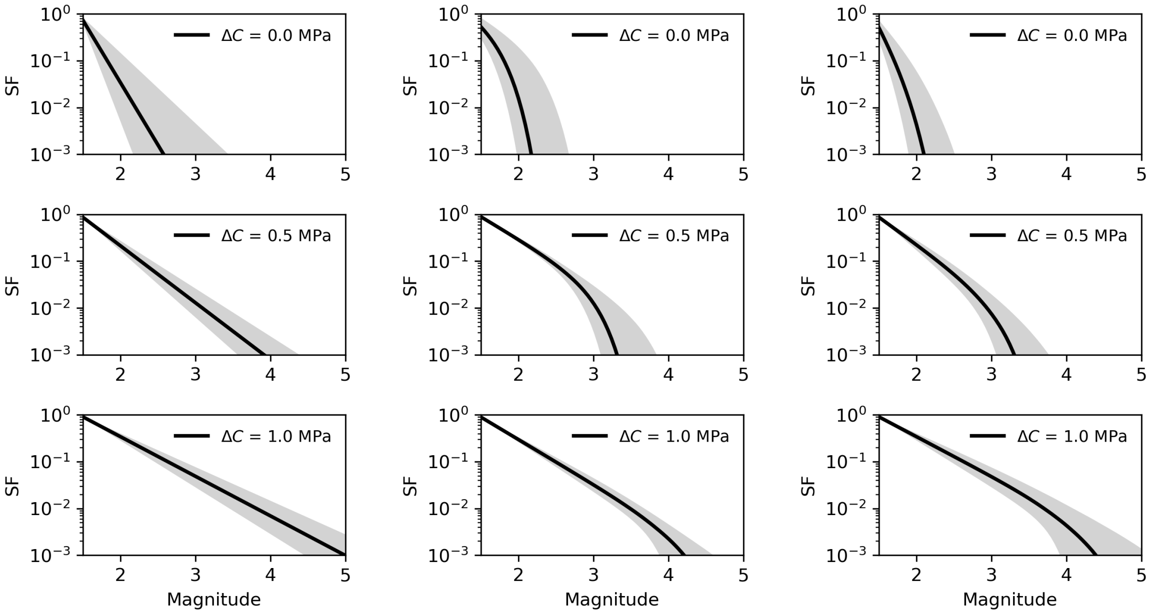


Figure 20. Evolution of the modelled magnitude survival function, SF, illustrated by three stress states, $\Delta C = 0, 0.5, 1$ MPa for a stress-dependent β-model (left, M_7), stress-dependent ζ-model (middle, M_{11}), and the stress dependent β-ζ-model (right, M_{13}). These ensemble functions are summarized by their medians (black), and 95% prediction intervals (grey).

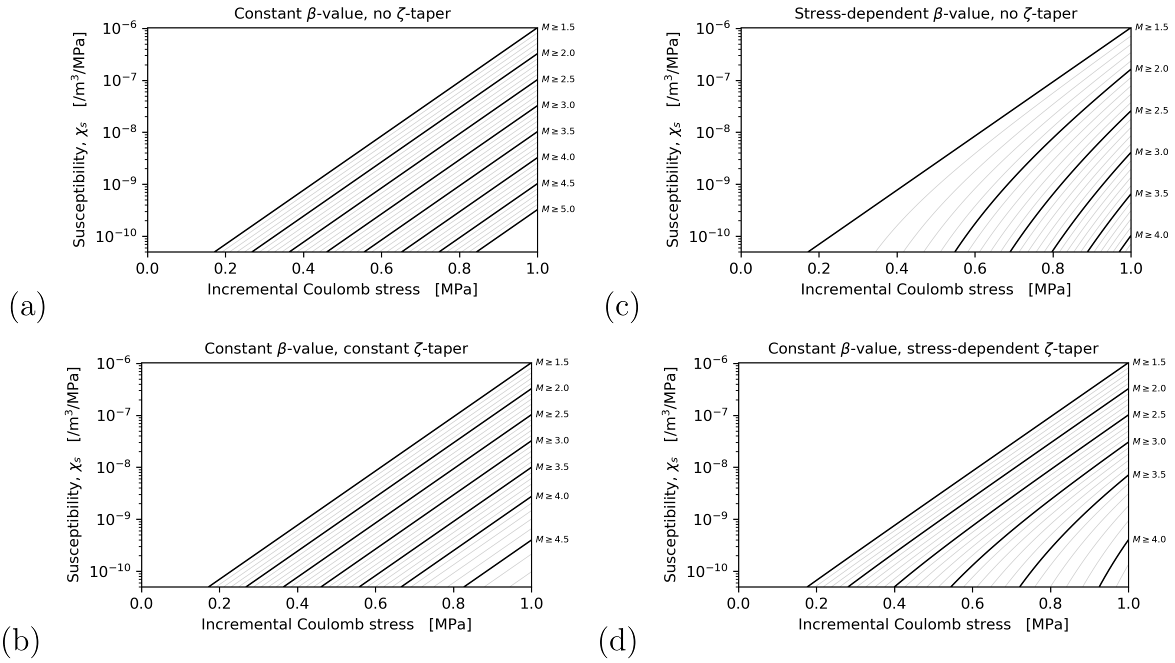


Figure 21. The expected rate, χ, of events of at least magnitude, M per unit reservoir volume and stress increment increases with incremental Coulomb stress under four alternative models: (a) stress-invariant model with $\beta = 2/3$ and $\zeta = 0$ (M_1), (b) stress-invariant model with $\beta = 2/3$ and $\zeta > 0$ (M_3), (c) stress-dependent β with $\zeta = 0$ (M_7), (d) stress-dependent ζ with $\beta = 2/3$ (M_{12}). Lines denote different magnitude thresholds from 1.5 to 5.0 in intervals of 0.1 (grey) and 0.5 (black). Each model (Table 1) is based on its MAP values inferred using the observed 1995–2019 $M \geq 1.5$ events and the poro-elastic thin-sheet stress model.

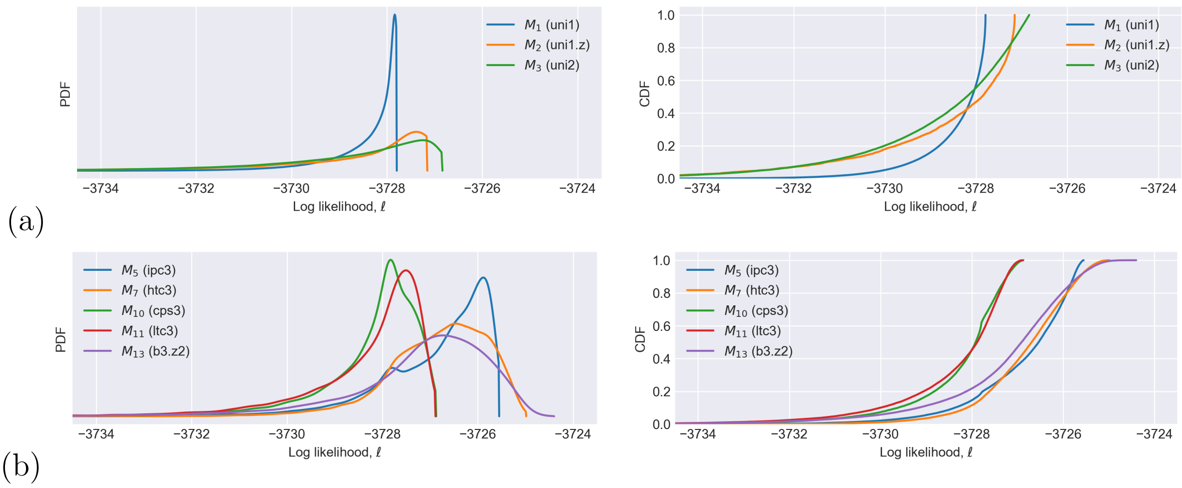
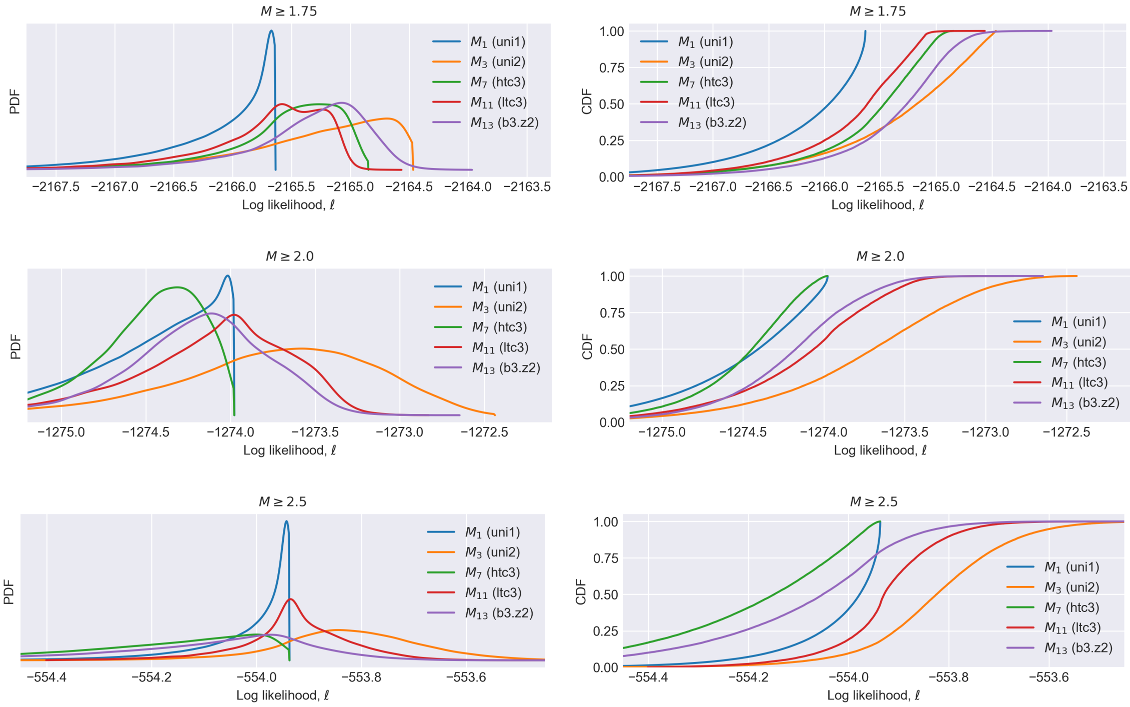


Figure 22. Out-of-sample forecast performance of each alternative magnitude-frequency models (Table 1) measured as the log likelihood distribution, ℓ, of the 1/1/2012 to 1/6/2019 observed $M \geq 1.5$ events, given the posterior distribution of models inferred from the observed 1995–2012 events. (a) Stress invariant models $\{M_1, M_2, M_3\}$. (b) Stress-dependent models $\{M_5, M_7, M_{10}, M_{11}, M_{13}\}$.



(a)

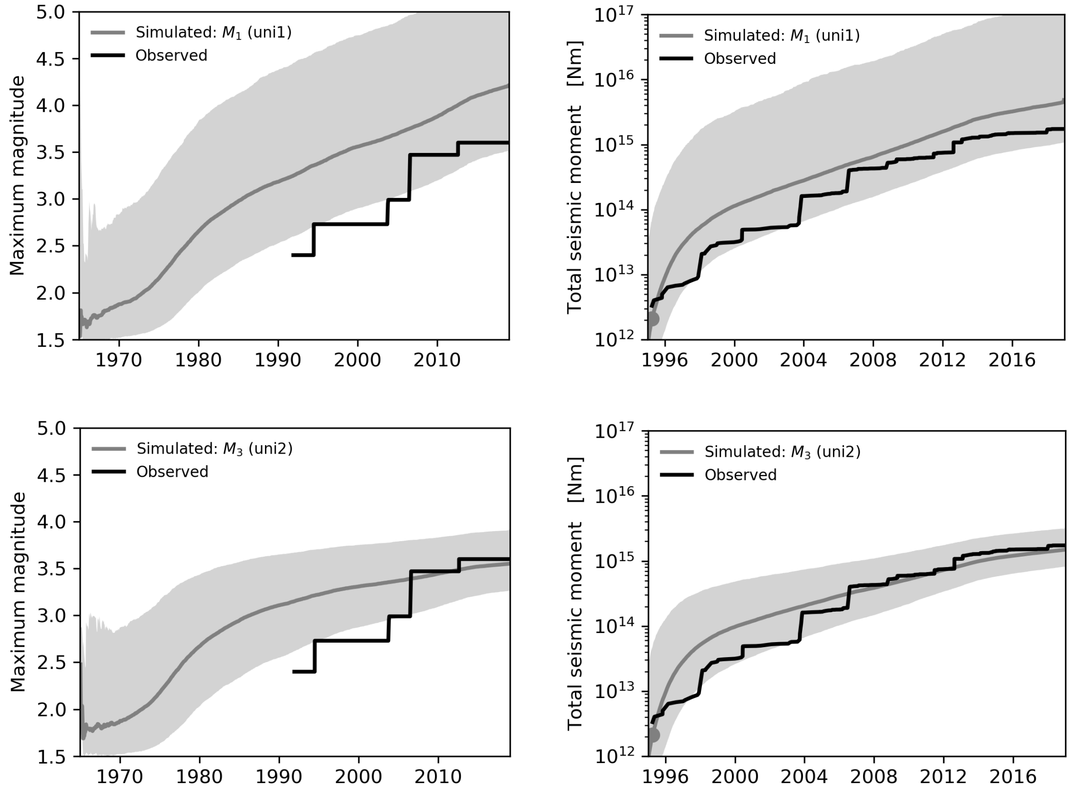
(b)

(c)

Figure 23. Out-of-sample forecast performance of each magnitude-frequency models (Table 1) measured as the the log likelihood distribution, ℓ, of the 1/1/2012 to 1/6/2019 observed (a) $M \geq 1.75$, (b) $M \geq 2.0$, and (c) $M \geq 2.5$ events, given the posterior distribution of models inferred given the observed 1995–2012 events.

Model class		M_i	P_{i1}, Probability of M_i out-performing M_1			
			$M \geq 1.5$	$M \geq 1.75$	$M \geq 2.0$	$M \geq 2.5$
Stress invariant	uni1	M_1	0.500	0.500	0.500	0.500
	uni1.zeta	M_2	0.587	0.950	0.955	0.996
	uni2.zeta	M_3	0.516	0.839	0.846	0.891
Stress-dependent β	ets0.ipc3	M_5	0.909	0.874	0.454	0.157
	ets0.htc3	M_7	0.901	0.836	0.571	0.212
Stress-dependent ζ	ets0.cps3	M_{10}	0.638	0.698	0.666	0.695
	ets0.ltc3	M_{11}	0.692	0.768	0.726	0.748
Stress-dependent β-ζ	ets0.b3.z2	M_{13}	0.872	0.899	0.754	0.414

Table 2. Out-of-sample magnitude forecast performance measured according to the probability of each model out-performing the baseline model of a stress-invariant β-values with ζ=0 (M_1). Models were trained using the observed 1/2012–6/2019 $M \geq 1.5$ events and evaluated using the observed 1/2012–6/2019 $M \geq M_t$ events, where $M_t = \{1.5, 1.75, 2.0, 2.5\}$. Colours vary from red to yellow to green denoting probabilities from 0 to 0.5 to 1 respectively.



(a)

(b)

Figure 24. Time series of observed and simulated maximum magnitudes (left) and total seismic moments (right) for two stress-invariant models: (a) stress-invariant β given ζ=0, M_1, and stress-invariant β-ζ-model, M_3. Given the absence of field-wide $M \geq 1.5$ earthquake monitoring until 1995, the total seismic moment time series start in 1995. Dark grey lines, and light grey regions denote the simulated median values and 95% prediction intervals respectively.

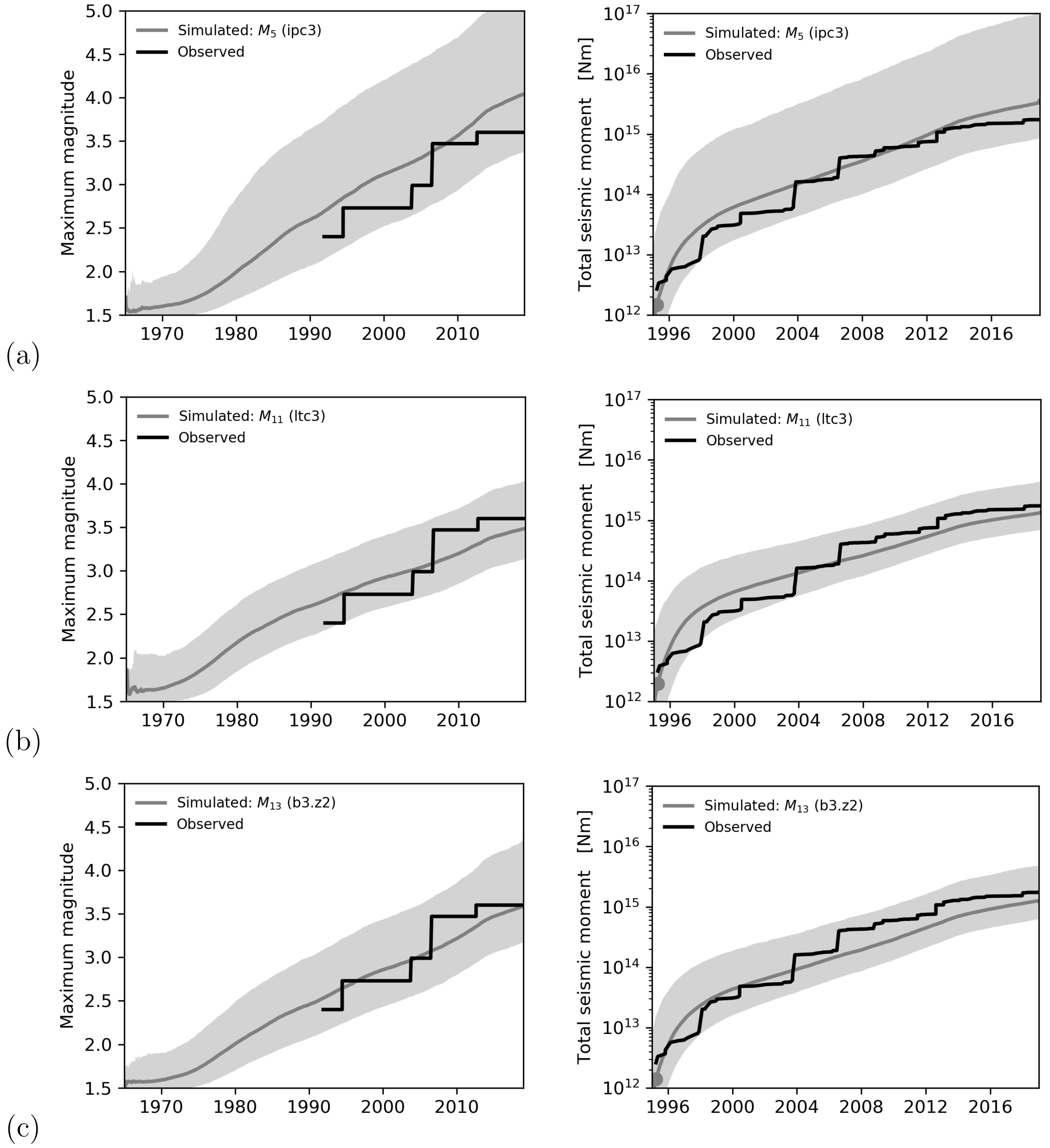


Figure 25. As Figure 24, except for three stress-dependent magnitude-frequency models: (a) stress-dependent β-values, M_5, (b) stress-dependent ζ-values, M_{11}, and (c) stress-dependent β-ζ-values, M_{13}.

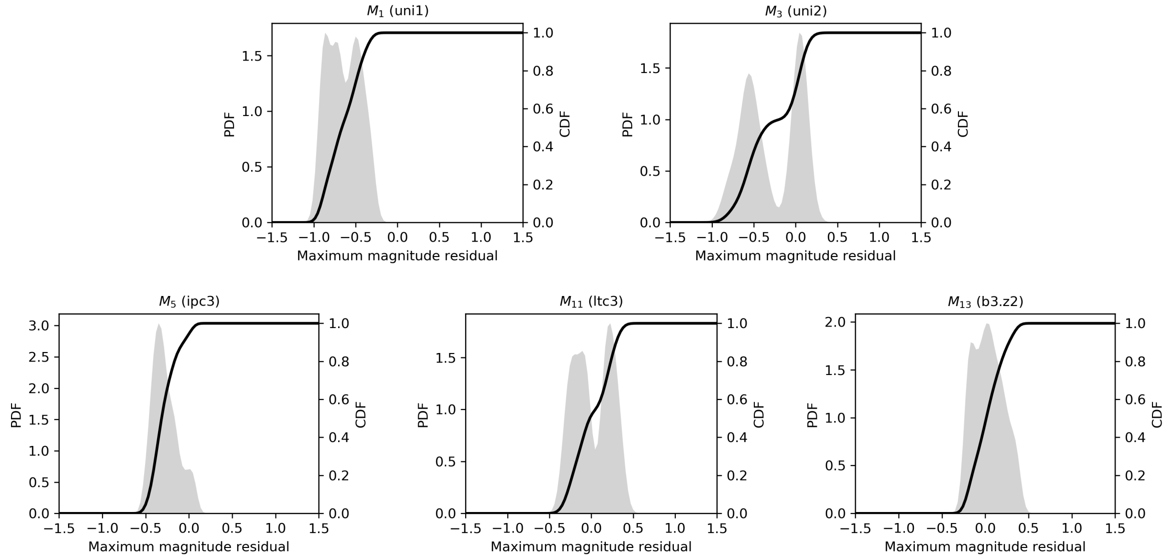


Figure 26. The distribution of all residuals between the observed and simulated time series of maximum magnitudes.

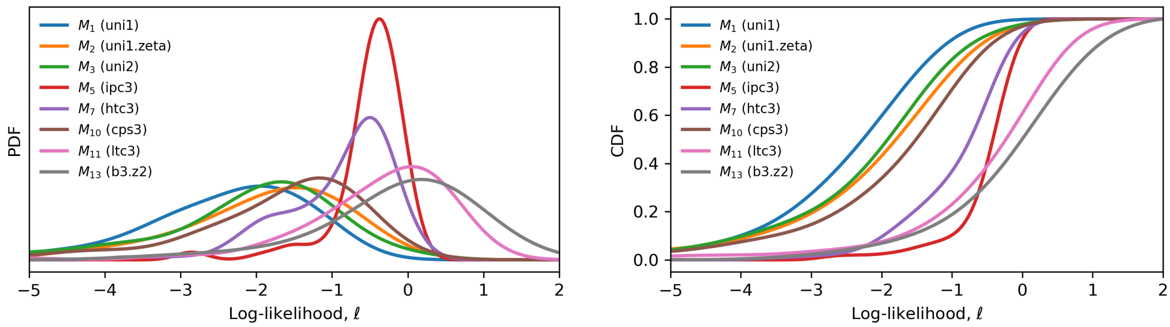
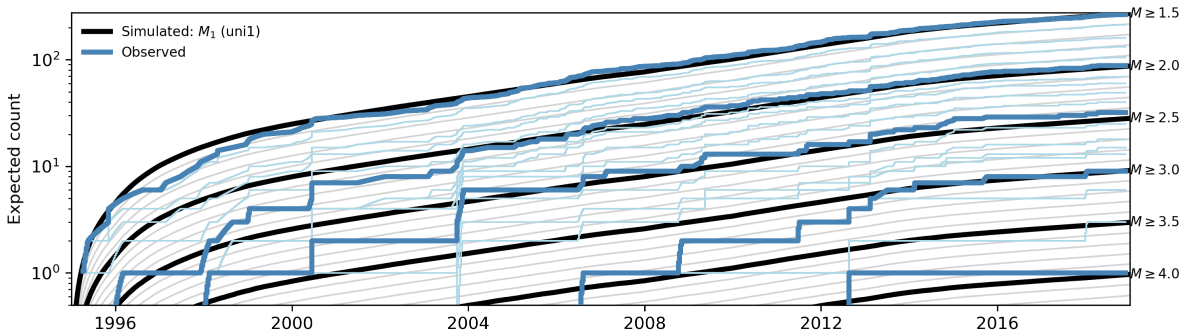


Figure 27. Out-of-sample forecast performance results for the maximum magnitude time series. These were obtained for a 1995 to 2012 training period, a 1965 to 2019 simulation period, and a evaluation period before 1995 and after 2012.

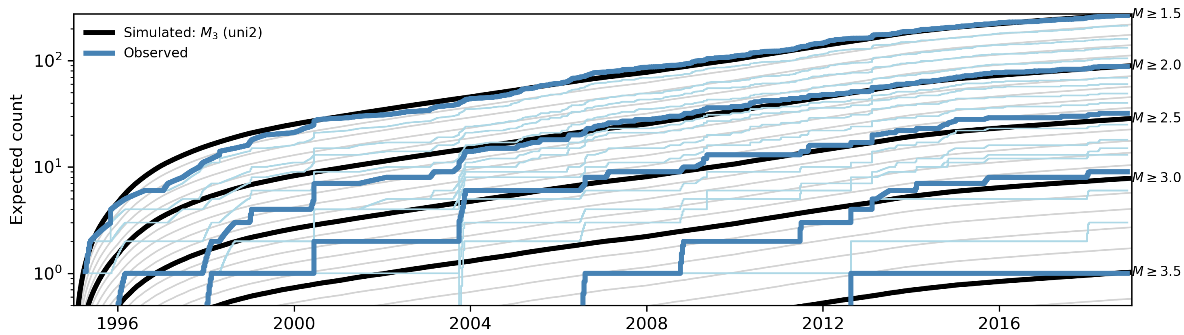
Model class		M_i	P_{ij}, Probability of M_i out-performing M_j							
			M_1	M_2	M_3	M_5	M_7	M_{10}	M_{11}	M_{13}
Stress invariant	uni1	M_1	0.500	0.400	0.370	0.283	0.091	0.343	0.060	0.060
	uni1.zeta	M_2	0.600	0.500	0.580	0.374	0.192	0.424	0.110	0.080
	uni2.zeta	M_3	0.630	0.420	0.500	0.364	0.162	0.394	0.110	0.120
Stress-dependent β	ets0.ipc3	M_5	0.717	0.626	0.636	0.500	0.273	0.576	0.152	0.111
	ets0.htc3	M_7	0.909	0.808	0.838	0.727	0.500	0.727	0.253	0.192
Stress-dependent ζ	ets0.cps3	M_{10}	0.657	0.576	0.606	0.424	0.273	0.500	0.141	0.121
	ets0.ltc3	M_{11}	0.940	0.890	0.890	0.848	0.747	0.859	0.500	0.400
Stress-dependent β-ζ	ets0.b3.z2	M_{13}	0.940	0.920	0.880	0.889	0.808	0.879	0.600	0.500

M_j

Table 3. Relative pairwise model forecast performance for the observed out-of-sample maximum magnitude time series as measured by the probability, P_{ij}, of model M_i out-performing model M_j. Colours vary from red to yellow to green denoting probabilities from 0 to 0.5 to 1 respectively. By definition, $P_{ij} + P_{ji} = 1$, so the above diagonal cells contain the same information as their below diagonal counterparts.

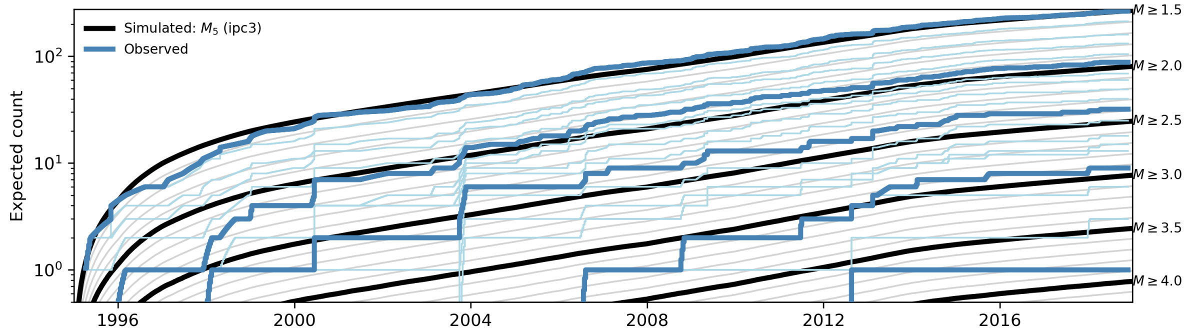


(a)

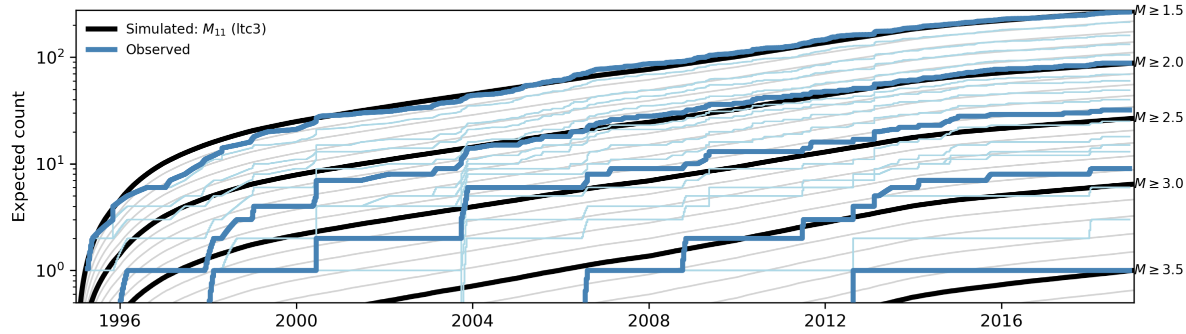


(b)

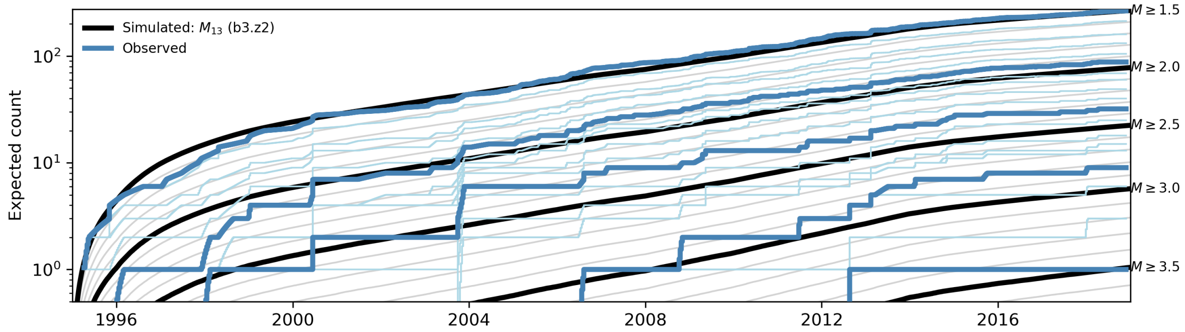
Figure 28. Observed event count time series compared to expected event counts obtained by model simulations from 1995 to 2019 for (a) stress-invariant β-values,M_1, and (b) stress-invariant β-ζ-values, M_3. See Table 1 for model details.



(a)

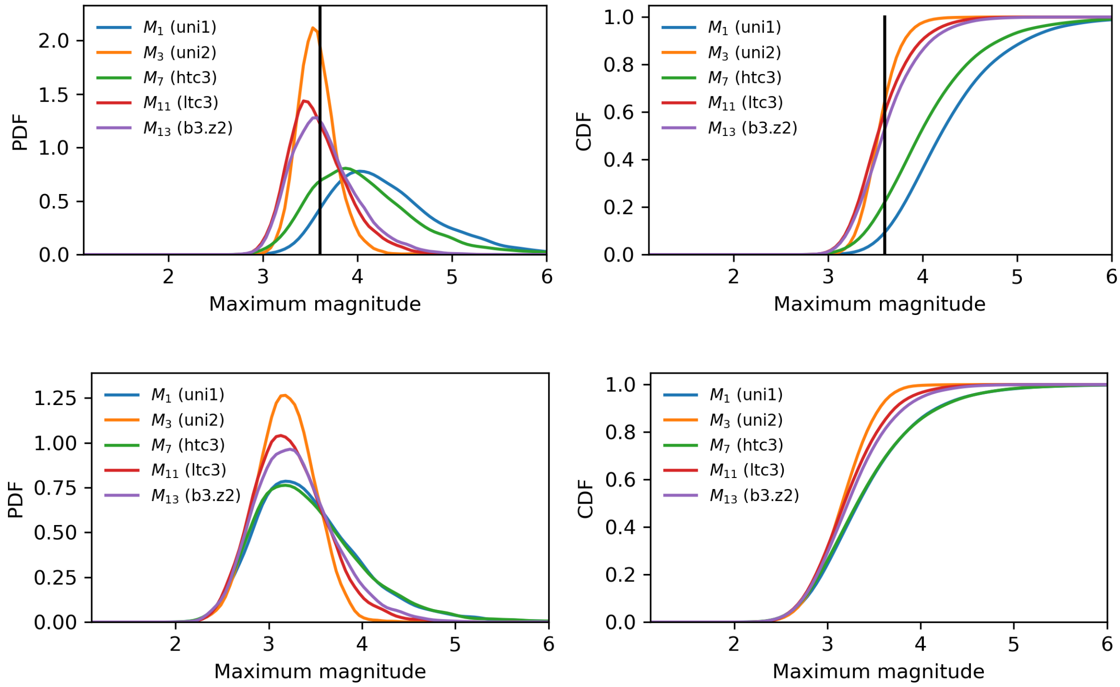


(b)



(c)

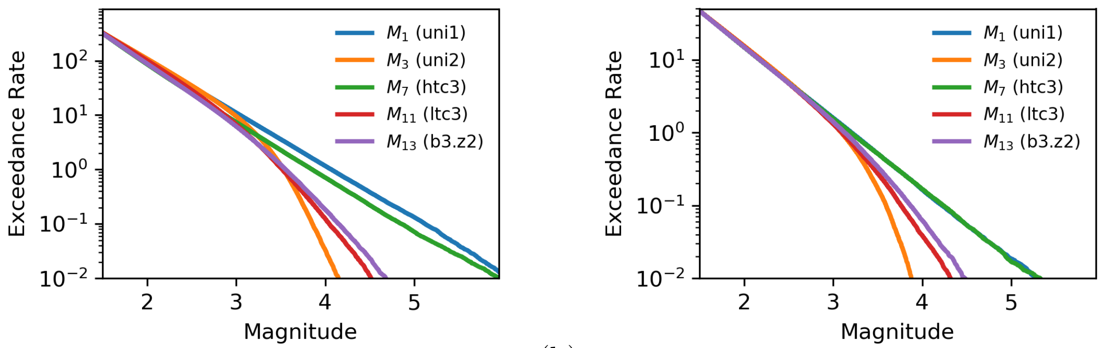
Figure 29. As Figure 28, except for (a) stress-dependent β-values, M_7, (b) stress dependent ζ-values, M_{11}, and (c) stress-dependent β- and ζ-values, M_{13}. See Table 1 for model details.



(a)

(b)

Figure 30. The distribution of maximum magnitudes according to model simulations for the periods (a) 1965–2019, and (b) 2019–2024. The vertical black line denotes the observed $M = 3.6$ maximum magnitude. The expected maximum magnitudes are 4.3, 3.6, 4.1, 3.6, 3.7 for the M_1, M_2, M_7, M_{11}, and M_{13} models respectively.



(a) (b)

Figure 31. Magnitude exceedance rate according to model simulations over the periods (a) 1965 to 2019, (b) 2019 to 2024.

Model	**50%**	**10%**	**1%**
M_1 (uni1)	3.36	4.18	5.29
M_2 (uni1.zeta)	3.16	3.58	3.95
M_3 (uni2)	3.18	3.58	3.88
M_5 (ipc3)	3.33	4.25	5.48
M_7 (htc3)	3.34	4.21	5.32
M_{10} (cps3)	3.29	3.81	4.45
M_{11} (ltc3)	3.20	3.74	4.32
M_{13} (b3.z2)	3.24	3.84	4.47

Table 4. Comparison of the magnitudes with a 50%, 10%, and 1% chance over exceedance between 2019 and 2024 according to simulations of the different magnitude-frequency models without imposing an upper bound, $M_{\max}$, to these probability distributions. These results are based on the *2019 GTS Raming* production scenario.

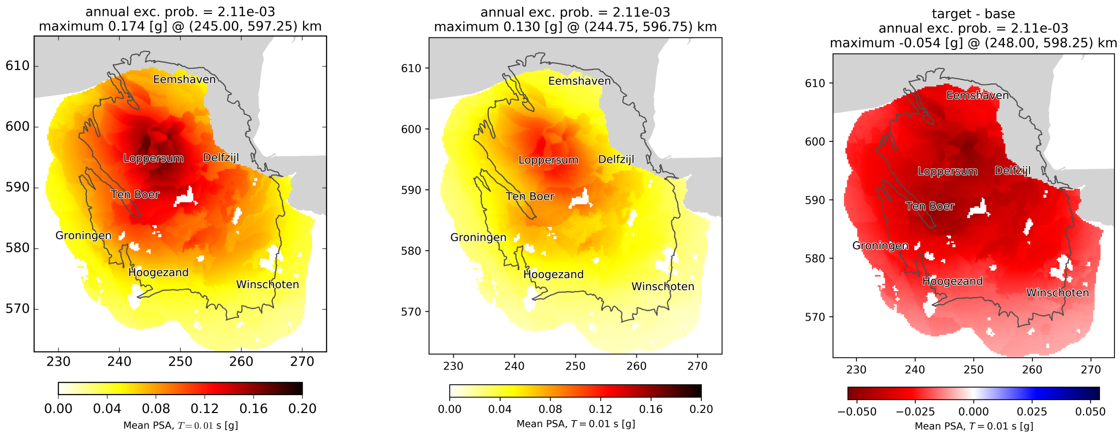


Figure 32. Seismic hazard represented as the peak ground motion acceleration (PSA, $T = 0.01$s) maps with a 2.1% annual chance of exceedance in 2020 computed using the M_5 magnitude model (left), verses a 0.2, 0.8 weighted combination of the M_5 and M_{11} models respectively (right), and the difference between these two maps (right). This represents a 31% reduction in the maximum seismic hazard by allowing for the possibility of an exponential taper in the power-law distribution of seismic moments.

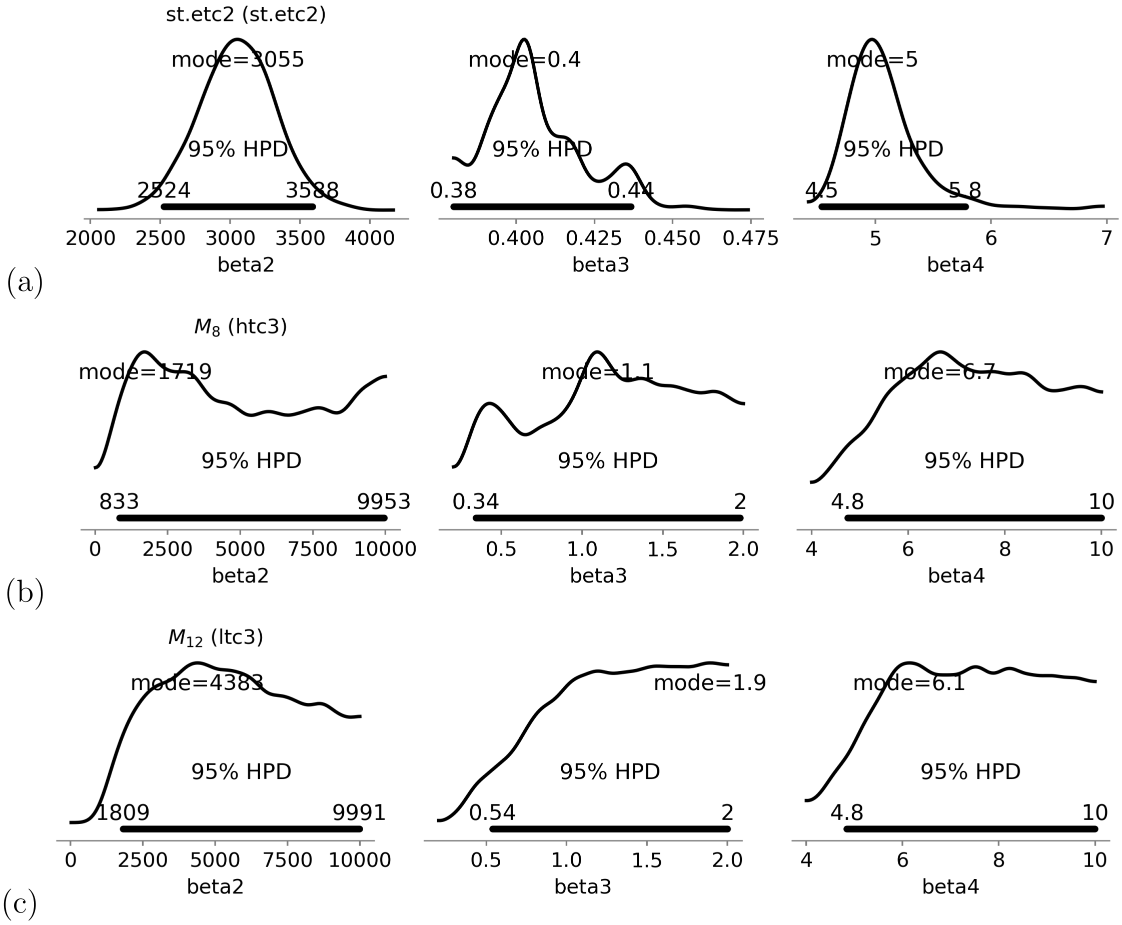


Figure 33. Marginal posterior distributions of the the poro-elastic thin-sheet intra-reservoir stress model parameters defined in section 5 obtained in combination with (a) the extreme threshold failures model for space-time distribution of earthquake occurrence (st.etc2), (b) the stress-dependent β-model of earthquake magnitudes (M_8, htc3), and (c) the stress-dependent ζ-model of earthquake magnitudes (M_{11}, ltc3).