WP2: Review of existing alternative approaches and knowledge gaps

Review of seismogenic source models for the Groningen gas reservoir

April 2020

GFZ
Helmholtz-Zentrum
POTSDAM

NORSAR

Contents

1 Executive summary

The WP2 report constitutes a follow-up of the WP1 report on seismogenic source models for Groningen. While the WP1 report was dedicated to reviewing existing NAM models, the WP2 report inspects existing alternative approaches and knowledge gaps. To this end, we collected more than 20 alternative models representing members of established approaches and techniques. Further variations of each model type may exist, but have not been included for the sake of simplicity.

The range of model types may be grouped into (1) deterministic physical models, (2) machine learning approaches, (3) purely statistical and empirical models and (4) so-called hybrid models integrating physical and statistical approaches into one framework. Additionally to reviewing and discussing the individual models, we present knowledge gaps that affect almost all of the existing seismogenic source models. The final discussion concentrates on evaluating the strengths and weaknesses of each model class rather than trying to rank individual models.

Physical models and *deterministic scenarios* were mostly developed to investigate specific aspects of the triggering and rupture process of induced earthquakes. They cannot be directly applied to seismic hazard studies, because of limited computational power as well as unknown model parameters and initial conditions. However, they serve to improve the process understanding and to enhance concepts for seismicity models. Typically, quasi-static triggering approaches are implemented. For Groningen, almost all seismicity occurs within or close to the reservoir layer. Many of the suggested models assume that the earthquakes are triggered at normal fault offsets, where differential field compaction and pore pressure variations lead to increased Coulomb stress that may exceed the frictional strength on the fault. A key question addressed by physical models is how fault geometry and throw impact the probability of earthquake triggering, and whether and how so-called runaway events can rupture the overburden or basement. Other studies investigated how large the remaining elastic deformation can become if inelastic deformation and creep is accounted for in order to better understand the time-dependent potential for larger magnitude earthquakes.

Machine learning (ML) could present a possibility to bypass the problem of incomplete and potentially biased process understanding, as it extracts relevant patterns and information directly from the available multivariate data sets without requiring insight into the underlying physical mechanisms. Although first tests and applications of machine learning are available for Groningen, the comparison to baseline models shows no significant improvement and the ML model accuracy and forecasting power remains limited by the small number of earthquakes available for model training. Another drawback of the machine learning approaches is that predictions of the seismicity are problematic due to the fact that future stress scenarios that are not covered in the historic training dataset.

Empirical statistical models avoid any physical constraint as well, but implement - in contrast to ML - parametric relations between observables and use basic statistical models as e.g. the Poisson model to account for aleatoric variability. Examples of such empirical relations are the Gutenberg-Richter

magnitude distribution, the Omori aftershock decay and linear or nonlinear relations between production volume and seismicity rate. Statistical models can be strong in making short term forecasts, as for instance the prediction of aftershock activity from ETAS-type models or the prediction of the cumulative seismicity. However, a drawback similarly to machine learning is that purely statistical models are often unable to provide reliable predictions for system states evolving outside the range covered during the learning period. In practice, statistical forecasts are often only reasonable for slow system changes, where model parameter can be adapted appropriately.

Hybrid models combine the physical model results with statistical approaches. In particular, they are based on modelled stress values in the seismogenic zone and a triggering criterion leading to a forecast of the average rate. The natural variability in magnitude and time is accounted for by the statistical Gutenberg-Richter and Poisson models, while aftershocks are at times included employing the ETAS model. Hybrid models are the preferential model class towards time-dependent hazard assessment for induced seismicity, as they are able to forecast induced seismicity for loading scenarios outside the range of previous production. The newest NAM model represents an entity of such hybrid models.

Hybrid models depend on their specific input parameters. Among them, reservoir loading, fault structure and initial stress state play a key role. For instance, the level of the initial stress state on faults is crucial for the triggering and rupture of induced earthquakes. If the initial stress is low, no earthquakes will be triggered. In contrast, faults which are already critically stressed in the beginning can be immediately triggered. Some of the seismicity models assume that the delayed onset in Groningen's induced seismicity is because of the low stress state on faults. Subcritical and variable initial stress conditions are assumed to explain the observed delayed and then accelerated seismicity. In particular, the NAM model assumes that the stress level is always in the tail of the distribution of stress values on faults, which allows the application of the extreme value theory. Alternatively, rate-and-state models assume that nonlinear evolving frictional instabilities lead to the triggering of induced earthquakes. They implicitly consider a time-dependent nucleation process, and therefore can naturally explain a time delay in the seismic response on stress loading scenarios. Which of those models is more appropriate remains unclear as long as the absolute stress cannot be precisely measured.

Almost all of the alternative hybrid models assume a time-independent magnitude distribution, based on the empirical Gutenberg-Richter relation, which only involves one model parameter. In contrast, the newest NAM model suggests a stress-dependent magnitude distribution with variable b-value and tapering at large magnitudes. The impact of a modified, more complex but less constrained frequency-magnitude distribution on the prediction of larger magnitudes can be significant, and therefore the question is of high relevance for seismic hazard assessment. More research is required before a final conclusion may be reached.

All hybrid models possess strengths and weaknesses and cannot be easily compared. For instance, they have been often developed with regard to different aims and tasks: some are specialised to forecast the cumulative number of events, others to reproduce the spatiotemporal event distribution; some are strong in predicting aftershocks, others to recover the long term trend of seismicity. A majority of the hybrid models is based on reservoir models and reasonable physical considerations, but the validity of those assumptions is not always clear. So far, it is unknown whether the assumption of the NAM model holds that the stress level is in the tail of the on-fault stress distribution or whether rate-state triggering is applicable. The lack of knowledge and uncertainties need to be considered by seismicity models, a topic which should receive increased attention in future refinements of seismogenic models.

For a quantitative comparison of the alternative models, a systematic testing approach would be essential. In such a testing infrastructure, models could be run and tested for unseen future data by researchers who are independent of the model developers. Such an approach could ensure the best usage of existing seismogenic models for Groningen and could help to implement a proper weighting in so-called ensemble models for seismic hazard assessments.

2   Introduction

Seismogenic models describe the seismicity, in particular the rate of earthquakes as a function of time, space, and magnitude. These models, also called seismicity models, are thus the backbone of any seismic hazard assessment. Traditional seismic hazard calculations assume time-independent models, specifically the Poisson process with constant seismicity rate, while the spatial component is described either by geologically defined areal sources with uniform rate, or by smoothing of historic seismicity assuming that earthquakes will occur in the future in the same location with the same rate. Furthermore, magnitudes are traditionally assumed to follow a time-independent trend, a Gutenberg-Richter distribution with fixed a and b-value, as observed in the past. The underlying assumptions are the independence of earthquakes and the stationarity of the earthquake generation process.

However, earthquakes cannot be treated as independent events. The most important types of short- and intermediate-time earthquake interactions are aftershock sequences. Aftershocks are associated with a main shock. Their occurrence rate generally scales exponentially with the main shock magnitude and decays with time according to a power-law, the so-called Omori-Utsu law (Utsu et al., 1995). The duration of aftershock sequences is typically in the range of weeks to months, but it is expected to inversely scale with the tectonic stressing rate and can thus be much longer in low seismicity zones and shorter for induced seismicity with high stressing rates (Dieterich, 1994). The spatiotemporal patterns of aftershocks can be well modelled by the statistical Epidemic-Type Aftershock Sequence (ETAS) model, which explains the activity by a superposition of ongoing Omori-Utsu type aftershock sequences triggered by past events and a time-independent background rate (Ogata, 1988; Ogata, 1998). Thus, in a formal sense, ETAS type models do not distinguish between fore-, main- and aftershocks, and their forecasts are only based on the number and magnitudes of previous events and a tectonic background rate of mainshocks.

While the underlying mechanisms are still unclear, statistical seismicity models, particularly the ETAS model, are sufficiently well tested to be used for official earthquake warnings. There is often a demand for authoritative forecasts of triggered events in the aftermath of a major damaging earthquake (Jordan et al., 2014). Aftershock models incorporating the Omori-Utsu law, such as ETAS or simpler alternatives, are generally used to provide such short term forecasts, with a time horizon of weeks, months or perhaps a year; as done e.g. in Italy, California, and New Zealand. An ideal Operational Earthquake Forecasting (OEF) system would be integrated with seismic hazard and risk tools. However, so far, the time-dependence of the seismicity is usually ignored for seismic hazard assessment.

For seismic hazard assessment, therefore, the earthquake catalogues are usually declustered before the parameters of the Poisson and Gutenberg-Richter models are estimated. The

> Seismic hazard assessments of induced seismicity require sophisticated models accounting for time-dependent driving forces.

declustering is necessary in order to approximate a stationary Poisson model for the tectonic mainshock activity. Declustering methods separate an earthquake catalogue into foreshocks, mainshocks, and aftershocks. However, there is no unique way to do this and any algorithm has potential pitfalls (van Stiphout et al., 2012). In any way, the assumption of a stationary driving force in standard hazard estimations for natural seismicity is obviously inadequate in the context of anthropogenically induced seismicity as in the Groningen gas production site. In this region, the tectonic forcing is negligible leading to almost no activity outside the gas fields, while the activity inside the gas fields is directly related to the time-dependent production of gas. Therefore, seismic hazard assessments of induced seismicity require more sophisticated models incorporating the time-dependency of the driving force.

To consider non-stationary driving forces, physical models have to be developed, implementing earthquake trigger processes and rock failure criteria. In particular, the Coulomb failure stress (CFS), i.e. the difference between the

> The change of Coulomb failure stress (ΔCFS) controls earthquake triggering.

shear stress and the product of the friction coefficient and compressive normal stress, is a key parameter controlling rock failure and earthquake triggering (Okada, 1992; King et al., 1994). Therefore, physics-based seismicity models are almost all based on the CFS parameter, although in some cases other variables are implemented as proxi for CFS, as for instance pore pressure change or reservoir depletion. However, the specific approach how the trigger criterion is defined and how rupture nucleates varies among models. For instance, a Coulomb failure model assumes a simple threshold criterion for CFS, which after exceedance instantaneously leads to a mechanically triggered earthquake. The rate-and-state seismicity model assumes a frictional instability, where the co-seismic rupture develops after an aseismic nucleation phase over a finite time depending on the level of the CFS. Thus, the rate-and-state model can simulate time-dependent earthquake rates other than the time function of the CFS loading, as for instance aftershocks. In contrast, the Coulomb failure model cannot explain aftershock occurrence and is often combined with a statistical model to overcome this limitation. Such a hybrid model, i.e. a combination of a physical and a statistical model, has for instance been developed by NAM (Bourne and Oates, 2017b).

Independent of the question whether a threshold or a rate-and-state model is used, the knowledge of the time dependent CFS function, or its time derivative (the change in CFS), is crucial to understand time-dependent earthquake rates. This is clearly relevant for anthropogenic seismicity, but even at plate boundaries, where the earthquakes are mainly related to the constant stressing rate from plate motions, we see indications of non-stationarity triggered by different processes (Vidale and Shearer, 2006). Examples include so-called static CFS changes from mainshocks (King et al., 1994; Stein, 1999), dynamic stress perturbations from seismic waves (Kilb et al., 2002; Felzer and Brodsky, 2006), or

stress changes induced by postseismic deformations such as afterslip (Perfettini and Avouac, 2004; Perfettini et al., 2019), poroelastic (Nur and Booker, 1972; Jonsson et al., 2003) or visco-elastic relaxation (Pollitz et al., 1998). Some of these processes are active on overlapping time scales. Hence, for example, both dynamic and static stress changes seem to be very important at short time scales, while large scale poro- and visco-elastic changes may lead to temporally delayed triggering. Other factors such as rupture type and small-scale slip heterogeneity may also have a first order effect on the resulting seismicity patterns (Marsan, 2006; Narteau et al., 2009).

The NAM seismogenic model (Bourne and Oates, 2017a; Bourne et al., 2018) is a hybrid approach with a combination of physical and statistical model components and was developed specifically for Groningen. In this gas field, production takes place since 1963 and more than 270 earthquakes with magnitude $M_L \geq 1.5$ have been recorded since 1986. In particular, the NAM model is based on calculations of the time-dependent compaction strain and the assumption of statistically distributed pre-stress values, which leads to a time-dependent seismicity rate modelled by an inhomogeneous Poisson model as described in detail in our first report (WP1, 2019). However, although the model is based on reasonable assumptions and explains the observed seismicity well within the confidence intervals, it involves a number of assumptions which limits its predictive power. Therefore, the purpose of this report is to review and discuss alternative model approaches and their advantages and disadvantages.

In chapter 3, we introduce and discuss alternative models already applied or generally applicable in the Groningen field with special emphasis on the involved physics, required input parameters and uncertainties, particularly considering data availability. In section 3.1, we firstly focus on purely physical geomechanical models and their results. These models are useful to understand the physical processes leading to induced seismicity in the Groningen field. However, they cannot account for the unpredictability of small-scale processes and heterogeneities. Partly this is due to computational costs, which prevent modelling of the earthquake process on scales of millimetres (or less) up to the scale of tens of kilometres, and partly it is due to an insufficient knowledge of properties, initial conditions and relevant processes in the reservoir and surrounding crust. Thus, for direct applications in seismic hazard assessments, statistical models are needed, which can account for natural variability and stochastic processes. Machine learning is one approach that can deal with the complex relations between input variables and earthquake occurrences. In section 3.2, we introduce and discuss such a purely data-driven approach to relate compaction to earthquake data in Groningen. Parametric statistical models relevant for the Groningen field are discussed in section 3.3, where purely empirical and hybrid models are analysed. While the empirical models ignore important physical knowledge and constraints, the model class of hybrid models, which includes the NAM model, combines the understanding of physical processes and stochastic components, and is thus most promising for seismic hazard applications in Groningen.

In chapter 4, knowledge gaps and open questions are discussed. Among others, these gaps include

the missing knowledge concerning the actual stress state, the likelihood to trigger neighbouring faults and potential changes of the frequency-magnitude distribution. On the other hand, open questions are related to the potential use of currently available high-precision data and the possibility of comparative model tests as currently ongoing in the context of natural seismicity within the Collaboratory for the Study of Earthquake Predictability (CSEP, Michael and Werner, 2018).

Finally, in chapter 5, we summarise and discuss the strengths and weaknesses of the alternative model approaches and the most important knowledge gaps that should be addressed in the future.

3   Alternative models

The occurrence of felt earthquakes due to gas production in Groningen has initiated numerous studies and model attempts to understand and quantify triggered seismicity in this region. Partly, the understanding of processes and model approaches, which have been developed for natural earthquakes and induced seismicity in other contexts and environments, have been used. In addition, some model concepts have been developed specifically for the Groningen case. The whole bandwidth of available models spans the range from fully deterministic models to purely empirical and stochastic models. In this chapter, we summarise the most important model approaches. However, due to the large number and variety of publications, we were forced to restrict our analysis and cannot discuss all approaches in depth. In particular, we focus on differences to the NAM model approach and the models' applicability to the Groningen field with respect to data availability. In table 1, we provide an overview of the discussed model approaches and their key components, parameters and previous application areas. Please note that this categorisation requires simplifications that can be questioned. However, the purpose of this table is to provide a brief overview, while more detailed descriptions of the models are given in the following sections.

Section 3.1 introduces mechanical models, which are built to understand the triggering mechanism. The subsequent two sections focus on models that account for the generally incomplete understanding of the process, the limited knowledge of the conditions in the crust, and the randomness of the involved processes. In particular, a machine learning approach for Groningen is described in section 3.2, and statistical and hybrid models are discussed in section 3.3.

> Table 1 provides a brief overview on features of the alternative models discussed in this report.

Table 1 – *Overview of the analysed models with their main properties*

Name (page)	Type[a]	Model components[b]	No. of parameters[c]	Applications	References
basicRes (p.12)	A	rectangular reservoir, poroelasticity, CFS	6	generic	Segall (1989)
basicRes+fault (p.12)	A	fault offsetting reservoir, poroelasticity, CFS	>8	generic	Jansen et al. (2019)
3D-Res (p.13)	D	3D FEM model, faults, quasi-static	large	Groningen	Sanz et al. (2015)
basicRes+quasistaticRup (p.14)	D	fault offsetting 2D reservoir, quasi-static rupture simulation	>6	generic	Van Wees et al. (2014, 2017a,b)
basicRes+dynRup (p.16)	D	fault offsetting 2D reservoir, dynamic rupture simulation	>6	generic	van den Bogert (2018), Buijze et al. (2017, 2019)
basicRes+quasidynRup-RS (p.20)	D	fault offsetting 1.5D reservoir, quasi-dynamic rupture simulation,RS	17	Groningen	DeDontney and Lele (2018)
basicRes+flow (p.22)	D	fault offsetting 2D reservoir, fluid flow	>6	generic	Zbinden et al. (2017)
basicRes+salt (p.22)	D	2D DEM model, faults, salt effects	>8	generic	Kettermann et al. (2017)
ML (p.23)	ML	features/predictors, ML-algorithms, out-of-sample tests	≥ 40 (predictors)	Groningen	Limbeck et al. (2018), Lanz et al. (2019)
K2/K3 (p.26)	S	adaptive smoothing in space, time & magnitude	6-7	California	Helmstetter & Werner (2012, 2014)
ETAS (p.27)	S or H	background events & aftershocks, Poisson, GR	9	worldwide	Ogata (1988), Hainzl and Ogata (2005)
NC-model (p.30)	S	compaction/fault slip, activity rate	3-7	Groningen	DeDontney (2017)
NV-model (p.31)	S	number-volume relation, Poisson, GR	3	Groningen	Hettema et al. (2017), Vlek (2019)
SI (p.33)	H	p, criticality, Poisson, GR	2	Groningen	Shapiro et al. (2010), Shapiro (2018)
Wentinck (p.35)	H	p, Poisson, Weibull, GR	6	Groningen	Wentinck (2015)
NAM (WP1)	H	ϵ_{zz}, p, CFS, Poisson, pdf(S_0), ζ, ETAS, GR	≥ 13	Groningen	Bourne and Oates (2017a), Bourne et al. (2018)
CMsub (p.37)	H	p, A, CFS, 1-D rupture propagation, ensemble simulation, GR	>16	Groningen	Dempsey and Suckale (2017)
visco-NAM (p.39)	H	ϵ_{zz}, CFS, Poisson, pdf(S_0), ζ	5	Groningen	Bourne and Oates (2018)
NAM-RS (p.41)	H	ϵ_{zz}, CFS, Poisson, RS	5	Groningen	Bourne and Oates (2018)
Candela-RS (p.42)	H	p, CFS, RS	15	Groningen	Candela et al. (2019)
Richter-RS (p.44)	H	p, ϵ_{zz}, CFS, RS	4	Groningen	Richter et al. (2019)

[a]: A: Analytic, D: purely deterministic, ML: machine learning; S: purely stochastic, H: hybrid

[b] p: pore pressure; ϵ_{zz}: compaction strain; CFS: Coulomb failure stress; pdf(S_0): pre-stress distribution; ζ: extreme threshold failure;

GR: Gutenberg-Richter distribution; A: poroelastic path coefficient; RS: rate-and-state fault friction

[c]: The number of parameters is partly difficult to estimate because of incomplete or unclear explanations. Parameters of the static/dynamic reservoir model are not included.

3.1 Physical models

To understand why earthquakes are triggered by fluid extraction, Segall (1989) applied the theory of poroelasticity (e.g. Geertsma, 1973) to a simplified model setup (**basicRes**). For a horizontal, simplified reservoir as shown in Fig. 3.1, he derived an analytic expression describing the induced stress changes. In response to a change in fluid mass per unit volume Δm inside the reservoir, he showed that the components of the induced stress tensor are given by

$$\sigma_{ij}(\vec{x},t) = \frac{\mu(1+\nu_u)B}{3(1-\nu_u)\pi\rho_0}\left[\int_V \Delta m(\vec{y},t)G_{ij}(\vec{x},\vec{y})d\vec{y} - \Delta m(\vec{x},t)\right] \tag{3.1}$$

with the shear modulus μ, the fluid density ρ_0, the undrained Poisson's ratio ν_u, and Skempton's coefficient B. The Green's function G_{ij} is related to a point source of dilatation. This simple model predict stress changes inside the reservoir, where $\Delta m \neq 0$, and outside of the reservoir.

Earthquakes can be triggered by such stress changes in the seismogenic crust. In particular, an indicative quantity for triggering is a positive change of the Coulomb failure stress (CFS),

$$\Delta CFS = \Delta\tau + f(\Delta\sigma_n + \Delta p) \tag{3.2}$$

with τ, σ_n, p, f being the shear stress, the normal stress (positive in case of extension), the pore pressure, and the friction coefficient, respectively (Okada, 1992; King et al., 1994). Equation 3.1 indicates $\Delta CFS > 0$ and thus, assuming a negligible tectonic background stress, can explain potential triggering of reverse faulting both above and below the reservoir, as well as normal faulting on its flanks. This result is illustrated by the focal mechanisms displayed in Fig. 3.1.

Earthquakes are usually associated with slip on preexisting faults (Scholz, 2002). In Groningen, seismic profiles indicate the existence of faults with variability in fault density, orientations, and throws. The updated static reservoir model of 2015 comprises around 700 interpreted faults,

> Analytic poroelastic solutions of induced stresses exist only for simplified model setups.

some with significant throws (SGS Horizon B.V., 2016). Most of the faults in the Groningen field have offsets less than half the reservoir thickness (Buijze et al., 2017; Wentinck, 2015). Recently, Jansen et al. (2019) provided analytic expressions for a normal fault offsetting a homogeneous reservoir of infinite extension (**basicRes+fault**). The authors demonstrated that the development of infinite elastic shear stresses on a displaced fault, at the internal and external reservoir-fault corners, implies that even small amounts of injection or production will result in some amount of slip or other nonelastic deformation. They also showed that there is a marked difference between the shear stress patterns resulting from injection and production in a reservoir. In both situations two slip patches emerge, but at the start of injection slip occurs immediately in the overburden and underburden, whereas during production the slip may remain inside the reservoir.
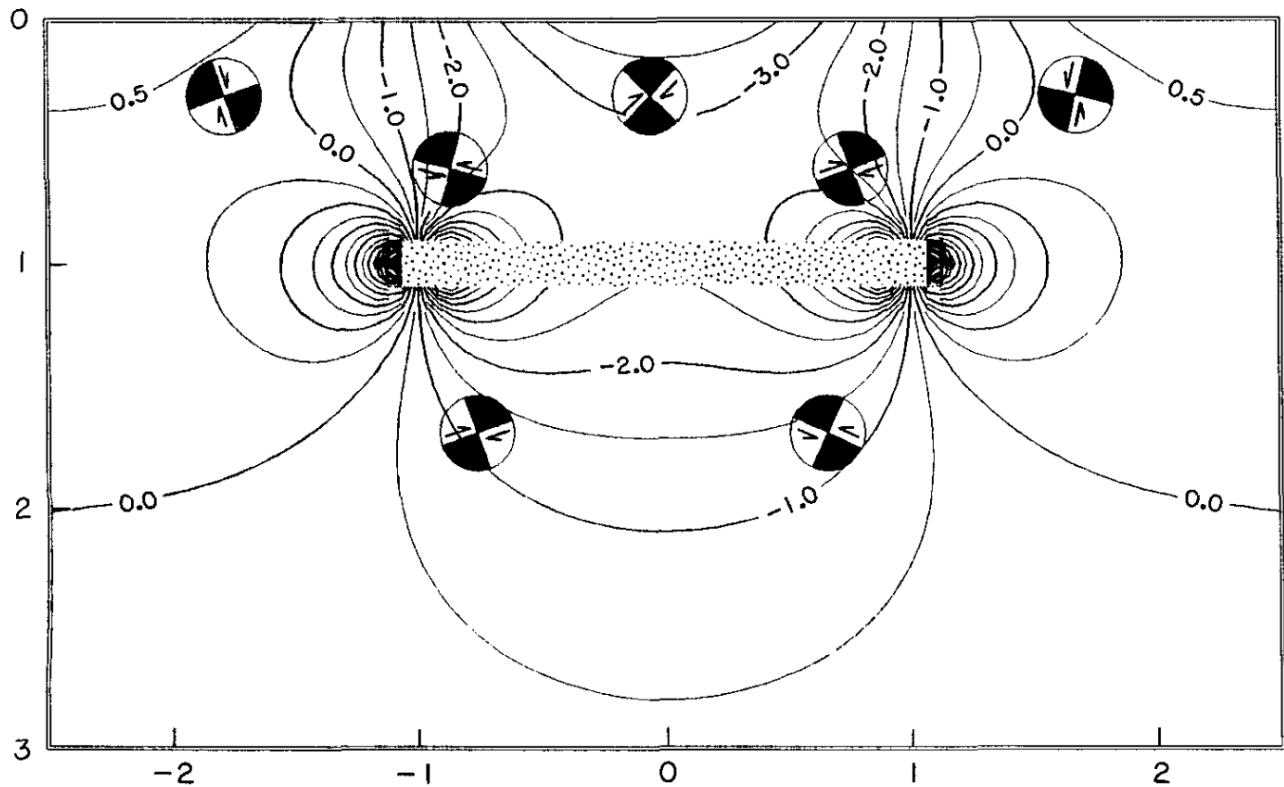
Fig. 3.1 – *Poroelastic solution for a simplified reservoir ($-1 < x < 1, 1-\Delta < z < 1+\Delta$) and plane-strain conditions (no strain in y-direction). Contour lines indicate the normalised horizontal normal stress component σ_{xx} due to fluid extraction (positive in case of extension). Contour lines are not displayed in the production zone. Focal mechanisms indicate planes of maximum shear stress, and sense of maximum shear on those planes. Horizontal distance x and depth z are plotted in units of reservoir depth. The figure is modified from Fig. 6 of Segall (1989).*

For finite reservoirs, stresses on faults offsetting the reservoir cannot be calculated by analytic expressions particularly during rupture nucleation and propagation, which requires numerical models. However, a three-dimensional modelling of the complete Groningen reservoir with sufficient resolution including dynamic effects of seismic events is not feasible due to the computational costs. Thus, three-dimensional reservoir models are so far restricted to quasi-static approaches. An example is the finite-element model of Sanz et al. (2015), which consists of two detailed submodels with horizontal sizes of 16 x 20 km and 13 x 19 km, respectively, including faults and being embedded in a global model of size 180 x 190 km consisting of 21 vertical layers (**3D-Res**). Such models can predict the cumulative fault slip, but do not account for the nucleation and propagation of individual events and thus cannot predict the radiated seismic energy (earthquake magnitudes) or the relation between seismic and aseismic slip. Dynamic models require simulations with a much higher spatial resolution. Furthermore, the lack of detailed information of the true subsurface parameters and the initial stress conditions as function of space would require the sampling of a huge number of alternative model setups to quantify the epistemic uncertainties, which is impossible for the time being. Thus, numerical geomechanical models focus so far on simplified, individual faults, mostly in 2D.

van Wees et al. (2014; 2017a) studied such a simplified model of a fault offsetting a reservoir (**basicRes+quasistaticRup**). The model, illustrated in Fig. 3.2A, incorporates two reservoir compartments, separated by a fault with a certain roughness (characterised by parameter β)

> Van Wees et al. (2014; 2017a,b) studied quasi-static earthquake nucleation and rupture propagation for a simplified numerical model of a fault offsetting a reservoir.

and a throw of half the reservoir thickness. This setup is considered to be representative for major fault zones in the central area of the Groningen field, in the vicinity of which most seismic events occurred. Their plane-strain finite element model had dimensions of 6 x 6 x 6 km. In this model, the continuously declining pore-pressure in the reservoir induces a ΔCFS on fault elements potentially leading to the initiation of an earthquake rupture. After each load step, the stresses on the fault patches are evaluated. If the yield stress is exceeded on a patch, slip occurs inducing stress changes on the other patches. The rupture extension and propagation is calculated by an iterative procedure until all stresses become subcritical. It should be noted that in the approach of van Wees et al. (2017a), the earthquake initiation and propagation is described by a static-kinematic friction law with quasi-static stress interaction, where stress changes occur instantaneously. Particularly, their model belongs to the class of *inherently discrete* models, which have the property that individual cells may fail independently, with the problem that they do not have a well-defined continuum limit (Rice, 1993). In contrast, fully dynamic and continuous rupture models can only reproduce the observed richness of slip complexity if the nucleation dimension is chosen sufficiently small (Cattania, 2019). Therefore, comparable fully dynamic rupture simulations would be computationally very costly, which justifies the approach of van Wees et al. (2017a).

The study of van Wees et al. (2014) and van Wees et al. (2017a) showed that ΔCFS-values are far more pronounced on faults bounding or offsetting the reservoir than elsewhere. Furthermore, the model is found to explain the delay in the onset of induced seismicity as well as the non-linear increase in seismic moment observed in Groningen, if the faults producing induced seismicity are not critically stressed at the onset of depletion. However, they also find a few important discrepancies to the observations:

1. The observed b-value of $b \approx 1$ and maximum magnitudes of $M_{max} \approx 3.6$ can only be reproduced if the fault is critically stressed from the beginning. However, in this case, the model would predict that induced seismicity occurred immediately after the start of production, which is in contradiction to the observations. For an initially subcritical stress state, the b-value is larger and the maximum magnitude smaller.

2. Quickly after the onset of the seismicity, the model predicts an almost linear increase of the cumulative seismic moment with time, equivalent to a constant seismic moment release rate (see yellow curves in Fig. 3.2B). This is not in agreement with the observed increase of the seismic moment rate with time (Bourne et al., 2014).
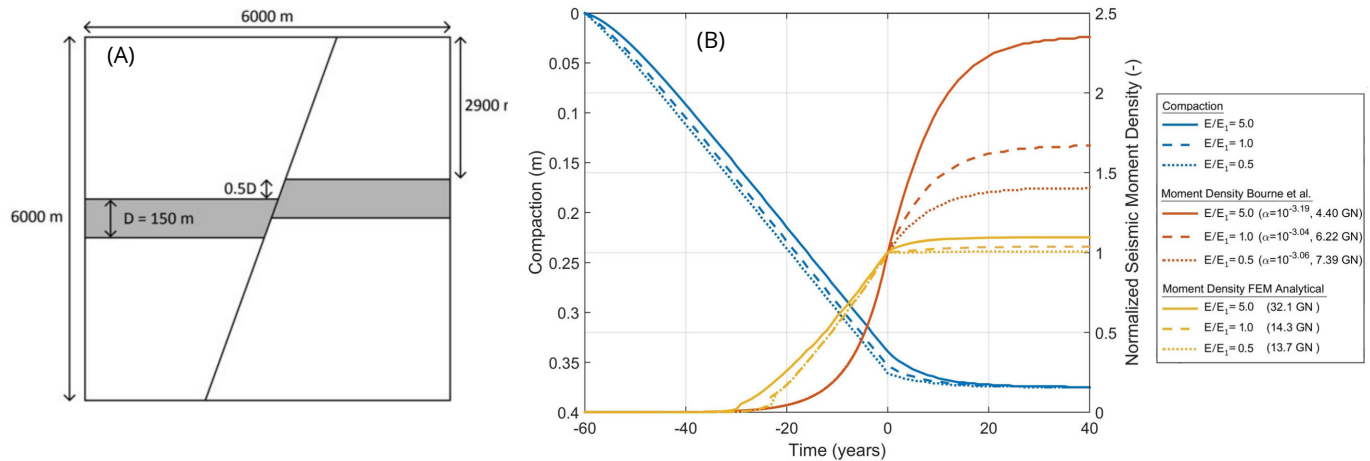
Fig. 3.2 – *(A) Schematic model of a normal fault offsetting the reservoir (modified from Fig. 5 of van Wees et al., 2017a). (B) Seismicity response to a shut-in of production at time 0: the shape of the cumulative seismic moment release (red and yellow lines, right scale) and the corresponding compaction (blue lines, left scale) are shown as function of time. The red curve is based on the heuristic correlation between compaction and seismic moment of Bourne et al. (2014), while the yellow curve represents the prediction of the model of van Wees et al. (2017b). The latter predicts much less activity after the shut-in compared to Bourne et al. (2014). (B) is a modified version of Fig. 8 of van Wees et al. (2017b).*

3. In the geomechanical model, 4 to 40 km of fault length would suffice to generate the observed cumulative moment. This is about one order lower than expected from the distribution of seismicity and faults in the field.

The authors discuss several possible explanations, including:

- The model only considers a single fault, while in reality, multiple faults with different orientation are present. At first, only optimally oriented faults will be activated. However, at later stages the progressive contribution of an increased number of reactivated faults will contribute to the enhancement and nonlinear growth of seismic moment release.

- The model neglects the impact of absorption of stress by the salt cap rock and energy losses through aseismic slip, which might explain the overestimation of the seismic moment release.

The first explanation is similar to the assumption of Bourne and Oates (2017b) that pre-stress values are broadly distributed and that only the tail of this distribution is close to the failure stress. As Bourne and Oates (2017b) showed, this assumption can indeed explain the observed acceleration of the seismic moment release. The second assumption is empirically used by Bourne et al. (2014) by considering the strain partitioning coefficient α. However, the extrapolation of the empirical strain partitioning function of Bourne et al. (2014) is questioned by the model results of van Wees et al. (2017b), who extended their previous model to account for viscoelastic material behaviour. In particular, van Wees et al. (2017b) considered Kelvin-Chain rheology leading to ongoing compaction after cessation of reservoir production. After shut-in, the model predicts an abrupt decrease of seismicity,

in contrast to the smooth prolongation predicted by Bourne et al. (2014). This point is highlighted in Fig. 3.2B, where the shape of the predicted cumulative seismic moment release as function of time is compared for both models.

For the same model setup (Fig. 3.2A), van den Bogert (2018) and Buijze et al. (2017, 2019), explored the fully dynamic initiation and propagation of ruptures, including their waveforms and source-time functions (**basicRes+dynRup**), in order to understand the impact of reservoir depletion, reservoir offset and thickness on the occurrence of fault instability and the magnitude of the simulated earthquakes. For this purpose, they used a 2D crustal model with dimensions of 6 x 6 km, described as a linear-elastic half-space, incorporating a horizontal homogeneously depleting reservoir with offset. The cross-cutting 1D fault is straight and behaves according to a Mohr-Coulomb friction law with linear slip-weakening relationship. For the model setup adapted to the Groningen field, van den Bogert (2018) conducted more than 250 dynamic rupture simulations with varying reservoir offsets, reservoir thicknesses, and different parameters of the linear slip-weakening relationship.

> For the same model setup of a single fault offsetting a reservoir, van den Bogert (2018) and Buijze et al. (2017, 2019) studied dynamic earthquake nucleation and rupture propagation.

In the simulations, fault slip occurs at a fault patch, once the shear stress reaches the shear strength. The slip reduces the shear strength (slip-weakening) and leads to stress changes on the whole fault. The fault slip is aseismic as long as incremental reservoir depletion is required to increase the slip displacement and propagate the slip patch. A seismic event occurs if no further incremental reservoir depletion is required and slip increases autonomously. The dynamic rupture simulations are conducted using an implicit solution scheme in a commercially available finite-element software. The simulations showed that seismic ruptures are always preceded by aseismic slip (nucleation phase), which finally accelerates to seismic slip (rupture) when its size exceeds a critical length L_c which adheres to the analytical expression derived by Uenishi and Rice (2003).

The pressure drop ΔP_S necessary for triggering seismic ruptures is found to be independent of the reservoir width, but dependent on the slip-weakening relationship and the reservoir offset. In particular, it is found to strongly decrease for increasing normalised reservoir offset (absolute reservoir offset divided by reservoir thickness) in the range between 0 and 1. Reservoirs with a normalised offset of about 1 are most prone to fault slip and seismic rupture, while reservoirs with a small normalised offset sustain a significantly larger reservoir depletion before the onset of seismic

> Larger fault offsets promote early rupture nucleation, but have reduced down-dip rupture sizes. In contrast, the smaller the offset, the larger is the potential for rupture propagation into the basement.

rupture (assuming the same in-situ stress and fault slip properties). Depletion causes critical stresses at both the top of the hanging wall and base of the footwall. Whereas larger fault offsets promote the nucleation of ruptures, they reduce the down-dip rupture size, because ruptures are more easily arrested in more stable stress regions, which formed during depletion. On the other hand, the smaller the offset, the larger is the potential for rupture propagation into the underburden. Propagation outside the reservoir interval is additionally promoted by critical in situ stress, a large stress drop, and small fracture energy (Buijze et al., 2019). Apart from the offset, the onset of fault slip is influenced by the orientation of the fault relative to the background stress field and the initial friction coefficient. Furthermore, the necessary pressure drop decreases with increasing slope of the linear slip-weakening relationship. The required pressure drop is found to influence the overall seismic coupling: The larger the required pressure drop, the more ductile is the structural response to reservoir depletion; that is, the fraction of aseismic to total slip increases.

Based on his findings, van den Bogert (2018) suggested to separate the depletion pressure ΔP_S into two parts

> The absence of seismic events in some places of the Groningen field might be the result of varying friction parameters.

$$\Delta P_s = \Delta P_e(S_0, f_s) + \Delta P_a(W_f) \qquad (3.3)$$

where ΔP_e is the depletion pressure drop that is accommodated elastically and finally leads to the onset of aseismic fault slip, and ΔP_a is the depletion pressure drop that occurs during aseismic fault slip until the initiation of the seismic rupture. ΔP_e is found to be mainly dependent on the initial stress condition S_0 and the static friction coefficient f_s. In particular, ΔP_e increases with the ratio between horizontal and vertical in situ stress, $K_0 = \sigma_h / \sigma_v$ (Buijze et al., 2019). In contrast, ΔP_a is mainly influenced by the slope W_f of the slip-weakening relation. Based on this result, van den Bogert (2018) speculates that the absence of seismic events in some places of the Groningen field is a consequence of specific slip-weakening parameters, in particular:

  i) a large value for the static friction coefficient,

  ii) a relatively small absolute value of the dynamic reduction of the friction coefficient,

  iii) or a small value for W_f such that the critical slip length L_c is larger than the reservoir thickness.

Based on these and similar considerations, Dempsey and Suckale (2017) developed a stochastic approach to account for variable friction properties in their seismogenic model for Groningen (see section 3.3.2) which is as well used by DeDontney and Lele (2018).

Despite the simple two-dimensional model setup with planar layers and a straight (one-dimensional) fault, the model already shows complexity in the rupture style with three different rupture mechanisms identified by van den Bogert (2018): one associated with instability of a single slip patch, one associated with the merging of two aseismic slip patches and the third associated with the merging of the two slip patches after one of the slip patches has become unstable. The source-time function

is distinctly different for each rupture mechanism, which may provide the possibility to associate observed seismic events with these three rupture mechanisms. van den Bogert (2018) estimated the corresponding moment magnitudes assuming that the slip distribution in strike direction is similar to the simulated one in dip-direction. He found that the estimated moment magnitudes for all three rupture mechanisms are in the range between 0.5 and 4, although that seismic rupture of a single slip patch tends to yield a smaller moment magnitude. Thus, the simulated event magnitudes are in agreement with observations. Furthermore, all three rupture mechanisms were shown to occur for any reservoir thickness and offset if the fault strength is in an appropriate range and they all can be described by universal scaling relations between moment magnitude, rupture length (L) and maximum relative slip (RSD), i.e., $L(M) \sim \exp(1.01M)$ and $RSD \sim \exp(1.35M)$. Assuming the validity of this model, the relations allow the estimation of the local L and RSD values in the Groningen field based on observed moment magnitudes.

While the model of van den Bogert (2018) and Buijze et al. (2017, 2019), represents an improvement to the model of van Wees et al. (2017a) by accounting for fully dynamic rupture nucleation, propagation, and termination, the study still suffers from a few shortcomings and simplifications, in particular:

- The study uses only a 2D instead of 3D modelling environment. A 3D modelling approach was deemed inappropriate, because either the element size would be too large or the computational costs would be too high.

- Non-linear and time-dependent behaviour of the reservoir and overburden formations (Sanz et al., 2015) have not been taken into account. However, these may be relevant, because of the salt body overlying the Groningen gas reservoir.

- Thermal aspects and (un)drained behaviour of the fault zone and adjacent formations are ignored, which may play a role for earthquake nucleation in the Groningen field.

Wentinck (2018a,b,c) used recorded ground motions from accelerometers to constrain the rupture process of the two largest earthquakes in Groningen. In particular, Wentinck (2018c) performed a kinematic inversion of the M_L3.6 Huizinge earthquake of 16th August 2012 based

> Some insights gathered from kinematic and dynamic source modelling of the two largest earthquakes in the Groningen field.

on recordings of six accelerometers. This earthquake occurred close to a junction of two faults with small to moderate throw (less than 30 m) and the slip plane was constrained to one of the two faults, where a modified Brune source time function was considered for each patch on this plane, embedded in a simplified version of NAM's seismic velocity model. Wentinck (2018c) found that the dimensions of the rupture area could not be reliably determined and might require a calibrated model for wave propagation and scattering in the shallow subsurface in the epicentral region. Further, the

study indicated that the observed second peak likely was produced by S-wave reflections in the shallow subsurface instead of the occurrence of a second event as suggested by a precursory study of Wentinck (2017). However, the analysis suffered in general from the large uncertainties related to the small number of recordings and the unreliable orientation of accelerometers.

Wentinck (2018b) applied the same technique to the more recent Zeerijp earthquake on the 8th January 2018 with a magnitude of M_L 3.4. In the meantime, the geophone and accelerometer network had been substantially expanded (Dando et al., 2018), such that the number of records available for analysis improved significantly. The hypocentre was located close to a minor fault in the reservoir with relatively small throw (40 m); its strike corresponds well to the one derived by moment tensor inversion, but not its dip. Ant-tracking of the fault geometry revealed differences to the values saved in the NAM fault database (Wentinck, 2018b). As the rupture plane dimensions were still not well constrained by seismic observations (Wentinck, 2018b), the source time function was extracted employing the empirical Green's function (EGF) method. A uni-directional rupture along fault strike was inferred (Wentinck, 2018b). From corner frequencies and apparent source time function, the rupture dimension was estimated to be less than 400 m and the slip area considerably smaller than 1 km^2, hinting to a large breakdown stress drop during rupture of 3-5 MPa (Wentinck, 2018a). However, it is well-known that such stress drop estimates involve large uncertainties.

Wentinck (2018a) also implemented a dynamic rupture modelling of the 2018 Zeerijp earthquake, constrained by the available recordings. In contrast to kinematic inversions, dynamic source inversions depend on the assumed pre-stress conditions, which was inferred from reservoir compaction and pressure diffusion into the Carboniferous underburden. Additionally, main input parameters were the static and dynamic friction coefficients as well as the so-called critical breakdown relative displacement of the employed slip-weakening model of Ohnaka (2013). Before rupture, the shear stress was estimated to be highest where the reservoir-anhydrite horizon is juxtaposed to the reservoir rock on the other fault side. However, a potential interplay with a reverse fault close to the hypocentre was disregarded. No evidence of substantial fault propagation into the underlying Carboniferous could be observed indicating a considerably higher static or residual friction coefficient of the Carboniferous compared with Rotliegend rocks. In addition, stress changes due to pressure depletion in the uppermost part of the Carboniferous turned out to be important (Wentinck, 2018a). The small peak rupture velocity implied, according to his modelling, that outside the nucleation patches of the earthquake, fault stress conditions were far from critical (Wentinck, 2018a).

While slip-weakening friction laws are often employed for simplicity as in the applications discussed above, laboratory experiments indicate that the force of resistance depends on the relative displacement velocity as well as the state of the fault surface after a previous slip event. Such a rate-and-state friction law was among oth-

DeDontney and Lele (2018) combined a quasi-dynamic model with rate-and-state friction.

ers explored by DeDontney and Lele (2018), who used three different physical models to analyse the impact of production fluctuation on the seismicity and above all on the seismic hazard (**basicRes+quasidynRup-RS**). First, the distance range influenced by daily and seasonal production was investigated with an analytic reservoir pressure model. Secondly, they used a rate-and-state earthquake activity rate model (Dieterich, 1994) to compare the seismicity rates for different production schemes similar to the models described in section 3.3.2. As third approach, they examined if the hazard at Groningen changes for different production schemes employing an earthquake cycle model. The latter considered the stress evolution on one fault described in 1.5D being subject to rate-and-state friction in a quasi-dynamic model to create earthquake catalogues.

The analytic model assumed a radial symmetric permeable medium around a production well. Different permeabilities were tested and daily as well as seasonal variations were simulated as sinusoidal changes of 10% and 40% amplitudes, respectively. Both amplitude ratios are taken from observed values. Modelled pressure variations were localised in distances up to 300 m around the well and were amplified at the well for lower permeability. A test of this model using a data set of seven months from a production well in the Groningen field with an additional depletion trend of 0.3 MPa/a showed localised pressure variations up to 500 m from the well and a dominant yearly depletion signal. The result of this approach was that the daily and monthly variation had only local influence with moderate amplitudes close to the wells, which applies only to a few faults. These findings are not completely supported by the full numerical dynamic model of NAM, which shows significant annual rate changes for the whole field (see also visco-NAM and NAM-RS model).

The second approach was a test of the influence of production fluctuations on seismicity rates modelled with a rate-and-state earthquake activity rate model. The model was based on the formula from Dieterich (1994) and seismicity rates were calculated for different stressing rate scenarios testing model parameter sets derived from observations. The stressing rates were again simulated as sinusoidal signals with 10% daily and 40% seasonal variation with a mean of 1 MPa/a over a period of 8 to 10 years. The resulting earthquake rates were compared to earthquake rates arising from a constant stressing rate. The background earthquake rate r was taken from the average observed seismicity rate in the last years for events with $M \geq 1.5$ and $M \geq 1.0$ (25 events/a and 90 events/a, respectively). According to this model setup, the earthquake rates showed no daily fluctuations, but clear seasonal changes. For the lower background earthquake rate of 25 events/a, an increase of 0.5-1.5 events per year was found, whereas for the higher background rate of 90 events/a, the increase was around 6 events per year. These findings can explain why the earthquake catalogue for $M \geq 1.5$ shows no clear seasonal variations, whereas for smaller events a clear seasonal signature is found. Overall the number of events is too small to come to statistically stable conclusions.

The third approach was investigated with an earthquake cycle model to examine if the hazard at Groningen changes due to production fluctuations. This model considered the stress evolution on one fault with a rate-and-state friction to derive slip velocities on fault patches similar to Lapusta

et al. (2000), and then build earthquake catalogues. The quasi-static stress changes on a fault in an elastic half space due to uniform slip over each rectangular fault patch was calculated. The stiffness matrix k controlled the shear stress caused by the slip and a dip-slip motion was assumed. This model is "quasi-dynamic" (Rice, 1993) including instantaneous stress transfer, but employing a radiative damping term for the energy flow that is proportional to the sliding velocity. The faults were modelled as 1.5D, meaning that the rough planar faults were examined in either strike or dip direction. For the model in strike direction uniform stress changes were modelled, whereas in dip direction non-uniform stress changes were modelled with the finite element software ABAQUS to include the effect of fault offset. The evolution of the state variable θ was described by two laws: the aging law (Dieterich, 1994)

$$\dot{\theta} = 1 - \frac{\theta v}{L_f} - \alpha \cdot \frac{a\theta}{b\overline{\sigma}} \cdot \dot{\overline{\sigma}} \tag{3.4}$$

or the slip law (Ruina, 1983)

$$\dot{\theta} = -\frac{\theta v}{L_f} \cdot ln\left(\frac{\theta v}{L_f}\right). \tag{3.5}$$

Here, L_f is the characteristic slip distance that is determined according to the cell size h, v is the sliding velocity, α is a constitutive parameter and the parameters a and b are experimentally derived coefficients describing nonlinear friction (for details see section 3.3.2 and equation 3.16). The initial shear stress heterogeneity was assigned following the procedure in Dempsey and Suckale (2016), where the shear stress is consistent with a fractal model.

The earthquake cycle model depends on several parameter assumptions and the influence of nine parameters was tested. These were the initial shear stress heterogeneity, the normal stress, the governing law (ageing or slip), the degree of rate weakening, the critical patch size,

> Daily and seasonal production fluctuations have only a local impact around the production wells and no change in the seismic hazard is found.

the minimum velocity threshold, the strike and dip geometries of the faults, position-dependent stress evolution and different fault loading scenarios. Aggregate catalogues are created consisting of 91 individual catalogues to obtain the common character of the seismicity. For these aggregated catalogues all tested parameters showed that the seismicity rates vary for fluctuating production rates and different parameters, but the cumulative number of events did not change significantly nor did the maximum magnitude for the tested time period. But after the majority of the fault patches had slipped at least once, the initially heterogeneous shear stress levels became relatively uniform over the slipped areas resulting in regular, large events. This loss of stress heterogeneity is not realistic and thus, the latter time period was not considered. Thus, there were two confining factors that questions the result:

- The flattening of the stress heterogeneity on the fault limits the time period of interpretation.

- The test of only a few fault dip and strike angles can not fully represent the highly heterogeneous fault system at Groningen as well as fault interaction.

The role of fluid flow has been addressed by Zbinden et al. (2017), who used a fully coupled fluid flow and geomechanics 2D simulator, which accounted for stress-dependent permeability and linear poroelasticity (**basicRes+flow**). Results showed that fluid flow played a major

> Fluid flow and the effect of salt embedded in fault zones can alter the forecasts significantly.

role in pore pressure and stress evolution within the fault. Fault strength was significantly reduced due to flow into the fault zone from the neighbouring reservoir compartment and other formations. In the case of well shut-in, a highly stressed fault zone could still be reactivated several decades after production had ceased, although on average the shut-in resulted in a reduction of seismicity.

Salt intrusions in the faults might affect dynamic ruptures significantly, because they may introduce a rate-dependency to fault movement and change the magnitude distribution of seismic events (b, M_{max}). Although there are currently no subsurface data indicating the existence of salt-filled faults, this might only be due to technical limitations. By means of analogue models, Kettermann et al. (2017) demonstrated that salt from the Zechstein formation may flow down-dip into opening faults due to gravitational flow. In particular, the authors analysed sandbox experiments with cohesive powder as analogue for brittle anhydrites and carbonates with viscous salt analogues to explore the developing fault geometry and the resulting distribution of salt in the faults. The analogue models also showed lateral salt flow filling jogs in deeper parts of the fault. Based on these results, Kettermann et al. (2017) analysed the effects of salt intrusions in a numerical discrete element model (**basicRes+salt**) approach able to simulate the mechanical behaviour of brittle–elastic and granular materials including ductile material behaviour (Abe and Urai, 2012). The simplified model consisted of a rough planar (1D) fault embedded in a 2D elastic medium. Salt was distributed in a patchy pattern within the fault zone of this model. Other fault zone processes which likely play a role, such as fault healing and rate-and-state friction, were not considered. First model results demonstrated that the behaviour of the fault becomes dependent on the shear loading rate due to the salt intrusions. This also resulted in a modification of the frequency–magnitude distribution of the generated seismic events. However, due to computational costs, the analysis could not explore larger parts of the parameter space of fault geometry, salt rheology and loading rate. Thus, the complex interplay between stick-slip dynamics of a fault and viscoelastic behaviour of salt contained within remains to be studied further.

Advantages and disadvantages of physical models are summarised as follows:

+ The conditions for rupture propagation into the over- or under-burden can be analysed, which potentially constitutes input into estimates of the maximum possible magnitude of triggered earthquakes.

+ Effects of stress concentrations, fluid flow and complex rheology can be addressed.

− Deterministic physical models are highly specific and only derivable for individual studies of earthquakes.

− Deterministic modelling of rupture nucleation and propagation comes with high computational costs.

− Unknown initial conditions and structural/material properties prevent a direct application to seismic hazard assessment.

3.2   Machine Learning (ML) approach

ML is a branch of statistical computer science, which has been applied successfully in a wide variety of domains over the last decade (Jordan and Mitchell, 2015). ML consists of a collection of algorithms and statistical models that enable computers to extract relevant patterns and information from large data sets. Unlike physical modelling approaches in which scientists develop theories based on physical laws to compare with real-world observations, ML approaches learn directly from data without explicitly reasoning about the underlying physical mechanisms (Bergen et al., 2019). However, the outcome depends on the selection of input parameters and features, the learning strategy and testing procedure as well as the setting of meta-parameters. The latter includes target quantity (e.g. earthquakes exceeding a preset magnitude cutoff), spatial and temporal grid size and choice of the ML algorithm, such as e.g. Decision Trees, K Nearest Neighbours, Support Vector Machines or Neural Networks. The setup of an appropriate ML system for forecasting is therefore crucial. An inappropriate use, such as in the recently published case for aftershock forecasting (DeVries et al., 2018), can lead to misleading results as demonstrated by Mignan and Broccardo (2019) for the same case. As every purely statistical method, ML should not be used to extrapolate behaviour outside the parameter range of the test dataset.

Limbeck et al. (2018) and Lanz et al. (2019) started to explore ML as an alternative approach for forecasting production-induced seismicity in the Groningen field. Both studies are closely related, since the later is an extension of the methodology developed in the former study. While the model of Limbeck et al. (2018) is lim-

> Unlike physical modelling approaches, ML approaches learn directly from large data sets without explicitly reasoning about the underlying physical mechanisms.

ited to a cumulative prediction of the earthquake numbers in the whole field, Lanz et al. (2019) extended the model for predictions in time and space. Furthermore, the former study did not reproduce the strong decrease in seismicity after shut-in in the Groningen reservoir, and showed a physically unlikely behaviour of an elevated activity in the very long term after production stopped. In the extended model, the last shortcoming is avoided by introducing artificial ultimate states (with

zero activity for the case that all time-derivatives of the features are zero).

The authors used a large set of potential predictors (features) created from all available data sources, including earthquake event catalogues, static, dynamic and geomechanical reservoir model data, fault information and seismic data. In particular, they used initially 45 different features, which were finally down-selected to 22 features following cross-correlation threshold filtering and grouping; the particular quantities are provided in tables 12 and 17 of Limbeck et al. (2018). These features were calculated on a spatial grid with a spacing of 1.5 km. To account for the freedom in the meta-parameter selection, they analysed 300 different combinations, including different time periods (3, 6 or 12 months) and different spatial bandwidths for spatial kernel smoothing for feature generation as well as 5 different ML algorithms. For each of the meta-parameter sets, the ML-model was trained only with data recorded before the year 2013 and the hold-out set of observations between 2013 and 2016 were used for testing the forecast power of the models. This is necessary due to potential over-fitting of the training data, which leads to a better performance in training data than for a hold-out set. For the test data, the mean absolute error (MAE), the root mean square logarithmic error (RMSLE), and the mean Poisson loss (MPL) were determined and compared relative to two baseline models. The first assumed that the forecast rate was proportional to the change of depletion thickness, i.e., pressure multiplied by reservoir thickness; the other considered a proportionality to the change in strain thickness, i.e., vertical strain multiplied by reservoir thickness. The proportionality factor was calculated field-wide from the observations in a look-back time period.

The analysis showed that the most significant spatial features appear to be the topographic gradient of the reservoir, reservoir thickness variations along the major faults, and compressibility, which are features that are likely to impact seismicity from the view of the physical Coulomb strain model, particularly the NAM model (Bourne and Oates, 2017b). Overall, however, it was found that each of the static features (such as fault properties) have less impact than any of the dynamic ones, such as pore pressure. The results also demonstrated that ML models can qualitatively capture the negative trend in seismicity observed during the 2013-2016 hold-out period, like the baseline models. However, although there was a relative match in trends, the models systematically underestimated seismicity, meaning that extreme values in seismicity rates are not easily captured given the information and models available at the moment. In comparison to the baseline models, the ML-algorithms were found to be systematically better only with respect to the MAE value, while the RMSLE and MPL measures showed improvement only for parts of the ML-models. Nevertheless, the Random Forest (RF) algorithm yielded the best performance in all cases and a significantly better result than the two baseline models. The second best was the Support Vector Machine with kernel function (KSVM), which also showed significant improvements.

Based on the comprehensive analysis presented by Limbeck et al. (2018) and Lanz et al. (2019) the advantages and problems can be summarised as follows:

+ ML can extract relevant patterns and information from large and multivariate data sets of potential predictors (features) recorded in the Groningen gas field.

+ Unlike physical modelling approaches, ML approaches learn directly from data without explicitly assuming any underlying physical mechanisms as the NAM model and thus can avoid model biases due to incomplete process understanding.

− The ML-models' accuracy and forecasting power remain limited by the small training data set with regard to the observed number of earthquakes.

− If the pressure development in the long term does not fall within the pressure range used in the training data set, the ML-model will likely lead to wrong forecasts.

− Important features are based on the results of the static and dynamic reservoir models, which are based on reservoir-flow history matches and forecasts. Thus, the quality of the ML-model depends on the quality of these input data, similarly to the NAM model.

− Confidence bands for the forecasts cannot be easily derived for the space-time ML model, in contrast to the NAM model.

− A separate magnitude model is required to complement the current spatiotemporal model. This could be derived by a non-parametric ML-approach or by assuming a Gutenberg-Richter distribution as done for the NAM model.

3.3 Statistical and hybrid models

Despite the extraordinary data sets available for the Groningen field (e.g., geodetic measurements, drill cores and seismicity data), the information about the stress state and material properties are limited, which prevents a fully deterministic approach for seismic hazard assessment. Thus, models accounting for these unknowns are required. In this context, two fundamental models play a dominant role in statistical seismology: the Poisson and the Gutenberg-Richter models.

Independent of the spatial scale, region and the tectonic environment, the distribution of earthquake magnitudes m has been found to have a similar shape and can be well approximated by the Gutenberg-Richter (GR) distribution (Gutenberg and Richter, 1956). The magnitude probability density function (pdf) related to the GR law is given by

$$pdf(m) = \frac{b \ln(10) 10^{-b(m-M_c)}}{1 - 10^{-b(M_{max}-M_c)}} \tag{3.6}$$

for m distributed between M_c and M_{max}, where the b-value is typically close to one. From a physical point of view, this distribution must have an upper and lower bound (doubly truncated) due to finite fault length and grain size, respectively. However, the physical lower bound is usually not observable, because of the limited detection capability of the seismic networks, which is related to the completeness magnitude M_c. The upper bound M_{max} is difficult to estimate from earthquake data

alone (Zöller and Holschneider, 2016a), but can be deduced from other observations such as the fault length distribution or the seismic moment budget (Schmedes et al., 2005). As alternative to a sharp truncation at M_{max}, a tapered distribution at large magnitudes is also often used (Jackson and Kagan, 1999).

The Poisson process is known as the model of complete randomness. In a stationary Poisson point process, events are independent of each other and the average rate λ is constant. The corresponding interevent-time distribution is then given by an exponential distribution, and the probability to observe a number of N events in a time interval Δt is given by

$$Pr(N) = \frac{(\lambda \Delta t)^N}{N!} e^{-\lambda \Delta t} \tag{3.7}$$

The probability to observe at least one event is calculated by $1 - e^{-\lambda \Delta t}$ (e.g. Zhuang et al., 2012). Most of today's probabilistic seismic hazard analyses are based on this model (Cornell, 1968). In an inhomogeneous Poisson process, events are still independent of each other, but the average rate changes over time. In this case, the following, more general equation applies (Zhuang et al., 2012)

$$Pr(N) = \frac{(\int_t^{t+\Delta t} \lambda(s)ds)^N}{N!} e^{-\int_t^{t+\Delta t} \lambda(s)ds} \tag{3.8}$$

which is, e.g., used in the NAM model. All statistical models with relevance for the Groningen case, which are described below, are based to some extent on these two fundamental models for seismicity.

3.3.1 Purely data-driven models

Nonparametric kernel smoothing models (K2/K3)

The nonparametric kernel smoothing models represent statistical end members on the spectrum of statistical seismicity models. None of the usual assumptions about earthquake clustering are explicitly included such as the Omori-Utsu law (see Eq. 3.9) or the Utsu–Seki clustering law (large earthquakes generate exponentially more aftershocks). Instead, the models use Gaussian kernels to estimate seismicity as a function of time, space and magnitude (Helmstetter and Werner, 2012; Helmstetter and Werner, 2014).

In standard seismic hazard assessments, adaptive smoothing is well-known for determining the homogeneous spatial earthquake distribution. In this case, the bandwidth (standard deviation) of the Gaussian kernel is set individually for each earthquake in the data set; in particular, it is proportional to the distance to the nth nearest neighbour, where the value of n is a model parameter. This concept can be extended to space-time (K2) and even to space-time-magnitude (K3)

> Nonparametric kernel smoothing is a flexible tool relying only on observed seismicity, while ignoring any empirical or physical laws.

as demonstrated by Helmstetter and Werner (2012, 2014). Whereas model K2 assumes additionally a parameterised distribution for the magnitudes, specifically the Gutenberg–Richter magnitude distribution with $b = 1$, model K3 uses kernels to estimate the (space–time-dependent) magnitude distribution. The widths of the kernels adapt to the activity level: sparse seismicity (in space and time) widens kernels; concentrated seismicity narrows kernels. The models thereby adjust to the current seismicity rate, which is extrapolated over the forecast horizon. These nonparametric kernel models offer maximum flexibility, at the cost of ignoring commonly observed empirical laws. Although the smoothing method is simpler and requires fewer parameters than other statistical models as e.g. the ETAS model, Helmstetter and Werner (2014) demonstrated for the application to seismicity in California that the obtained probability gains are surprisingly close to those of the ETAS model described in the next section.

To our knowledge, the space-time-magnitude smoothing methods have yet not been applied to induced seismicity. However, for short term forecasts, they might be a useful alternative in the context of ensemble forecasts, because this model class does not rely on any physical or empirical law, which might be not valid in the specific application.

The advantages and problems of implementing adaptive smoothing models to the Groningen case can be summarised as follows:

+ The models automatically adapt to the existing spatial, temporal and magnitude resolution of the input earthquake catalogue.

+ The models do not rely on any empirical or physical laws that might be wrong, biased, or incomplete.

– The forecasts are limited to short time scales, because the forecasted rate is a decaying function of the past activity.

– Forecasts are not meaningful if production will strongly vary in the forecast period.

ETAS models

Aftershock triggering following large earthquakes is ubiquitous in natural seismicity. Most aftershocks occur close to the mainshock rupture with an occurrence rate R, which can be well described by the Omori-Utsu law

$$R(t) = K \cdot (t + c)^{-p} \tag{3.9}$$

where t indicates the elapsed time since the mainshock (see Utsu et al., 1995, for a review). The parameter K is known to depend on the mainshock magnitude M_{main}, while p is typically in the range 0.8-1.2 and independent of M_{main} (Utsu et al., 1995). The time-offset parameter c is generally much less than one day, and is usually related to reduced detection ability of the operating seismic network immediately after large events (Kagan, 2004; Hainzl, 2016).

Large aftershocks trigger their own local aftershock sequence. To account for this secondary triggering, the Epidemic Type Aftershock Sequence (ETAS) model has been developed. It is a stochastic point process model that builds on the Omori-Utsu law and also takes stationary background seismicity and secondary aftershocks into account (Ogata, 1988; Ogata, 1998). In the ETAS model, each earthquake with occurrence time t_i, location $\vec{x}_i$, and magnitudes m_i has a magnitude-dependent ability to trigger aftershocks according to $K_0\, 10^{\alpha(m_i - M_c)}$ (Utsu–Seki law), where α and K_0 are constants and M_c is the lower magnitude cut-off of the earthquakes under consideration (which must be equal or larger than the completeness magnitude). In this model, the total rate is the sum of background seismicity and ongoing aftershock sequences triggered by all past events. The total occurrence rate of earthquakes is given at time t and location $\vec{x}$ by

$$R(t, \vec{x}) = \mu(\vec{x}) + \sum_{i:t_i<t} \frac{K_0\, 10^{\alpha(m_i - M_c)}}{(t - t_i + c)^p} g(|\vec{x} - \vec{x}_i|) \tag{3.10}$$

with μ being the background rate and $g(r)$ a normalized spatial kernel, typically given by a power-law decay as function of distance. A necessary condition for stability of forward simulations is that the aftershock sequences decay sufficiently fast, namely with $p > 1$. Otherwise, the total number of aftershocks would become infinite in the long term and seismicity would escalate with time. The ETAS model represents an inhomogeneous Poisson model, where the local event rate depends on the earthquake history (Zhuang et al., 2012). For forward simulations, an independent frequency-magnitude distribution has to be assumed. For this purpose, the GR distribution (Eq. 3.6) is generally used.

It is important to note that the model rate depends in a nonlinear way on the model parameters and thus ETAS forecasts sensitively depend on the values of the ETAS-parameters. The branching ratio n, defined as the average number of triggered aftershocks per event, depends on the four triggering parameters K_0, α, c and p, as well as on the parameters of the frequency-magnitude distribution b, M_c and M_{max} (Helmstetter and Sornette, 2002). If the branching ratio is larger than one, the simulations are unstable and will lead to an escalating predicted event rate with time. However, fits of the ETAS model to empirical data are found to often yield parameter estimates that correspond to an overcritical branching ratio, $n > 1$ (Helmstetter and Sornette, 2002). The reasons are likely too small sample sizes or catalogue lengths, catalogue incompleteness or inappropriateness of the ETAS model to describe the observed data.

The ETAS model has been developed for natural seismicity, which is driven by tectonic plate motion. In this case, the stress loading due to external forcing can be assumed to linearly increase in time with a constant stressing rate. Thus, the assumption of a constant background rate μ, combined with aftershocks resulting from earthquake-induced stress changes, is reasonable. However, even at plate boundaries, transient aseismic forcing such as fluid flow (Vidale and Shearer, 2006) or slow slip events (Holtkamp and Brudzinski, 2011) are frequently occurring, leading to transient aseismic stress-changes and triggered seismicity, which is not accounted for in the standard ETAS model. In

particular, the assumption of a constant driving force becomes obviously inadequate in the context of anthropogenically induced seismicity. For that reason, Hainzl and Ogata (2005) modified the ETAS model to account for a time-dependent background forcing term $\mu(t)$, which can be estimated from earthquake catalogue data. An ETAS-model with an additional time-dependence of the triggering parameters was applied by Llenos and Michael (2013) to induced activity in Oklahoma and Arkansas. More sophisticated methods were introduced later to optimise the fitting procedure of the time-dependent parameters (Kumazawa and Ogata, 2014; Marsan et al., 2013).

A fundamental problem of forecasting future seismicity with the modified ETAS model remains in the unknown development of time-dependence of $\mu(t)$ and other parameters. For the purpose of seismic hazard assessments of earthquake swarms, USGS scientists currently investigate the use of empirical models for the duration of swarms (Llenos and van der Elst, 2019). On the other hand, they use the assumption of slowly changing forcing rates and estimate the future background rate by the extrapolation of the inverted one in the recent past (Llenos and Michael, 2019). The latter is similar to the adaptive smoothing approach (see subsection 3.3.1), but with distinction between externally-triggered seismicity and earthquake-induced aftershocks.

Instead of a purely empirical approach, the future forcing term can also be estimated based on production data in the case of anthropogenically induced activity. For the year 2006 stimulation of a reservoir for an Enhanced Geothermal System (EGS) in Basel, Bachmann et al. (2011) assumed that the background rate $\mu(t)$ was simply

> It is straightforward to combine the ETAS model with any deterministic stress model, assuming that the background rate is simply proportional to the stressing rate.

proportional to the injection flow rate, in agreement with the assumption of the Seismogenic Index model (see below). They found that this model performs best in comparison to other statistical models. While a very simple approach was used in this case, the same approach can in principle be based on more sophisticated forecasts of modelled induced stresses. Assuming a simple model, the so-called *Coulomb failure model*, where earthquakes are instantaneously triggered if stress exceeds a critical threshold and pre-stress (criticality values) is uniformly distributed, the induced activity rate is proportional to the stressing rate (Hainzl et al., 2010a). The background rate of the ETAS model can thus be assumed to be simply proportional to the calculated stressing rate $\dot{S}$ with $\mu(\vec{x}, t) = \kappa \dot{S}(\vec{x}, t)$, where the proportionality factor κ is a free model parameter (instead of μ), which can be estimated from the past activity. In this way, it is straightforward to combine the ETAS model with any deterministic stress model. The most recent NAM model by Bourne et al. (2018) represents such a combination. However, in this case the background term is not simply proportional to the stressing rate, because the pre-stress is not assumed to be uniform. Instead, the failure stress is assumed to be in the tail of the pre-stress distribution, which leads to the application of the extreme value theory (for details please see WP1, 2019).

The advantages and problems of an ETAS application to the Groningen case can be summarised as follows:

+ The ETAS model is an empirical model which can be simply fitted to earthquake catalogues.

+ It is successful in reproducing the most conspicuous spatiotemporal patterns of natural seismicity on time scales from minutes to decades.

+ It can be extended to account for time-dependent aseismic forcing.

+ The modified ETAS model can be easily combined with production-based stress models.

– If only based on catalogue information, the forecast power is limited due to the unknown future forcing.

– The ETAS model depends non-linearly on the model parameters which leads to high sensitivity of the forecasts to the parameters estimated from the data. Thus, the input catalogue needs to include a sufficient number of aftershocks, otherwise the parameters are not well constrained and forward simulations can become dubious. This necessary data requirement likely is not fulfilled for the Groningen case, where only approximately 5% of the activity is classified as aftershocks (Muntendam-Bos, 2020); see further discussion in Sec. 5.

– The approach needs an additional model for the magnitude probability distribution, where typically a GR distribution is assumed (Eq. 3.6).

Models based on empirical relations between seismicity and compaction (NC-model)

A number of different functional relations between the local compaction of the reservoir and the cumulative number of triggered earthquakes has been tested by DeDontney (2017). In particular, the author tested linear, quadratic, cubic and exponential relations, where the maximum likelihood method and the Akaike Information Criterion (AIC) were used for data fitting and model comparison, respectively. To explain that no earthquakes occurred in the beginning of gas production, a threshold was imposed in the models as either a year or a deformation value, so only deformation that occurred above this threshold was considered in generating the statistical model. As pointed out by DeDontney (2017) it makes conceptually sense that a large amount of compaction has an effect, but a small amount is negligible (a threshold value), but it is hard to elaborate an argument about how everything changed during just one year, especially given the amount of variability in compaction across the field in any given year. Nevertheless, regardless of the physical intuition, a reference year was found to be consistently the better option (DeDontney, 2017).

Based on the fault slip model of Lele et al. (2016) based on deformation, DeDontney (2017) tested the same set of relationships between seismicity rate and fault slip instead of compaction. However, fault slip models were only available in sub-regions of the Groningen field. Furthermore, a spatial smoothing was required to account for the uncertainties of the epicentre locations, which seldom

can be directly associated with a given fault. Several kernel widths were tested with the result that larger Gaussian kernels (standard deviation of 7 km) were found to yield better results. For the corresponding seismicity forecasts based on linear, quadratic, cubic or exponential relations, a temporal threshold was found to provide better results compared to a deformation threshold, similar to the results based on compaction data.

In the area where a fault slip model exists, the fault slip based metric was found to yield a significantly better representation of the observed seismicity than the compaction-based models. The best fault slip-based seismological model forecasts lower activity rates over the next 30 years than a compaction-based model (using an exponential relationship), similar to what was previously used by NAM (Bourne et al., 2014). However, because the fault slip model does not exist in the entire domain of the field, currently only compaction can be used to develop a field-wide activity model. Among the variety of functions which were considered, the model based on an exponential relationship, i.e., similar to Bourne et al. (2014) was found to be one of the better models. However, many models fitted the observed seismicity almost equally well, but these models yielded very different activity forecasts over the next 10 years, which highlights the large epistemic uncertainties of predictions based on empirical functions.

The advantages and problems of this approach can be summarised as follows:

+ The approach represents a fully data-driven approach to find the functional relationship between seismicity and compaction or fault slip. Thus, it avoids any bias related to an incomplete understanding of the earthquake triggering mechanism.

+ The analysis shows that non-linear relationships between compaction (fault slip) and seismicity lead to comparable or best fits of the earthquake history; in particular, an exponential relation as used by NAM (Bourne et al., 2014) is found to be well suited.

– The similar data fits for the observed earthquake data show that the observed data set is not large enough to discriminate between different models. Thus, the proposed purely data-driven selection of the functional dependence leads to huge uncertainties with respect to future seismicity.

– The model yields the best results when a threshold in time (instead of a deformation value) is used, which cannot be simply explained and might indicate a physical inconsistency of the approach.

Models based on empirical relations between seismicity and extraction volume (NV-model)

The empirical relationship between the cumulative number of seismic events and cumulative gas production can be exploited to predict the expected number of future earthquakes based on production plans. In combination with the Poisson model and the GR distribution, this approach yields a

probabilistic forecast of earthquake activity.

Hettema et al. (2017) showed that the cumulative number of triggered earthquakes accelerates as function of the cumulative volume extracted in the Groningen field. By normalising the earthquake numbers in time bins by the extracted gas volume in the same time bins, the authors recognised that the number of triggered earthquakes per extracted unit of gas volume

> The empirical observation of a quadratic dependence of the cumulative earthquake number on the cumulative extracted gas volume can be exploited for forecasting future earthquake activity.

has linearly increased as function of the cumulative gas production since the time of the first recorded earthquake in December 1991. A linear regression fit yields variance reductions of approximately 70%. Vlek (2019) used the same approach and observed a variance reduction of even 90% for two-year bins.

Based on this simple approach, the mean number of earthquakes can be predicted for any future time interval, given a production scenario. For that purpose, Hettema et al. (2017) and Vlek (2019) assumed a Poisson model. It is noteworthy that this model predicts a fixed number of triggered future events, if the Groningen reservoir is completely exploited, independently of the production scenario. Lowering the annual production would only postpone earthquakes, but the cumulative number of events at the end of depletion would be fixed regardless of the way in which the field is depleted. This prediction assumes that the empirical relationship holds in the future.

Assuming a GR magnitude distribution, the probabilities for any magnitude range can be easily estimated for any time interval. In particular, the maximum expected magnitude can be estimated (Vlek, 2019; Zöller and Holschneider, 2016b).

The advantages and problems of this approach can be summarised as follows:

+ The simple empirical relationship between the cumulative earthquake number and the cumulative extracted gas volume has the advantage that time is eliminated as a variable, and short-term fluctuations (months) are suppressed.

+ It involves only three model parameters (two parameters for the linear regression and one for the GR distribution, namely the b-value).

− The empirical relation is not based on any process understanding, which might question its reliability for the future. For example, the hydraulic diffusivity is known to decrease with increasing gas extraction, which is expected to lead to a delayed earthquake triggering, which is not accounted for in the model (Hettema et al., 2017).

− The approach has been only tested for the entire field without spatial resolution. A first test for a subdivision into different sub-regions of the Groningen field indicated that the empirical

relation is not uniform and varies significantly in space, which limits the model's applicability (Hettema et al., 2017).

3.3.2   Hybrid models

In contrast to previous models described in this section, which are solely based on statistical model components and empirical relations to production data, the class of hybrid models is partly build on deterministic calculations or physical considerations or both, which are finally combined with statistical components. Thus, they aim to account for both the statistical nature of the earthquake process and our physical process understanding. Insofar they are the most promising candidates for seismic hazard assessments.

The NAM model belongs to this model class. Here, we will discuss five alternative models of this type that have been already applied to the Groningen field and elsewhere. In particular, we introduce (i) the Seismogenic-Index model, (ii) the model of Wentinck (2015), (iii) the model of Dempsey & Suckale (2016a, 2016b, 2017), (iv) the viscoelastic model from NAM, and (v) the rate-and-state model with three applications. The first four models implicitly assume that earthquake nucleation occurs instantaneously if stress reaches a threshold value (similarly to the NAM model), while the last model accounts for delayed triggering.

Seismogenic Index (SI) model

The seismogenic index quantifies a seismogenic reaction of rocks to a unit volume fluid change, which is determined by empirical fits of fluid volumes versus triggered seismicity rates. The assumption that the ratio between earthquake number and fluid volume is constant was originally based on physical considerations concerning pore pressure diffusion and Coulomb failure stress (CFS) triggering of earthquakes during fluid injections. The seismogenic index was proposed to forecast seismogenic effects of planned subsurface operations (Shapiro et al., 2010; Langenbruch and Zoback, 2016). It facilitates a quantitative comparison of tectonic conditions and fluid operations in terms of their potential to induce earthquakes and to produce seismic hazard. The higher the seismogenic index, the higher the induced seismic hazard.

The model is based on the GR distribution and the assumption of an inhomogeneous Poisson model, where the time-dependent seismicity rate is proportional to the injection fluid rate $\dot{V}(t)$ at time t. Both assumptions are combined in the frequency-magnitude distribution for the

> The SI model is based on the assumption of a critical stress state and instantaneous triggering due to pore pressure changes.

expected cumulative number $N_{\geq M}(t)$ of earthquakes with magnitudes larger than M induced until

the time t by the cumulative injected fluid volume $V(t)$

$$\log_{10}\left[N_{\geq M}(t)\right] = \log_{10}\left[V(t)\right] + \Sigma - bM \qquad (3.11)$$

Fundamentally, the seismogenic index value Σ incorporates the volumetric concentration of preexisting faults and the state of stress (termed the seismotectonic state). For different areas of induced seismicity, Σ is found to remain constant through time, which supports the physical basis of the model, i.e., that Σ is a characteristic of the seismogenic state of a crustal volume, independent of the anthropogenic perturbation.

Langenbruch and Zoback (2016) calibrated the SI model with a uniform Σ-value to the seismicity rates observed in Oklahoma in order to predict future event rates based on regulatory scenarios. For that purpose, the authors fitted Eq. 3.11 to the observed seismicity data and the injected volumes of waste water in the Arbuckle Group. In order to improve the fit, Langenbruch et al. (2018) based their revised model on a 3-D pore pressure diffusion model for waste water disposal and allowed Σ to vary spatially. This is in agreement with the fact that equation 3.11 has been derived from the understanding that induced seismicity is triggered by pore pressure diffusion in the pore and fracture space of rocks. Particularly, the model was developed for fluid injections into critically stressed fractured rock, but the model can also be assumed to be applicable for other processes such as poroelasticity or static overpressure (Broccardo et al., 2017).

Shapiro (2018) modified the seismogenic index for fluid productions. In particular, he analysed fluid extraction in a thin hydraulically isolated poroelastic layer as an analogue to the Groningen field. Similarly to the previous derivations of the seismogenic index for injection-induced pore pressure changes, he assumed that the number of induced seismic events is proportional to the change of the Coulomb failure stress (ΔCFS) for positive perturbations, while no earthquakes are expected to be triggered for $\Delta CFS \leq 0$. He found that the seismogenic index model for injection can also be adopted for production. His estimation for the Groningen gas field under normal faulting indicates that the combination of the friction coefficient and the poroelastic coupling seems to be unfavourable for inducing seismicity by injection, while it is favourable for earthquake triggering during production.

The advantages and problems of SI model applications to the Groningen case can be summarised as follows:

+ The model is simple (essentially integrating GR and CFS triggering) and has been demonstrated to work well for many injection sites.

+ The empirical model is in agreement with physical considerations of poroelasticity and Coulomb failure stress triggering assuming that the region of interest is in a critical stress state.

– The assumed constant ratio between the number of earthquakes and fluid volume contradicts the observed increase of the ratio with cumulative production volume; see subsection *Models based on empirical relations between seismicity and extraction volume (NV-model)*.

- The applicability to Groningen might be questionable, because the hydraulic diffusivity is expected to change over time due to compaction (Hettema et al., 2017).

- So far, the model has not been demonstrated to fit the (spatial-)temporal distribution of seismicity in Groningen.

- The model does not account for aftershocks.

Wentinck model

Wentinck (2015) analysed a model that assumes a uniform and constant pressure drop in the Groningen and other gas production fields in the Netherlands. Based on simple mechanical considerations, he postulated that the increase of shear stress on faults is only proportional to the pressure drop Δp during interseismic time periods. In contrast, earthquakes are assumed to lead to shear stress drops on faults which are proportional to the seismic moments $M_{0,i}$ of the occurred events ($i = 1, \ldots, N$). The key model parameter is the stress state η, which is given as function of time t by

$$\eta(t) = c_1 \Delta p(t) - c_2 \sum_{i=1}^{N} M_{0,i} \qquad (3.12)$$

where c_1 and c_2 are fit parameters, and $\Delta p(t) = p_0 - p(t)$ is the total stress drop until time t. Note that we have renamed the two fit parameters for simplicity.

The relative likelihood for the occurrence of an earthquake is assumed to be proportional to the probability given by a Weibull probability distribution function, $p_W = 1 - \exp(-\eta^k)$, with the additional model parameter k (shape parameter).

> The Wentinck model assumes unloading due to earthquakes leading to decreased activity after main events.

For justification, Wentinck (2015) pointed out that the Weibull distribution function is frequently applied (i) to describe the lifetime distribution of products and particle sizes, (ii) to analyse point load tests on rock cores, (iii) to describe the heterogeneous distribution of micro-fractures in natural materials, or (iv) to analyse other systems with many similar components (such as a chain). For $\eta \ll 1$, the probability p_W is approximately 0, while p_W converges to 1 for increasing η-values.

The last free model parameter is the proportionality factor λ_{PS}, which determines the earthquake rate described by a Poisson process at high stress levels. The earthquake rate R of the assumed inhomogeneous Poisson process becomes

$$R(t) = \lambda_{PS}(1 - e^{-\eta^k}), \qquad (3.13)$$

where the fit in the different analysed subregions yielded k-values in the range between 6 and 7. Thus, the predicted event rate depends on the absolute stress value, similarly to the NAM model of Bourne et al. (2014) and Bourne and Oates (2017b). In particular, the stress dependence is very
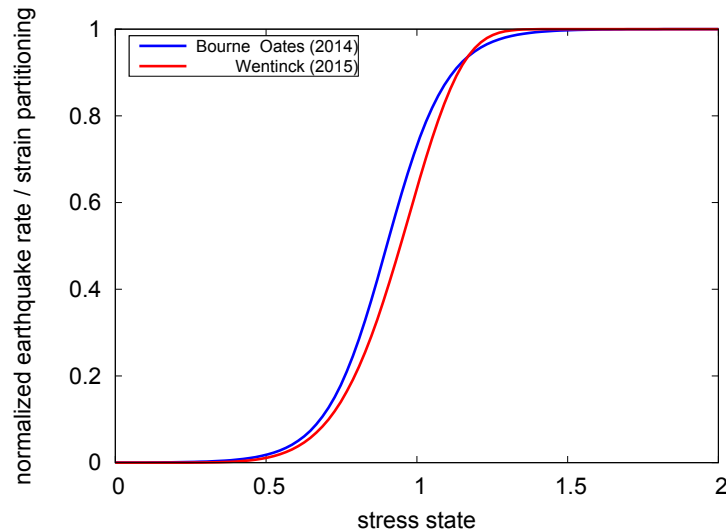
Fig. 3.3 – *The red curve represents the functional form of the normalised earthquake rate, $1 - e^{-x^k}$, using the fit value $k = 6.5$ of Wentinck (2015). For comparison, the shape of the strain partitioning function of Bourne et al. (2014), $\exp(f + gx)/(1 + \exp(f + gx))$ with the fit value $f = -9$, is shown in blue. For the latter, the second parameter is set to $g = 10$ to be on the same scale.*

similar to the strain-partitioning model of Bourne et al. (2014), where it is assumed that only a stress-dependent fraction $\exp(f + g\Delta h)/(1 + \exp(f + g\Delta h))$ of the geodetic deformation is released by earthquakes. Here, Δh represents the absolute reservoir compaction, which is proportional to the pore pressure drop. The comparison of the shape of both stress-dependencies is shown in Fig. 3.3. The two model curves show only minor differences. Both models predict a saturation of the earthquake rate for high stress levels. In contrast the extreme threshold failure model of Bourne and Oates (2017a) predicts an acceleration of the activity with increasing stress. The main difference of the Wentinck-model to the strain-partitioning model is that earthquakes reduce the stress state in the former, which means that the predicted number of earthquakes is reduced after the occurrence of an event leading to a degree of quasi-periodicity in contrast to aftershock clustering assumed by the ETAS-model. Furthermore, after a shut-in, the model still predicts a high number of earthquakes, which only slowly decreases due to the earthquake-induced stress decrease.

The Wentinck-model was applied to six relatively small regions (circular areas with radius of ≈ 5 km) in the Groningen field and in the Annerveen and Eleveld fields south of the Groningen field. The fits yielded almost the same value of k (6 or 7) in all regions, but the parameters c_1, c_2, λ_{PS} were found to significantly vary between regions. Overall, the model was found to fit the observed data in each region well, only a few deviations were observed. One significant difference to the observations was found for the region around Eleveld, where the reservoir pressure reduction has almost completely stopped after 2005. In this case, the modelled event rate was significantly higher than the observed rate after 2005.

The advantages and problems of the Wentinck-model can be summarised as follows:

+ The model is rather simple assuming a Weibull-distribution for the dependence of the activation rate on the absolute stress drop value.

− The model does not account for aftershocks, rather it predicts a decreased activity level after an earthquake. This is problematic, because this is usually not observed.

− After production shut-in, the model predicts an almost constant high activity, which is only slowly decaying due to the stress release by the earthquakes themselves. Thus, this model likely fails to reproduce the observed decline of activity after the slow-down of the production from 2014 onward.

− So far, the model is only tested assuming uniform pore pressure changes with constant rate, which is an oversimplification. It is unclear whether more realistic spatiotemporal pressure data would lead to a good fit of all regions within the Groningen field with uniform model parameters.

− An additional model is needed for the magnitude probability distribution; here, a GR distribution is assumed (Eq. 3.6).

CMsub model

The model of Dempsey and Suckale (2017) is based on the poroelastic earthquake triggering mechanism introduced by Segall and Fitzgerald (1998) and modified by Zoback (2010):

$$\Delta\sigma_h = \alpha \frac{1 - 2\nu}{1 - \nu} \Delta p \qquad (3.14)$$

assuming poroelastic contraction as triggering mechanism with the stress path coefficient $A = \alpha(1 - 2\nu)/(1 - \nu)$, where α is the Biot coefficient and ν is Poisson's ratio. An earthquake nucleates if a critical stress level is exceeded (Mohr-Coulomb failure) over a critical slip distance resulting in a slip-weakening instability (Uenishi and Rice, 2003). Initially, a heterogeneous shear stress profile is created to simulate the frictional behaviour on each fault based on a fractal model and drawing the values according to a positively skewed Gumbel distribution. This is an improvement over the usual normal distribution, since the post-rupture stress changes follow an exponential trend similar to the Omori-type decay. This model is combined with a rupture propagation simulation in 1-D (Dempsey and Suckale, 2016), in which the propagation is calculated from the equation of motion for two tips of an expanding crack (Eshelby, 1969), accelerating if the shear stress is above the critical level (energy source) and slowing down for shear stress below the critical level (energy sink). To account for earthquake statistics and unknown heterogeneities, an ensemble approach for earthquake simulation is chosen. Each simulation is done with different realisations of the initial shear stress profile based on the same stress model and fault. The earthquake magnitudes are calculated according to the rupture length L obtained from the rupture propagation model and the average stress drop assuming a circular rupture area. The resolvable magnitude range is between $1.0 \leq M \geq 6.4$ where

$M = 6.4$ is the magnitude corresponding to a complete rupture of the largest fault and $M = 1.0$ depends on the fault discretization of $dx = 10$m. The authors use a bootstrap sampling procedure to select subcatalogues of the original catalogue that fit a GR model (Dempsey et al., 2016). Overall, the model comprises more than sixteen model parameters. Five of these parameters are free (A, f_S, h, H, Φ) and are calibrated to the observed seismicity with $M \geq 1.0$ in the time period from 1991-2017, assuming an absence of events before. These model parameters are: A the stress path coefficient, f_S the static friction coefficient, $h = \frac{\sigma_3 - \sigma_{crit}}{\sigma_1 - \sigma_{crit}}$ the minimum horizontal stress parameter, $H = \frac{\sigma_2 - \sigma_3}{\sigma_1 - \sigma_3}$ the maximum hor-

> The CMsub model by Dempsey and Suckale (2017) is based on poroelastic contraction with ruptures propagation on 1D faults from Groningen. 10^6 realisation of initial shear stress profiles create the variability of the catalogues.

izontal stress parameter and Φ the angle between σ_3 and north. Based on the model best describing the seismicity from 1991 to 2017, forecasts of seismicity rates and magnitudes are obtained for the time period 2017 to 2024 according to three production scenarios. The size of the earthquakes in the forecast is described by the number of felt earthquakes $N_{2.5}$ with $M \geq 2.5$ and the largest expected magnitude M_T found in the realisations. The pore pressure data are taken from the 2-D NAM model by selecting the local pore pressure at the faults and assuming the changes in the normal stress are negligible in order to obtain the shear stress on the fault. Thus, only local stresses are considered. The critical shear stress σ_{crit} is represented by the normalised parameter h. To confine computational costs, the fault system is restricted to the 325 largest faults represented as 1-D line segments. The key parameters are determined by a Bayesian approach, where the prior value ranges are calibrated in a log likelihood manner to fit the observed seismicity rates. Using a Markov chain Monte Carlo method (MCMC), a parameter set is selected for the creation of a catalogue. The variability of the catalogue is taken into account by producing 10^6 realisations for each production scenario. The model is able to reconstruct the observed seismicity from 1995 to 2017 in terms of rate and magnitude. It forecasts a rapid drop in seismicity rates after production reduction and results in a widely varying range of felt earthquakes with almost the same M_T for the three production scenarios in the time period to 2024.

The advantages and problems of the CMsub model can be summarised as follows:

+ The model combines a deterministic description of the earthquake nucleation and rupture propagation with a probabilistic determination of uncertain parameters and the seismic catalogue.

+ The model parameters obtained from calibration can be interpreted in terms of properties of the reservoir, e.g. the initially not critically stressed faults ($h > 0.06$), or may hint to model limitations regarding the higher than typical stress path coefficient.

− The model is limited to the reservoir layer, because only local stresses are considered.

– The selection of the faults and their orientations is a crucial step in the model procedure, which is not investigated, although only local stresses at the faults are considered.

– The total field seismicity rate is modelled; therefore, no spatial resolution is given by the model.

– Aftershock triggering is not captured in the model.

Viscoelastic NAM model (visco-NAM)

The viscoelastic model presented by Bourne and Oates (2018) is introduced to analyse the influence of seasonal depletion rates on the observed seismicity rates and to compare the results to the NAM model (Bourne and Oates, 2017b). Therefore, the viscoelastic model is implemented in the framework of the thin-sheet model and the extreme threshold failure model to achieve seismicity rates. Additionally, instantaneous creep is assumed because the correlation analysis of the observed event rates with the event rates from the NAM model point to a zero time-lag.

Another assumption is that the material ratio $H(x)/\eta(x)$ is constant, meaning that the spatial variations in the elastic modulus $H(x)$ mirror the variations in viscosity $\eta(x)$. This model is opti-

> The viscoelastic NAM model (Bourne and Oates, 2018) combines the thin-sheet and extreme threshold failure model with a viscoelastic coupling between stresses and seismicity rates.

mised with the same Bayesian inference as the NAM model having one more parameter, θ_2. The time series have a 1-week sampling rate in the time period from 1st April 1995 to 1st January 2018 and the catalogue consists of events with $M \geq 1.5$. The seismicity rate is given by

$$\lambda_m(\vec{x}, t) = h\theta_0 \Gamma H_p \cdot (\theta_1 \Delta \dot{P} + \theta_2 \Delta \ddot{P}) \cdot e^{\Gamma H_p (\theta_1 \Delta P + \theta_2 \dot{P})}, \tag{3.15}$$

where $H_p = H/H_r$, and $H_r = \Delta P/\epsilon_{zz}$. From Bourne and Oates (2017b), H is defined as $1/H = 1/H_r + 1/H_s$ with $H_s = 3K_s(\frac{1-\nu}{1+\nu})$ being the skeleton reservoir modulus based on observed strain and K_s the skeleton bulk modulus. Γ is the topographic gradient and h the reservoir thickness. ΔP and $\Delta \dot{P}$ are non-negative. The resulting seismicity rates overestimate the seismicity rates from 1995 to 1998 and for the short smoothing windows do not reflect the long term trends in the seismicity rates, but are dominated by the annual depletion rate variations. Correspondingly, the model parameter θ_2, that reflects the influence of the viscosity, approaches zero in the model optimisation.

The advantages and problems of the viscoelastic NAM model can be summarised as follows:

+ The viscoelastic model is a generalisation of the elastic model.

– The potential of the viscoelastic model is to capture the delayed and buffered reaction of the reservoir to stresses. Therefore, the calibration of the model should cover the whole production period. When the optimisation period covers only part of the production-induced stress loading, the long term effects cannot be modelled.

– The assumption of the extreme threshold failure model was introduced for the linear elastic model. The validity for the viscoelastic model is not straightforward.

– The model does not account for aftershocks or magnitudes.

The Rate-and-State model

The rate-and-state model introduced by Dieterich (1994) connects results from rock mechanics and seismology. The basic concept is employing laboratory-derived friction laws and the Coulomb failure stress (Eq. 3.2 in section. 3.1) to describe, when slip occurs on a plane or fault according to the local stress state (see also section 3.1). Experimental studies revealed that rock frictional behaviour can only be described in first-order with the linear Coulomb failure stress, but that the friction coefficient f has a non-linear behaviour depending on slip, slip rate, contact time and stressing history. A widely used form of the constitutive friction law was introduced by Ruina (1983) and is given by:

$$f = f_0 + A \cdot \ln\left(\frac{v}{v_s}\right) + B \cdot \ln\left(\frac{\theta}{\theta_s}\right), \qquad (3.16)$$

where f_0, A and B are experimentally determined coefficients, v is the slip velocity, v_s the steady-state slip velocity, θ a state variable and θ_s the according steady-state value. The rate-and-state dependent friction has been used to reproduce a range of sliding regimes, including creep processes and earthquake nucleation.

The seismicity is modelled as a sequence of earthquake nucleation events depending on the initial conditions and the stressing history. In order to calculate the seismicity rate caused by applied stresses S, Dieterich (1994) considered the collective behaviour of an infinite population of faults. The seismicity rate R evolves according to

$$R(\vec{x}, t) = \frac{r(\vec{x})}{\gamma(t)\dot{\tau}_r} \ , \ d\gamma = \frac{1}{A\sigma}[dt - \gamma dS], \qquad (3.17)$$

where r is the steady-state earthquake rate for the reference stressing rate $\dot{\tau}_r$, γ is a state variable, $\sigma = \sigma_n + p$ is the effective normal stress that consists of normal stress and pore pressure and S is the modified Coulomb failure stress based on the shear and effective normal stress (Dieterich et al., 2000).

The solution of equation 3.17 was given for step changes (Dieterich et al., 2000; Hainzl et al., 2010c) and later on applied to several cases (Toda et al., 2002; Green et al., 2015; Norbeck and Rubinstein, 2018). Therefore, Heimisson and Segall (2018) performed a revision of the model and confirmed the validity for arbitrary

The Rate-and-State model (Dieterich, 1994) considers the seismicity caused by a stressing history as the collective behaviour of an infinite population of faults. Heimisson and Segall (2018) confirmed the model and provided solutions for changing normal stresses.

stress histories if the normal stress is time inde-
pendent. If the normal stress varies strongly in
time, they provide an alternative solution, which can better describe the resulting seismicity:

$$\frac{R}{r} = \frac{K(t)}{1 + \frac{1}{t_a}\int_0^t K(t')dt'} = \frac{K(t)}{1 + \frac{\dot{\tau}_r}{A\sigma_0}\int_0^t K(t')dt'} \ . \tag{3.18}$$

Here, the function $K(t)$ is a kernel that describes the slip dependency on stress and stressing rate
according to

$$K(t) = \exp\left(\frac{\tau(t)}{A\sigma(t)} - \frac{\tau_0}{A\sigma_0}\right)\left(\frac{\sigma(t)}{\sigma_0}\right)^{\alpha/A} . \tag{3.19}$$

R is the seismicity rate, r is the background seismicity rate, t_a is the characteristic aftershock decay
time ($t_a = A\sigma_0/\dot{\tau}_r$, introduced by Dieterich, 1994), $\tau(t)$ is the shear stress, $\sigma(t)$ is the normal stress
that varies with time, τ_0 and σ_0 are the respective values for t=0 and α is a constitutive parameter
relating changes in normal stress to changes in state. The rate-and-state model has two more pa-
rameters compared to the Coulomb failure model, but can additionally explain:

- the occurrence and rate decline of aftershocks,

- a delay of seismicity compared to the loading rate,

- the low probability of earthquakes occurring in a stress shadow.

NAM-RS model

Bourne and Oates (2018) tested the rate-and-state model based on the formulation from Heimisson
and Segall (2018) and implemented it in the framework of the thin-sheet model to analyse the influ-
ence of seasonal depletion rates on the observed seismicity rates and to compare the results to the
NAM model (Bourne and Oates, 2017b). Here, the seismicity rates are given by

$$\lambda(\vec{x}, t) = \theta_0 h(\vec{x}) \frac{e^{S(\vec{x},t)/\theta_1}}{1 + \frac{\theta_2}{\theta_1}\int_0^t e^{S(\vec{x},t')/\theta_1}dt'} \tag{3.20}$$

based on weekly maximum Coulomb stress $S(\vec{x}, t)$ achieved from the thin-sheet model. The θ-
parameter refer to $\theta_0 = r$, $\theta_1 = A\sigma$ and $\theta_2 = \dot{\tau}_r$, while $h(\vec{x})$ is the reservoir thickness.
The model parameters $\theta_0, \theta_1, \theta_2$ are optimised
as in the other models using the seismic cata-
logue from 1st April 1995 to 1st January 2018 for
events with $M \geq 1.5$. In comparison to the NAM
model, the rate-and-state model neither does
explain large seasonal variations nor the abrupt
decrease in seismicity in 2014, when the produc-
tion was drastically diminished.

> The NAM-RS model (Bourne and Oates, 2018) combines the thin-sheet model with the Rate-and-State model modified by Heimisson and Segall (2018).

The advantages and problems of the NAM-RS model can be summarised as follows:

± The comparison of the rate-and-state model with the elastic model points to a better description of seismicity by the elastic model, but only in combination with the extreme threshold failure; whereas the rate-and-state model explains the overall seismic evolution at once.

\+ The formula from Heimisson and Segall (2018) differs only from Dieterich (1994) if significant normal stress changes occur. This most probably takes place at Groningen, but changes in normal stress are not included in the thin-sheet model.

\+ The small seasonal amplitudes are consistent with findings from Candela et al. (2019) for stress changes within periods $t \ll t_a = \theta_1/\theta_2$.

– The results from Richter et al. (2019) point to better results of the rate-and-state model based on poroelastic triggering than on compaction strain which is the main driving force in the thin-sheet model.

Candela-RS model

Candela et al. (2019) proposed a semi-analytic approach to calculate the Coulomb stress on faults that divide the reservoir into compartments and result in differential compaction. They took the complex 3-D structures into account, especially the vertical offset at the faults (van Wees et al., 2019). The input data for the geomechanical model is the 3-D pressure field for Groningen with a yearly sampling rate. The poroelastic stress change along the faults is computed on a metre scale by integration of all grid blocks considered as compaction nuclei of strain (Okada, 1992). The stress changes honour the effect of differential compaction caused by the fault offsets. At the faults, the stress values are sampled and the maximum Coulomb stressing rate is taken for each vertical section (pillar). The modified Coulomb stress function considers changes in the effective normal stress as well, nonetheless the seismicity rate is calculated by Eq. 3.17. The authors explain that the observed variations in normal stress for the time period 1993 to 2016 are in the range of 3 MPa, whereas the average value is 12.5 MPa; thus, the expected deviation is not significant. The normalised seismicity rate R depends on the stressing rate $\dot{S}$ as follows:

$$\frac{dR_N}{dt} = \frac{R_N}{t_a} \left[\frac{\dot{S}}{\dot{\tau}_r} - R_N \right] , \ R_N = \frac{R}{r} . \tag{3.21}$$

Here the state variable γ is eliminated by the introduction of the normalised seismicity rate relative to the background rate r as in Segall and Lu (2015). The computed rates are compared to observed seismicity after Gaussian smoothing of the 2-D field of R with the parameter σ_s. The assessment of the best model is performed by a maximum log likelihood estimate to determine the posterior probability distribution of the four

> Candela et al. (2019) calculate Coulomb stresses on the 3-D fault system for two sub-areas based on differential compaction combined with the Rate-and-State model (formulated as by Segall and Lu, 2015) starting in 1993 to achieve seismicity rates.

model parameters $[\sigma_s, \dot{\tau}_r, r, A]$ using two methods (a brute-force grid search and a Markov chain Monte Carlo method). Two time periods are compared as training periods: (i) 1st January 1993 to 31 December 2015 and (ii) 1st January 1993 to 31 December 2010, but both time periods start long after production at Groningen in 1968. The long characteristic relaxation time $t_a = A\sigma/\dot{\tau}_r$ (>80 years) found for the training periods indicates that the seismicity is in unsteady-state in 1993, although the solution of Eq. 3.17 assumes initial steady-state conditions. The authors argue that the redefinition of $R' = \beta R$ satisfies the original equation if the background stressing rate $\dot{\tau}_r$ and the background seismicity rate r are as well redefined ($\dot{\tau}_r' = \dot{\tau}_r/\beta$, $r' = r/\beta$). Consequently, the background rates in 1968 are $1/\beta$ times lower than in 1993 and the values in 1993 are "apparent background values" while the whole fault system is in transition between two steady-state phases. In this model, the same earthquake catalogue is used as in the other studies (Bourne and Oates, 2018; Dempsey and Suckale, 2017), i.e., the earthquake catalogue provided by KNMI. The results from two subsets of the catalogue are compared: one subset with a minimum magnitude of $M = 1.0$ and the other with $M = 1.3$. Both subsets lead to similar results.

The influence of seasonal stress variations is discussed as a minor factor, because the associated stress changes have periods $t < 1a$ meaning $t \ll t_a$. Therefore, the seismicity rates have only little variability and the overall seismicity rates are proportional to the stress changes and not to the stressing rates in the observation time, which agrees with the findings of Bourne and Oates (2018). The influence of aftershocks is analysed by using the declustering method from Zaliapin et al. (2008) and Zaliapin and Ben-Zion (2016) detecting a small number of aftershocks for the southwestern area (3 of 67 events) and around 18% for the central area (39 of 219 events). Nevertheless, the catalogue was not declustered following the results from Ziv and Rubin (2003) and Heimisson (2019) that the effect of stress interaction caused by aftershock sequences does not significantly perturb the rate-and-state results.

Two sub-areas of the Groningen gas field are analysed by Candela et al. (2019), the one exhibiting the highest seismic activity within the central area and the other in the southwest featuring much lower activity. Candela et al. (2019) successfully modelled the first-order spatiotemporal distribution of the observed seismicity in both areas for differing model parameters.

The advantages and problems of the Candela-RS model can be summarised as follows:

+ The model describes the stresses in a very detailed semi-analytic approach on 3-D faults according to a 3-D pressure history considering the shear and the normal stresses. This is combined with the seismicity rate theory from Dieterich (1994) to obtain the spatiotemporal distribution of seismicity rates.

+ The study is based on 2-D stress results, but can as well be applied in 3-D if the earthquake locations are better constrained.

+ The model can explain well the spatiotemporal distribution of observed earthquake rates with

minor deviations.

- The modelled seismicity rates are represented by the seismicity rate of the mean model parameters and the 95% confidence interval for a stochastic Poisson process. For some years, the number of observed earthquakes lies outside the 95% confidence interval, which may be caused by aftershocks, short period stress perturbations or uncertainties in the stress or seismicity model. Also, the selection of the area may have an influence if there is stress interaction with regions outside the selected area or errors in event locations distort the results.

- The introduction of the "apparent" background stressing and seismicity rate in 1993 is a critical assumption because the most intensive production period lies before 1993. Therefore, a large pre-stressing history that is most probably spatially heterogeneous is not included.

- The model does not account for the magnitudes of the earthquakes.

Richter-RS model

The study from Richter et al. (2019) modelled the seismicity rates for the complete Groningen field based on the rate-and-state friction behaviour described by Dieterich (1994). The changes of the spatiotemporal stress field were calculated by testing two simple models: (A) proportionality to pore pressure changes according to poroelasticity and (B) proportionality to compaction strain changes as the ratio of compaction change over reservoir thickness. The rate-and-state formula was solved by describing the stress history as a succession of steps and iteratively determining the state variable as provided by Hainzl et al. (2010b) assuming constant normal stresses. The stress field was based on the yearly pore pressure and compaction data from the 2-D-NAM model. The location of faults was taken into account by a fault density map that was implemented as a factor varying the background seismicity rate r. The fault density was calculated based on the 325 largest faults as provided by Dempsey and Suckale (2017) and smoothed by means of a Gaussian function with a standard deviation of 1 km. The model parameters r, $A\sigma$ and t_a were optimised by a maximum log likelihood method compared to the same catalogue as in the previous studies considering the events with $M \geq 1.5$ and assuming that no such events occurred before 1991.

Four different model assumptions were compared and evaluated by the Akaike Information Criterion: (i) stresses are proportional to pore pressure with Coulomb failure model, (ii) stresses are proportional to pore pressure with rate-and-state frictional behaviour, (iii) stresses are proportional to compaction strain with Coulomb failure model, and (iv) stresses are proportional to compaction strain with rate-and-state frictional behaviour. Additionally, the influence of

> Richter et al. (2019) model the spatiotemporal seismicity rates in the entire Groningen field with the Rate-and-State model by Dieterich (1994) based on the 2-D pore pressure and compaction data.

smoothing the earthquake locations (0 km to fieldwide) and fault strike direction were tested. The best-fitting model was found to be the rate-and-state model based on pore pressure (model ii) with a location error of 3 km and taking the fault strike direction into account. This model well described the long term trends in seismicity as well as the spatiotemporal seismicity pattern in the whole field. Deviations occur e.g. in the southwestern region, where the seismicity was overestimated similar to Candela et al. (2019), who find different model parameters there compared to the central region. Other results were that taking the fault density into account improved all models significantly and explains prominent patterns in the seismicity. Comparing the models based on pore pressure and on compaction strain the temporal changes in the spatial seismicity pattern were not as clearly displayed by the compaction strain model no matter if the linear or rate-and-state frictional behaviour was chosen. On the other hand the rate-and-state frictional behaviour improved clearly the temporal changes in seismicity pattern of the pore pressure model to the observed patterns.

The advantages and problems of the Richter-RS model can be summarised as follows:

+ The model is a simple approach to simulate the main features of the spatiotemporal seismicity and evaluate the influencing parameters as e.g. the type of frictional behaviour (linear or rate-and-state), the driving force (pore pressure or compaction strain) or considering the faults and their orientation.

– The parameterisation is too coarse to model detailed geomechanical interactions.

– The model gives no information on the magnitude of the induced earthquakes.

Brief summary of RS models: The three applications of rate-and-state models to Groningen (NAM-RS, Candela-RS, Richter-RS) partly lead to different conclusions: (i) The combination with the thin-sheet model in the NAM-RS model leads to worse results compared to the original NAM model. (ii) The Candela-RS model show that the spatiotemporal pattern after 1993 can be well explained based on detailed Coulomb stress modelling on the 3-D faults in two sub-areas. (iii) The Richter-RS model is able to fit the spatiotemporal seismicity for the complete field and production period and yields best results based on pore pressure compared to the linear models or the RS-model based on compaction strain. This is in contradiction to the NAM model which is mainly driven by compaction strain. The different conclusions of the three model applications are likely related to the different stress calculations and model assumptions. However, a direct comparison to the NAM model is not possible.

In summary the advantages and problems of the implemented rate-and-state approaches for Groningen can be stated as follows:

+ The nonlinear seismicity response in the reservoir can be well reproduced by rate-and-state models both on large- and small-scale.

– The influence of significant changes in the normal stress is not yet fully explored.

– The assumption of the unknown initial steady-state condition is critical.

4   Gaps

The information on the 3D reservoir and fault structures, pressure depletion, reservoir compaction and seismicity in Groningen is comprehensive and detailed. The seismicity models for Groningen have considered most of these unique and outstanding information and are rather advanced. However, there are knowledge gaps in terms of process understanding as well as the role of structural and dynamic parameters controlling earthquake triggering and rupture in the Groningen field. In the following, we discuss these knowledge gaps and their potential impact without claiming completeness. We will first examine the question of background stresses together with the question of whether earthquake rupture outside the actual reservoir layer can be expected. Subsequently, we discuss knowledge gaps in the magnitude frequency distribution and whether or how detailed information on earthquake locations, mechanisms and rupture should be incorporated into the seismicity models.

4.1   Impact of background deviatoric stress and the triggering of neighboring faults

Earthquakes are triggered by CFS and ruptures are driven by the shear stress acting on pre-existing faults. Therefore, the knowledge of all three contributing factors - Coulomb stress (including pore pressure and friction), acting shear stress and existence of faults - is needed to approach the triggering problem and to study induced seismicity.

Since stresses cannot be measured directly with remote sensors and are difficult to determine in situ, the magnitude and orientation of stress as a function of space and time is often least known. This is a major knowledge gap. For example, the extend of highly stressed fault

> The poor knowledge of magnitude and orientation of background stress is a major knowledge gap

patches at reservoir level probably limits the maximum magnitude of the induced earthquakes within the Groningen field, and is itself controlled by poroelastic depletion, pre-stress and the inelastic strain in the field.

A key question till today is how large the partitioning of deformation between poroelastic and time-dependent inelastic components is, and how heterogeneous the remaining stored elastic stress is today. For instance, Spiers et al. (2017) performed tri-axial lab deformation tests on reservoir rocks to study the amount of elastic versus inelastic strain and to establish upscaling rules to field scale. They found that up to 40-50% of the deformation occurring in samples with porosity of 19-20% (Rotliegend sandstone) was inelastic and irreversible. Time-dependent inelastic deformation (creep) was observed as well and was clearly affected by pore fluid composition (e.g. stress corrosion cracking processes). Spiers et al. (2017) speculated that the partitioning of deformation between poroe-

lastic and time-dependent inelastic components may influence the nucleation and size of ruptures. Additionally, marked accelerated and stress-dependent reservoir compaction may control the nucleation of induced seismicity in Groningen. A somehow similar concern was raised by SodM in 2015 (Mundendamm-Bos, per. comm.), and later taken up by Jansen and Herber (2017), suggesting that fluctuation in production (opening or closing of wells, seasonal fluctuations in productions, etc.) may dynamically trigger seismic events, as for instance (poro-)elastic stress waves may travel through the reservoirs. Kortekaas and Jaarsma (2017) pointed out that the permeability structure of the Groningen reservoir is complex and highly anisotropic, for instance due to fault transmissibility. Therefore, most clusters of production wells have their own specific area from which they produce (so-called drainage areas). This may indicate that pressure drop, compaction and differential compaction may be heterogeneous and promote failure.

Another process potentially influencing the pore pressures and thus the Coulomb stress inside and outside of the reservoir is related to the underlying aquifer system reacting to production and injections in Groningen. The Groningen field contains only a single injection cluster at Borgseer in the east of the field. To our knowledge, injections there are not performed with larger rates, and pore pressure diffusion from these injections are considered in the dynamic reservoir model (Mundendamm-Bos, pers. comm.). But it is unclear whether intermediate and long-term effects from injections, even at larger distance, may impact the seismicity at Groningen. Worldwide, examples of causual relationsships and far-distance and strong-delayed effects on induced and triggered seismicity are documented (van der Elst et al., 2016; Galis et al., 2017; Grigoli et al., 2018).

The state of stress on faults and their stress history is assumed to be a key factor controlling the trigger potential of earthquakes. However, both values are difficult to measure. While much work has been done to estimate and forecast

> The stresses outside of the field and within basement rocks are rarely studied.

pressure and stress within the reservoir layer, the stresses outside of the field and within basement rocks are rarely studied. How large is the stored tectonic background stress today and can this lead to larger triggered earthquakes by runaway ruptures that are difficult to control? The existing fault population outside the reservoir may control the b- and M_{max}-values of such events, which plays a key role for size forecasts. Interestingly, the induced seismicity in the Groningen area occurred only after a pore pressure reduction of $\sim 10\,MPa$ was reached (e.g. Candela et al., 2018). This is often taken as evidence that before production most regional faults were far from critical tectonic Coulomb stresses. On the other hand, gas fields in Germany, which are placed in very similar Rotliegend formations and show a delayed onset of seismicity as well, experience significant earthquakes that occur on faults outside the reservoirs (e.g. the Rotenburg/Völkersen fields). This behaviour has not yet been explained, and we believe that this question should receive more attention in further developments of seismicity models as a delayed onset of deeper boundary faults cannot be excluded in the future

for Groningen. However, so far the vast majority of the earthquakes at Groningen are located within the reservoir layer.

As stated above, earthquakes beyond the reservoir depth may occur as runaway ruptures which are initiated within the reservoir itself. Buijze et al. (2019) developed a generic 2D offset-fault stressing model and combined this with a simulation of dynamic ruptures. The aim was to study controls on the triggering and the size of induced earthquakes in Groningen. The stress pattern output from the numerical simulations were confirmed by a closed-form analytical solution for a displaced fault crossing a depleting reservoir interval (Jansen et al., 2019). In a depleting reservoir, fault slip is promoted if the offset is becoming larger. The rupture typically propagates from the upper reservoir level downward into the reservoir for depletion scenarios. However, the likelihood that the rupture propagates into the under- or overburden is larger if the offset is small (i.e. if the stress perturbation is smoother). According to Buijze et al. (2019), a transition to a runaway rupture, i.e. a rupture propagating far outside the reservoir interval, is observed if the excess strength ratio, which covers the effect of in-situ stress and the friction drop on rupture potential, is high enough. The excess strength ratio had been assessed by dynamic rupture simulations, an approach that may be extended for the Groningen field region.

4.2   Stress dependence of the M-f distribution

Seismicity models for Groningen have been designed to reproduce the historical seismicity for the given production and the derived Coulomb stress loading. Forecasting larger magnitude events, potentially outside the range of previously measured magnitudes, is a very critical issue and this is where the differences between the various suggested seismogenic models increase. A magnitude frequency distribution has to be extrapolated to realise such forecasts, and thus a decision on the parameterisation of the distribution has to be taken.

While the most simple and conservative assumption is to use a static Gutenberg-Richter trend with a constant b-value combined with a load-dependent productivity, newer models suggest to implement time- or stress-dependent shapes of the distribution. For instance, Bourne and Oates (2017a) proposed the use of a stress-dependent b-value, $b = b(\sigma)$ and in a more recent internal report (Bourne and Oates, 2019), the use of a stress-dependent tapered power law for larger magnitudes, $\zeta = \zeta(\sigma)$, or both. The impact of stress-dependent magnitude frequency distributions on seismic hazard can be large.

> The impact of stress-dependent $M - f$ distributions on seismic hazard can be large, but such models are not established.

To date no established theory exists to favour one of the distributions. Whether more complex parameterisations of the magnitude distribution can be justified depends on the seismicity database in combination with the history of Coulomb stress forcing. A problem with the seismicity in the Gronin-

gen field is the limited number of events, which renders any purely statistical approach obsolete. However, the analyses by Muntendam-Bos et al. (2017), Bourne et al. (2018) and Bourne and Oates (2019) indicate that in the Groningen case, the b-value may become smaller with increasing stressing rate or increasing depletion. The internal report by Bourne and Oates (2019) investigates the consequences of stress-dependent b- and ζ-values on the earthquake probability. The modification of the standard Gutenberg-Richter approach is partly justified by theoretical crack and trigger models.

In this context, the theoretical study by Almakari et al. (2019) is of interest, studying the dependence of earthquake rate and frequency magnitude distribution for injection-induced seismicity. A 2D fault model was employed in a fluid diffusion scenario involving a rate-and-state seismicity model. The authors showed that the frequency of larger magnitude events is increased with increasing injection rate, and that the proportion of large to small earthquakes depends on the loading. An increased loading rate can lead likewise to a higher proportion of large magnitude events. This would support the research approach started in the most recent NAM models. As the modification of the Gutenberg-Richter approach is of high relevance for the forecast of large magnitude induced events, we feel such a step should be further investigated and more comparisons to other induced seismicity cases should be drawn.

4.3   Exploitation of high precision source and field parameters

The enhancement of correlation of earthquake hypocentres with faults and variability of co-seismic source mechanism and isotropic compaction components has not yet been considered in seismicity models of Groningen. However, recent studies have the potential to further improve the forecasts on seismicity and to identify critical faults in a deterministic way. Future production plans may use such information to mitigate risk. Especially, highly resolved fault models are of interest to understand the stress build-up at the two sides of a sealing fault (Jansen and Herber, 2017). Additionally, fault properties such as transmissibility and slip resistance may vary along individual faults, depending on orientation, stress state, strain, throw, free water levels and local diagenesis. Alternatively, a fault with a significant throw such that compacting reservoir rock is in direct contact with stationary non-reservoir rock is viewed as a critical situation.

Highly resolved fault models can be derived from 3D seismics. Kortekaas and Jaarsma (2017) applied such a technique involving seismic attributes to the Groningen field. Several of the identified faults were not included in previous seismicity models, but seem to correlate with recent earthquake hypocentres. For instance, Bourne et al. (2015) compared earthquake hypocentres with mapped faults and found a rather uncertain relationship because of the

> The combination of high precision source and rupture parameters with high resolved fault models is able to close knowledge gaps and to provide better physical constraints to improve seismicity models.

combined uncertainty of hypocentre and fault locations. The improved fault model of Kortekaas and Jaarsma (2017) recovered interesting features. For instance, the M 3.6 Huizinge earthquake ruptured a fault with relatively small throw and the authors questioned whether a minor throw is promoting the rupture of larger magnitude events. In addition, intersecting faults, as being present close to the 2018 Zeerijp event (Wentinck, 2018b), may accumulate stress and fault complexity may be related to larger induced events (see also Candela et al., 2018). Further, recovered fault attributes potentially differ from the NAM fault model (Wentinck, 2018b).

An improved location of the small earthquakes within the Groningen field is necessary to obtain a better seismicity statistics and to characterise active weak zones prior to the occurrence of large earthquakes. KNMI is currently implementing NAM's 3D velocity model into the earthquake location process (Dando et al., 2018). Further, several novel methods have recently been developed to improve the absolute and relative locations of induced earthquakes. For instance, Perol et al. (2018) presented a highly scalable convolutional neural network for earthquake detection and location. Applying the technique to induced seismicity in Oklahoma, USA, a 17-fold number of earthquakes could be detected as previously catalogued by the Oklahoma Geological Survey. Similar methods could be employed for Groningen.

Multiplet analysis can help to identify clusters featuring similar waveforms and to recognise highly active, critical fault segments. Jagt et al. (2017) employed subsequent relative locations using high precision P- and S-wave arrival times and a master event technique and showed that in the time period from 2010 to mid-2014 only 10% of all earthquakes in Groningen are spatially clustered, even if they occur long periods apart. Muntendam-Bos, 2020 found that seismicity in Groningen is very diffuse and clustering comprises only 5% of fore- and aftershock activity and 10% of repeater events. This indicates that (1) seismicity typically occurs on divers fault segments with few kilometres distance and likely varying source mechanisms, (2) aftershocks are not very prominent and (3) a fault segment is only rupturing once. Whether the clusters occur at loci of geometrical complexities of fault systems (e.g. intersections, kinks, etc.) has not yet been resolved.

Another feature that may contribute significantly to improve source models is the events' source mechanisms. Willacy et al. (2018; 2019) analysed 100 events from 2015 to September 2017 to obtain moment tensor solutions based on full-waveform analysis. The authors found predominantly normal faulting on the major NNW-SSE trending faults. A few events with a higher proportion of strike-slip motion were located at fault junctions. In addition, they identified a cluster of four events that occurred within a time period of two months. In a common project, NORSAR and KNMI inverted for full moment tensors and - for their example event (on 11th March 2017) - detected a complex faulting behaviour involving normal faulting and a simultaneous volume reduction in the source region (equivalent to bulk density increase), which strengthens the opinion that gas extraction may lead to isotropic components in the source mechanism, although the credibility of these components still needs to be confirmed (Dando et al., 2018).

In addition to location and source mechanism kinematic rupture models have a potential to improve the understanding of induced seismicity, but are not fully considered in current seismicity models. Source models can be improved if the observation density is high, but needs also highly resolved velocity models. For instance Wentinck (2017) analysed the Huizinge earthquake of 16th August 2012 with a magnitude of M_w 3.4. Geometry and wave velocities of the overburden were derived from NAM's seismic velocity model (updated in 2015) by lumping several layers. The very low shear wave velocities within the uppermost, shallow clay and peat layers were not captured by the model, though. Simulations were therefore limited to a frequency content below 5 Hz. Wentinck (2017) noted that the difference between ground motions originating from an extended or a point source were subtle and could not easily be resolved with the present uncertainties in accelerometer orientation, and a double peak perceived in the data was not reproduced. Thus, Wentinck (2017) concluded that this double peak in radial displacement component was due to two seismic events occurring quickly after each other. A continuation work implementing additional accelerometers indicated that the observed second peak likely is produced by S-wave reflections of a single event in the shallow subsurface instead of the occurrence of a second event (Wentinck, 2018c).

Wentinck (2018b) applied the same technique of kinematic rupture modelling to the more recent Zeerijp earthquake on the 8th January 2018 with a magnitude of M_L 3.4, and could implement even more geophone and accelerometer sensors (see also Dando et al., 2018). Results are hinting to a large stress drop of 3-5 MPa (Wentinck, 2018a), a result that may impact seismicity models at Groningen, as boundaries for the pre-seismic state of stress on a fault can be constraint.

The examples show the potential of small earthquake rupture studies, but also indicate that despite the extension of the monitoring network in 2015, the reconstruction of the kinematic and dynamic rupture of such small earthquakes remains a challenge.

4.4 Comparative model testing

The Groningen gas field is exceptional not only with respect to the available measured data, but also with regard to the number of suggested seismicity models. To our knowledge, there exists no other case of gas or oil production with a comparable characterisation of the mechanical properties of the reservoir, sophisticated geomechanical modelling and comprehensive seismic hazard studies. Particularly in chapter 3 and table 1, we introduced and discussed several seismogenic source models, which have been specifically proposed for Groningen, as well as other model approaches, which could potentially be applicable as well. It has been already demonstrated that several of these models are able to fit the past activity within the expected natural variability. However, the forecast power of these models is not clear.

Each of the models were compared and tested with existing data by the models' authors, partly employing sophisticated test procedures.

However, some models were tested for the time-dependency of the overall seismicity, while others were tested for spatiotemporal patterns. Also, due to the complexity of most models and the fact that some models are specifically developed and adapted for the Groningen field, it remains unclear whether or not hidden parameters were set based on the knowledge of the full data set. Such a hidden model optimisation often occurs without explicit perception of the authors. Furthermore, models are at times evolving. An example is the development of NAM models (Bourne et al., 2014; Bourne et al., 2015; Bourne and Oates, 2017a; Bourne and Oates, 2017b; Bourne et al., 2018). Therefore, it is almost impossible for us to judge and compare the predictability of the different models in detail. A possible solution to this knowledge gap is presented by comparative prospective testing. For that purpose, the model codes would have to be implemented in an independent testing centre and tested against unseen data.

> The predictive power of alternative models is impossible to judge without comparative and prospective testing.

5   Discussion

In chapter 3, we have reviewed 21 models developed specifically for, or potentially applicable to, the Groningen case. The eight physical models consist of two analytical and six deterministic models investigating specific scenarios. Five models were purely statistical or involving machine learning techniques, and eight were classified as hybrid models combining physical principles with statistical concepts (see table 1). It is a challenge to compare and evaluate such a large number of models against each other, since they are developed for various aims and specific targets; some are specialised to forecast the cumulative number of events, others to reproduce the spatiotemporal distribution of magnitudes, some are adapted to predict aftershocks, others to recover the long term trend of seismicity. We believe there is a need for SoDM, the operators and other users to concentrate future developments on the most promising class of models and to be able to continuously evaluate model performances with regard to new developments, software releases and earthquake occurences. We therefore focus our discussion on two main aspects: (1) the strengths and weaknesses of model classes in view of the Groningen field and (2) the discussion of the potentials of a testing framework.

5.1   The strengths and weaknesses of model classes

Physical models and scenarios

The physical model class represents deterministic approaches that are important in order to advance the physical understanding of processes. Models in this class are often build to investigate specific aspects of rupture or seismicity, but cannot be implemented directly in seismic hazard studies. Examples in this report include the study whether and when earthquake ruptures are likely to grow beyond the extend of the reservoir or how the relation between aseismic and seismic deformation depends on structural and pre-stress conditions. The analytic expressions presented by Segall (1989), and Jansen et al. (2019), are of help to obtain a comprehensive understanding of processes as a function of model parameters and geometry, but are restricted to quasi-static solutions. Numerical simulations of simplified 2D reservoirs with one fault offsetting the reservoir (e.g. van Wees et al., 2014; van Wees et al., 2017a; van Wees et al., 2017b; DeDontney and Lele, 2018; van den Bogert, 2018; Buijze et al., 2017; Buijze et al., 2019; Zbinden et al., 2017) were performed to understand the quasi-static or dynamic rupture nucleation and rupture process.

Physical models will be developed further and applied to the Groningen field as well in the future in order to continuously improve the process understanding of induced seismicity. However, both analytical and numerical approaches have clear limitations considering the complex nature of reservoirs, rocks and earthquake ruptures. This is either because of computational costs or because of a lack of knowledge of structure and state of the subsurface. In the following, knowledge gaps are listed that may be approached by deterministic models in the future.

The role of the initial stress state on faults is crucial for the triggering and rupture of induced earth-quakes. If the initial stress is low, no earthquakes will be triggered. In contrast, faults that are already critically stressed can be triggered immediately. Some of the seismicity models assumed that the delayed onset in Groningen's induced seismicity was caused by the low state of stress on faults, whereas others explained the delay by the evolution of frictional instabilities. As long as the absolute stress cannot be measured precisely, e.g. from wells, the physical models may be better suited to investigate the impact of pre-stress on seismicity. Such studies could involve the effect of unknown structural and material inhomogeneities, which may play a key role in earthquake dynamics, as for example highlighted by the study of Kettermann et al. (2017). Specifically, the potential impact of salt intrusion on stress could be analysed.

Another controversy to be approached is the simplified description of the rupture nucleation and propagation. Quasi-static simulations, as performed by van Wees et al. (2014; 2017; 2017) are crit-icised because they do not possess a well-defined continuum limit (Rice, 1993). In contrast, fully dynamic and continuous rupture models can only reproduce the observed richness of slip complex-ity if the nucleation dimension is chosen sufficiently small (Cattania, 2019). The latter leads to high computational costs, which is why dynamic rupture simulations were only performed on a 1D fault by van den Bogert (2018) and Buijze et al. (2017; 2019). In addition, simplified friction laws are im-plemented; in particular, most available dynamic rupture models for Groningen rely on a linear slip-weakening law. More sophisticated constitutive laws, such as the rate-and-state-dependent friction law of Linker and Dieterich (1992) are often suggested and were implemented by DeDontney and Lele (2018). However, the importance of the choice of friction law has not yet been analysed.

Purely statistical and machine learning models

Statistical models analyse previous seismicity in response to loading forces to forecast future seis-micity in response to future loading scenarios. Data-driven modelling is based on empirical statistical (multi-parameter) functions and concepts, with the advantage of being independent of the accuracy of simplified physical relations. Statistical models are strong in making short term forecasts for uni-versal loading scenarios, as for instance the prediction of aftershock activity from ETAS-type models. However, a drawback of purely statistical models is that they are often unable to predict the system behaviour if the loading is outside the range covered during the learning period. Statistical forecasts are still reasonable if the production (loading) parameter do not change quickly over time, as statis-tical models can adapt to such slow changes. Any quick change of conditions, however, can only be reliably modelled.

An interesting statistical model not yet exploited for Groningen is the purely data driven K2/K3 ap-proach by Helmstetter and Werner (2012; 2014). In this case, only observed seismicity is used to forecast the future distribution of rates and magnitudes as function of space and time. The method relies on adaptive kernel smoothing of the recorded earthquake catalogue. This model bypasses any

dependence on physical assumptions and thus avoids potential biases based on misleading features or modelling approaches, but adapts only with delay to a quick changes of pattern. More importantly, K2/K3 forecasts are limited to short time scales, because they are solely based on past events and do not include the effect of events occurring in the forecasting period.

A well-established empirical model is the ETAS model, which is used in the latest NAM model (Bourne et al., 2018). The ETAS-type aftershock modelling can be combined with any physics-based estimate of the depletion-induced seismicity rate. However, it is difficult to constrain the parameters of the ETAS model based on small earthquake catalogues such as the Groningen catalogue. As a consequence, unrealistic parameter values might be estimated leading to unstable simulations in the long term.

The approach of DeDontney (2017) avoids any physical explanation and instead employs combinations of power functions to fit the data and predict the future seismicity rate. However, this model contains very large uncertainties because of the lack of any constraint. Alternatively, Hettema et al. (2017) and Vlek (2019) exploited the observed correlation between seismicity rate and cumulative volume. This approach is significantly better constrained, because the maximum number of future earthquakes is limited by the gas volume remaining in the reservoir. Nonetheless, the approach was only demonstrated for the cumulative seismicity observed in the whole reservoir. First studies for different sub-regions of the Groningen field indicate that the regression depends on the location and the model is not easily extendable to a space-time model - a challenge that may be further analysed in the future.

Machine learning (ML) could present a solution to bypass the problem of unknown conditions and incomplete process understanding. Unlike physical modelling approaches, ML can extract relevant patterns and information directly from the available multivariate data sets without requiring insight into the underlying physical mechanisms. Limbeck et al. (2018) and Lanz et al. (2019) started to explore ML as an alternative approach for forecasting production-induced seismicity in the Groningen field. They tested an ML approach, which uses a large number of static (e.g. fault properties) and dynamic (e.g. pore pressure) features as predictors. However, the comparison to baseline models shows no significant improvement and the ML model accuracy and forecasting power remains limited by the small number of earthquakes available for model training. Another drawback is that confidence bands for forecasts cannot easily be derived for the space-time ML model. Finally, similarly to purely statistical models, predictions may be problematic if the future stress scenario is not included in the training data set. In this context, the missing physical constraints become crucial.

Hybrid models

Hybrid models combine physics-based deterministic with statistical models. They are the preferential model class towards time-dependent hazard assessments for induced seismicity, as they are able to forecast seismicity for loading scenarios outside the range of previous production. However, it should

be noted that the models depend on the results of the static and dynamic reservoir models, which are based on reservoir-flow history matches and forecasts. Thus, the quality of the hybrid models depends on the quality of these input data. Among the hybrid models, there are different strategies to deal with a delayed and nonlinear onset of seismicity. In particular, it is assumed that (i) a critical stress state is reached at the beginning of the fitting period and the delayed onset is not considered explicitly (SI-model and RS-model of Candela et al., 2019); (ii) a critical stress state exists from the beginning and the delay results from a non-linear frictional response (NAM-RS and RS-model of Richter et al., 2019); (iii) a single subcritical initial stress state combined with a non-linear stress response function (Wentinck-model); or (iv) a distribution of subcritical initial stress values combined with a linear stress response function (NAM and CMsub models). While SI, NAM, and CMsub models assume that the stress response of faults is linear once the critical stress level has been reached, in the RS models the response depends on the stressing history in agreement with lab-experiments (Linker and Dieterich, 1992; Dieterich, 1994). In any case the delayed onset of seismicity is a prominent feature of the seismic response to production and thus should be considered in the model as subcritical initial stresses or non-linear stress response.

In contrast to all other models, in the Wentinck-model the earthquake rate only depends on the absolute stress level without any dependency on the stressing rate. As a consequence, after a shut-in this model predicts only a slow decay of seismicity, which is solely related to the stress release by the earthquakes themselves. Thus, it is unlikely that this model can reproduce the observed, rather rapid drop in seismicity in response to the reduction of production during the last years. Furthermore, the Wentinck-model predicts a decrease in seismicity rate in response to every earthquake, loosely speaking: it predicts *anti-aftershocks*, in contrast to observed aftershock triggering. Both observations disqualify the Wentinck-model. Similarly, the SI-model seems to be less applicable, because it predicts a linear relation between seismicity rate and pore pressure changes in contradiction to the observations.

A ranking of the remaining hybrid models is more difficult. All of them are based on reservoir models and reasonable physical considerations, but the validity of those assumptions is not always clear. For example, it is still debated whether or not the data are in favour of the rate-and-state dependent friction (Bourne and Oates, 2018; Richter et al., 2019; Candela et al., 2019) or if the driving forces for earthquake nucleation in the Groningen field are better represented by compaction strain or pore pressure changes (Bourne et al., 2018; Richter et al., 2019). So far, it is unknown whether the assumption of the NAM model holds that the stress level is in the tail of the distribution of stress values on faults justifying the application of the extreme value theory. In turn, the CMsub model is based on results of simplified model simulations of 1D faults disregarding the dip, offset and small faults, which might not be representative for prospective seismicity. On the other hand, the sophisticated 3D fault rupture modelling of Candela et al. (2019) entails the need of different frictional behaviour for the investigated regions. However, the difference might also be caused by the ignored production

history before 1993. The spatial deviations observed by Richter et al. (2019) could in turn point to ig-nored fault properties (dip, offset). It is likely that rate-and-state friction applies on rough faults with heterogeneous pre-stress values, which has not yet been considered in any of the models. Further-more, due to the simplified physical model setup, the predictive power of the frequency-magnitude distribution simulated by the CMsub model may be questioned. In contrast, all other hybrid models predict magnitudes independently based on the empirical GR law, which requires a fit to the ob-served data. This fit might be only constrained weakly, in particular with regard to the estimation of the maximum magnitude and the stress-dependence assumed in the NAM model.

5.2   Testing framework for induced seismicity at Groningen

Because of the large number of parameter sets, assumptions and input fields, the different purposes and the various output and forecast result structures, it is almost impossible to compare and judge the performance of the alternative models at the current stage. For a long time, a similar problem ex-isted for natural seismicity. Various models were suggested for the occurrence of earthquakes, some claiming the possibility of rather precise forecasts of location, time and magnitude of large earth-quakes. To tackle this problem, the Collaboratory for the Study of Earthquake Predictability (CSEP) was established (Jordan, 2006). Within the CSEP-framework, fully prospective tests of earthquake forecasts are carried out, using standardised statistical tests and authoritative data sets to assess the predictive skills of forecast models and to objectively compare them. CSEP conducts prospective experiments in four testing centres around the world; currently, more than 400 models are under evaluation (Michael and Werner, 2018). These models include a range of methods and scales, ex-tending from long term earthquake forecasts used for time-independent seismic hazard assessment to short term regional forecasts used for Operational Earthquake Forecasting ((OEF); Jordan et al., 2014).

Not much work has been done so far on model selection and comparison in the context of induced seismicity. In principle, it would be straightforward to implement such a CSEP-type testing frame-work in order to validate, compare and rank the models developed for Groningen. A first attempt for induced seismicity was performed by Kiraly-Proag et al. (2016) in the case of the geothermal stim-ulation projects in Basel (2006) and Soultz-sous-Forêts (2004). However, the test was only pseudo-prospective, because they were performed after the projects were finished, which does not exclude the danger of hidden model optimisations, where parameters were set based on the knowledge of the full data set. In contrast to short term injection operations, gas production is in general long-lasting, which offers the possibility of conducting fully prospective tests. However, in the case of the Groningen field, the production will be terminated in the near future as decided by the Dutch government. Whether or not such a test setup is feasible will depend on the following requirements:

- The expected number of future earthquakes is large enough to allow for a discrimination be-tween models within a meaningful time range. This depends on the production plan as well as

on the detection limits of the seismic network. The predictive power of the alternative models is impossible to judge without comparative and prospective testing.

- The developers of the different models are able and willing to participate and provide software codes that run automatically, given the required input.

- A test centre can be established, where the codes are implemented and tests are performed by independent researchers.

The feasibility of such tests should be evaluated carefully, because such a test framework could provide important indications about the forecast power of the different models. In particular, the test results could provide important information how different models should be weighted in a logic-tree approach for seismic hazard assessment in Groningen.

6 Outlook

The high quality of the current monitoring network, field observations, and reservoir models enable further improvements of seismicity models for seismic hazard assessment in Groningen and adjacent regions. In particular, the exploitation of the deep borehole data leads to more precise earthquake locations, a lower completeness magnitude and enhanced moment tensor solutions of recent and future seismicity, which offer the possibility to answer focused research questions, e.g. concerning the impact of fault throws or orientation on the activity rate and earthquake magnitude. Do earthquakes rupture increasingly deeper into the basement and overburden on faults with small throws as some physical models predict? Furthermore, the data allows for a detailed comparison between the central field and the southwestern region subject to less seismicity in order to explore the main influencing factors, for example regarding the salt layer. The high-quality data may also allow for quantitative tests of the predictive power of the existing seismicity models, the current NAM as well as alternative hybrid models discussed in this report. Additionally, ensemble models may be developed and tested. Based on the results of such a testing framework and infrastructure, reliable weighting factors may be obtained ultimately to balance the different branches (seismicity models) of the logic tree for probabilistic seismic hazard assessment.

In the subsequent WP3 report, details of these potential future studies will be presented and discussed in detail.

A   Literature

References

Abe, S. and J. L. Urai (2012). "Discrete element modeling of boudinage: insights on rock rheology, matrix flow, and evolution of geometry". In: *J. Geophys. Res.* B01407. DOI: 10.1029/2011JB008555.

Almakari, M., P. Dublanchet, H. Chauris, and F. Pellet (2019). "Effect of the injection scenario on the rate and magnitude content of injection-induced seismicity - Case of a heterogeneous fault". In: *J. Geophys. Res.* 124, pp. 1–10. DOI: 10.1029/2019JB017898.

Bachmann, C. E., S. Wiemer, J. Woessner, and S. Hainzl (2011). "Statistical analysis of the induced Basel 2006 earthquake sequence: Introducing a probability-based monitoring approach for Enhanced Geothermal Systems". In: *Geophys. J. Int.* Pp. 793–807.

Bergen, K. J., T. Chen, and Z. Li (2019). "Preface to the focus section on machine learning in seismology". In: *Seismol. Res. Lett.* (2A), pp. 477–480. DOI: 10.1785/0220190018.

Bourne, S. and S. Oates (2018). *The influence of stress rates on induced seismicity rates within the Groningen gas field*. Tech. rep. NAM.

Bourne, S.J. and S.J. Oates (2017a). "Development of statistical geomechanical models for forecasting seismicity induced by gas production from the Groningen field". In: *Netherlands Journal of Geosciences* 96.5, s175–s182.

Bourne, S.J. and S.J. Oates (2017b). "Extreme threshold failures within a heterogeneous elastic thin sheet and the spatial-temporal development of induced seismicity within the Groningen gas field". In: *Journal of Geophysical Research: Solid Earth* 122, pp. 10299–10320.

Bourne, S.J. and S.J. Oates (2019). "The evolution of induced earthquake magnitude distributions with increasing stress in the Groningen gas field". In: *Restricted Draft, NAM*.

Bourne, S.J., S.J. Oates, J.J. Bommer, B. Dost, J. Van Elk, and D. Doornhof (2015). "A Monte Carlo method for probabilistic hazard assessment of induced seismicity due to conventional natural gas production". In: *Bulletin of the Seismological Society of America* 105.3, pp. 1721–1738.

Bourne, S.J., S.J. Oates, and J. Van Elk (2018). "The exponential rise of induced seismicity with increasing stress levels in the Groningen gas field and its implications for controlling seismic risk". In: *Geophysical Journal International*.

Bourne, S.J., S.J. Oates, J. Van Elk, and D. Doornhof (2014). "A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir". In: *Journal of Geophysical Research: Solid Earth* 119.12, pp. 8991–9015.

Broccardo, M., A. Mignan, S. Wiemer, B. Stojadinovic, and D. Giardini (2017). "Hierarchical Bayesian modeling of fluid-induced seismicity". In: *Geophys. Res. Lett.* Pp. 11, 357–11, 367.

Buijze, L., P. A. J. van den Bogert, B. B. T. Wassing, and B. Orlic (2019). "Nucleation and arrest of dynamic rupture induced by reservoir depletion". In: *J. Geophys. Research*. DOI: 10.1029/2018JB016941.

Buijze, L., P.A.J. van den Bogert, B.B.T. Wassing, B. Orlic, and J. ten Veen (2017). "Fault reactivation mechanisms and dynamic rupture modelling of depletion-induced seismic events in a Rotliegend gas reservoir". In: *Netherlands Journal of Geosciences* 96.5, s131–s148.

Candela, T., S. Osinga, J.-P. Ampuero, B. Wassing, M. Pluymaekers, P. A. Fokker, J.-D. van Wees, H. A. de Waal, and A. G. Muntendam-Bos (2019). "Depletion-induced seismicity at the Groningen gas field: Coulomb rate-and-state models including differential compaction effect". In: *J. Geophys. Res.* DOI: 10.1029/2018JB016670.

Candela, T., B. Wassing, J. ter Heege, and L. Buijze (2018). "How earthquakes are induced: Conditions within Earth's crust determine whether human subsurface activities lead to earthquakes". In: *Science*. DOI: 10.1126/science.aat2776.

Cattania, C. (2019). "Complex earthquake sequences on simple faults". In: *Geophys. Res. Lett.* DOI: 10.1029/2019GL083628.

Cornell, C. A. (1968). "Engineering seismic risk analysis". In: *Bulletin of the seismological society of America* 58.5, pp. 1583–1606.

Dando, B., B. Goertz-Allmann, D. Kühn, N. Langet, V. Oye, and A. Wüstefeld (2018). *Review of the public KNMI induced earthquake catalogue from the Groningen gas field (report project phase 1, WP1: catalogue review*. Tech. rep. NORSAR.

DeDontney, N. (2017). *Groningen seismological models for activity rate forecasting*. Tech. rep. NAM.

DeDontney, N. and S. Lele (2018). *Impact of production fluctuations on Groningen seismicity – Geomechanical modelling using Rate and State friction*. Tech. rep. NAM.

Dempsey, D. and J. Suckale (2016). "Collective properties of injection-induced earthquakes sequences: 1. Model description and directivity bias". In: *Journal of Geophysical Research: Solid Earth* 121, pp. 3609–3637. DOI: 10.1002/2015JB012550.

Dempsey, D. and J. Suckale (2017). "Physics-based forecasting of induced seismicity at Groningen gas field, the Netherlands". In: *Geophysical Research Letters* 44.15, pp. 7773–7782.

Dempsey, D., J. Suckale, and Y. Huang (2016). "Collective properties of injection-induced earthquakes sequences: 2. Spatiotemporal evolution and magnitude frequemcy distributions". In: *Journal of Geophysical Research: Solid Earth* 121, pp. 3638–3665. DOI: 10.1002/2015JB012551.

DeVries, P. M. R., F. Viégas, M. Wattenberg, and B. J. Meade (2018). "Deep learning of aftershock patterns following large earthquakes." In: *Nature*. DOI: 10.1038/s41586-018-0438-y.

Dieterich, J. (1994). "A constitutive law for rate of earthquake production and its application to earthquake clustering". In: *J. Geophys. Res.* 99 (B2), pp. 2601–2618.

Dieterich, J., V. Cayol, and P. Okube (2000). "The use of earthquake rate changes as a stress meter at Kilauea volcano". In: *Nature* 408, pp. 457–460.

Eshelby, J.D. (1969). "The elastic field of a crack extending non-uniformly under general anti-plane loading". In: *Journal of the Mechanics and Physics of Solids* 17, pp. 177–199.

Felzer, K. R. and E. E. Brodsky (2006). "Decay of aftershock density with distance indicates triggering by dynamic stress". In: *Nature* 441, pp. 735–738.

Galis, M., J. P. Ampuero, P. M. Mai, and F. Cappa (2017). "Induced seismicity provides insight into why earthquake ruptures stop". In: *Science Advances* 3:eaap7528, pp. 1–10. DOI: `10.1126/sciadv.aap7528`.

Geertsma, J. (1973). "A basic theory of subsidence due to reservoir compaction: The homogeneous case". In: *Verhandelingen Kon. Ned. Geol. Mijnbouwk. Gen.* 28, pp. 43–62.

Green, R.G., T. Greenfield, and R.S. White (2015). "Triggered earthquakes suppressed by an evolving stress shadow from a propagating dyke". In: *Nature Geoscience* 8, pp. 629–633.

Grigoli, F., S. Cesca, A. P. Rinaldi, A. Manconi, J. A. Lopez-Comino, J. F. Clinton, R. Westaway, C. Cauzzi, T. Dahm, and S. Wiemer (2018). "The November 2017 Mw 5.5 Pohang earthquake: A possible case of induced seismicity in South Korea". In: *Science* 360.6392, pp. 1003–1006. DOI: `10.1126/science.aat2010`.

Gutenberg, B. and C. F. Richter (1956). "Earthquake magnitude, intensity, energy and acceleration". In: *Bull. Seis. Soc. Am.* Pp. 105–145.

Hainzl, S. (2016). "Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs". In: *J. Geophys. Res.* Pp. 6499–6509.

Hainzl, S., G. B. Brietzke, and G. Zöller (2010a). "Quantitative earthquake forecasts resulting from static stress triggering". In: *J. Geophys. Res.* B11311. DOI: `10.1029/2010JB007473`.

Hainzl, S. and Y. Ogata (2005). "Detecting fluid signals in seismicity data through statistical earthquake modeling". In: *J. Geophys. Res.* B05S07. DOI: `10.1029/2004JB003247`.

Hainzl, S., S. Steacy, and D. Marsan (2010b). *Seismicity models based on Coulomb stress calculations, Community Online Resource for Statistical Seismicity Analysis*. Available at http://www.corssa.org. DOI: `http://dx.doi.org/10.5078/corssa-32035809`.

Hainzl, S., G. Zöller, and R. Wang (2010c). "Impact of the reciever fault distribution on aftershock activity". In: *J. Geophys. Res.* B05315. DOI: `10.1029/2008JB006224`.

Heimisson, E.R. (2019). "Constitutive Law for Earthquake Production Based on Rate-and-State Friction: Theory and Application of Interacting Sources". In: *J. Geophys. Res.* Pp. 1802–1821. DOI: `10.1029/2018JB016823`.

Heimisson, E.R. and P. Segall (2018). "Constitutive law for earthquake production based on rate-and-state friction: Dieterich 1994 Revisited". In: *J. Geophys. Res.* Pp. 4141–4156. DOI: `10.1029/2018JB015656`.

Helmstetter, A. and D. Sornette (2002). "Subcritical and supercritical regimes in epidemic models of earthquake aftershocks". In: *J. Geophys. Res.* P. 2237. DOI: `10.1029/2001JB001580`.

Helmstetter, A. and M. J. Werner (2012). "Adaptive spatio-temporal smoothing of seismicity for long-term earthquake forecasts in California". In: *Bull. Seismol. Soc. Am.* Pp. 2518–2529.

Helmstetter, A. and M. J. Werner (2014). "Adaptive smoothing of seismicity in time, space, and magnitude for time-dependent earthquake forecasts for California". In: *Bull. Seismol. Soc. Am.* Pp. 809–822.

Hettema, M. H. H., B. Jaarsma, B. M. Schroot, and G. C. N. Yperen (2017). "An empirical relationship for the seismic activity rate of the Groningen gas field". In: *Netherlands Journal of Geosciences*, s149–s161.

Holtkamp, S. G. and M. R. Brudzinski (2011). "Earthquake swarms in circum-Pacific subduction zones". In: *Earth Planet. Sci. Lett.* Pp. 215–225.

Jackson, D. D. and Y. Y. Kagan (1999). "Testable Earthquake Forecasts for 1999". In: *Seismol. Res. Lett.* Pp. 393–403.

Jagt, L., E. Ruigrok, and H. Paulssen (2017). "Relocation of clustered earthquakes in the Groningen gas field". In: *Netherlands Journal of Geosciences* 96.5, s163–s173.

Jansen, J. D., P. Singhal, and F. Vossepoel (2019). "Insights from closed-form expressions for injection- and production-induced stresses in displaced faults". In: *J. Geophys. Res.* DOI: 10.1029/2019JB017932.

Jansen, J.-D. and R. Herber (2017). "Research into induced seismicity in the Groningen field–further studies". In: *Netherlands Journal of Geosciences* 96.5, s279–s284.

Jonsson, S., P. Segall, R. Pedersen, and G. Bjoernsson (2003). "Post-earthquake ground movements correlated to pore-pressure transients". In: *Nature*, pp. 179–183.

Jordan, M. and T. Mitchell (2015). "Machine learning: trends, perspectives, and prospects". In: *Science*, pp. 255–260.

Jordan, T. H. (2006). "Earthquake predictability, brick by brick". In: *Seismol. Res. Lett.* Pp. 3–6.

Jordan, T. H., W. Marzocchi, A. J. Michael, and M.C. Gerstenberger (2014). "Operational earthquake forecasting can enhance earthquake preparedness". In: *Seismol. Res. Lett.* (5), pp. 955–959.

Kagan, Y. Y. (2004). "Short-term properties of earthquake catalogs and models of earthquake source". In: *Bull. Seismol. Soc. Am.* Pp. 1207–1228.

Kettermann, M., S. Abe, A. F. Raith, J. de Jager, and J. L. Urai (2017). "The effect of salt in dilatant faults on rates and magnitudes of induced seismicity – first results building on the geological setting of the Groningen Rotliegend reservoirs". In: *Netherlands Journal of Geosciences*, s87–s104.

Kilb, D., J. Gomberg, and P. Bodin (2002). "Aftershock triggering by complete Coulomb stress changes". In: *J. Geophys. Res.* (B4), p. 2060.

King, G. C. P., R. S. Stein, and J. Lin (1994). "Static stress changes and the triggering of earthquakes". In: *Bull. Seis. Soc. Am.* 84, pp. 935–953.

Kiraly-Proag, E., J. D. Zechar, V. Gischig, S. Wiemer, D. Karvounis, and J. Doetsch (2016). "Validating induced seismicity forecast models – Induced Seismicity Test Bench". In: *J. Geophys. Res.* Pp. 6009–6029.

Kortekaas, M. and B. Jaarsma (2017). "Improved definition of faults in the Groningen field using seismic attributes". In: *Netherlands Journal of Geosciences* 96.5, s71–s85.

Kumazawa, T. and Y. Ogata (2014). "Nonstationary ETAS models for nonstandard earthquakes". In: *Annals Applied Statistics*, pp. 1825–1852.

Langenbruch, C., M. Weingarten, and M. D. Zoback (2018). "Physics-based forecasting of man-made earthquake hazards in Oklahoma and Kansas". In: *Nature Communications*. DOI: 10.1038/s41467-018-06167-4.

Langenbruch, C. and M. D. Zoback (2016). "How will induced seismicity in Oklahoma respond to decreased saltwater injection rates?" In: *Sci. Adv.* e1601542.

Lanz, F., K. Bisdom, E. Barbaro, J. Limbeck, T. Park, C. Harris, and K. Nevenzeel (2019). *Evaluation of a Machine Learning methodology for spatio-temporal induced seismicity forecasts within the Groningen field - Machine Learning*. Tech. rep. NAM.

Lapusta, N., J.R. Rice, Y. Ben-Zion, and G. Zheng (2000). "Elastodynamic analysis for slow tectonic loading with sponateous rupture episodes on faults with rate- and state-dependent friction". In: *J. Geophys. Res.* Pp. 23765–23789.

Lele, S. P., S.-Y. Hsu, J. L. Garzon, N. DeDontney, K. H. Searles, G. A. Gist, and B. A. Dale (2016). "Geomechanical modeling to evaluate production-induced seismicity at Groningen Gas Field". In: *Abu Dhabi International Petroleum Exhibition and Conference. Abu Dhabi, UAE: Society of Petroleum Engineers*.

Limbeck, J., F. Lanz, E. Barbaro, C. Harris, K. Bisdom, T. Park, W. Oosterbosch, H. Jamali-Rad, and K. Nevenzeel (2018). *Evaluation of a Machine Learning methodology to forecast induced seismicity event rates within the Groningen Field*. Tech. rep. NAM.

Linker, M. F. and J. H. Dieterich (1992). "Effects of variable normal stress on rock friction: Observations and constitutive equations". In: *J. Geophys.Res.* Pp. 4923–4940.

Llenos, A. L. and A. J. Michael (2013). "Modeling earthquake rate changes in Oklahoma and Arkansas: Possible signatures of induced seismicity". In: *Bull. Seismol. Soc. Am.* Pp. 2850–2861.

Llenos, A. L. and A. J. Michael (2019). "Ensembles of ETAS models provide optimal operational earthquake forecasting during swarms: Insights from the 2015 San Ramon, California Swarm". In: *Bull. Seismol. Soc. Am.* DOI: 10.1785/0120190020.

Llenos, A. L. and N. J. van der Elst (2019). "Improving earthquake forecasts during swarms with a duration model". In: *Bull. Seismol. Soc. Am.* Pp. 1148–1155.

Marsan, D. (2006). "Can coseismic stress variability suppress seismicity shadows? Insights from a rate-and-state friction model". In: *J. Geophys. Res.* B06305. DOI: 10.1029/2005JB004060.

Marsan, D., E. Prono, and A. Helmstetter (2013). "Monitoring aseismic forcing in fault zones using earthquake time series". In: *Bull. Seismol. Soc. Am.* Pp. 169–179.

Michael, A. J. and M. J. Werner (2018). "Preface to the focus section on the Collaboratory for the Study of Earthquake Predictability (CSEP): New results and future directions". In: *Seismol. Res. Lett.* (4), pp. 1226–1228.

Mignan, A. and M. Broccardo (2019). "One neuron versus deep learning in aftershock prediction". In: *Nature*. DOI: 10.1038/s41586-019-1582-8.

Muntendam-Bos, A. G. (2020). "Clustering characteristics of gas-extraction induced seismicity in the Groningen gas field". In: *Geophys. J. Int.* in press.
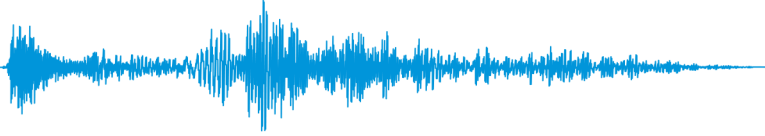
Muntendam-Bos, A.G., J.P.A. Roest, and H.A. de Waal (2017). "The effect of imposed production measures on gas extraction induced seismic risk". In: *Netherlands Journal of Geosciences* 96.5, s271–s278.

Narteau, C., S. Byrdina, P. Shebalin, and D. Schorlemmer (2009). "Common dependence on stress for the two fundamental laws of statistical seismology". In: *Nature*, pp. 642–645.

Norbeck, J.H. and J.L. Rubinstein (2018). "Hydromechanical earthquake nucleation model forecasts onset, peak, and falling rates of induced seismicity". In: *Geophysical Research Letters* 45, pp. 1–13.

Nur, A. and J. R. Booker (1972). "Aftershocks caused by pore fluid flow?" In: *Science* 175, p. 885.

Ogata, Y. (1988). "Statistical models for earthquake occurrence and residual analysis for point processes". In: *J. Am. Stat. Assoc.* 83, pp. 9–27.

Ogata, Y. (1998). "Space-time point-process models for earthquake occurrences". In: *Annals of the Institute of Statistical Mathematics* 50.2, pp. 379–402.

Ohnaka, M. (2013). *The physics of rock failure and earthquakes*. Cambridge University Press.

Okada, Y. (1992). "Internal deformation due to shear and tensile faults in a half-space". In: *Bull. Seism. Soc. Am.* 2.2, pp. 1018–1040.

Perfettini, H. and J.-P. Avouac (2004). "Postseismic relaxation driven by brittle creep: A possible mechanism to reconcile geodetic measurements and the decay rate of aftershocks, application to the Chi-Chi earthquake, Taiwan". In: *J. Geophys. Res.* 109 (B02304). DOI: 10.1029/2003JB002488.

Perfettini, H., W. B. Frank, D. Marsan, and M. Bouchon (2019). "Updip and along-strike aftershock migration model driven by afterslip: Application to the 2011 Tohoku-Oki aftershock sequence". In: *J. Geophys. Res.* 124. DOI: 10.1029/2018JB016490.

Perol, Th., M. Gharbi, and M. Denolle (2018). "Convolutional neural network for earthquake detection and location". In: *Science Advances* 2 e1700578, pp. 4575–4590. DOI: 10.1126/sciadv.1700578121.

Pollitz, F. F., R. Burgmann, and B. Romanowicz (1998). "Viscosity of oceanic asthenosphere inferred from remote triggering of earthquakes". In: *Science* 280, pp. 1245–1249.

Rice, J. R. (1993). "Spatio-temporal complexity of slip on a fault". In: *J. Geophys. Res.* Pp. 9885–9907.

Richter, G., S. Hainzl, T. Dahm, and G. Zöller (2019). *Stress-based, statistical modeling of the induced seismicity at the Groningen Gas Field*. submitted to Environmental Earth Sciences.

Ruina, A. (1983). "Slip instability and state variable friction laws". In: *J. Geophys.Res.* Pp. 359–370.

Sanz, P. F., S. P. Lele, K. H. Searles, S.-Y. Hsu, J. L. Garzon, J. A. Burdette, W. E. Kline, B. A. Dale, and P. D. Hector (2015). "Geomechanical analysis to evaluate production-induced fault reactivation at Groningen gas field". In: *SEG Annual Technical Conference and Exhibition, Houston, Texas*.

Schmedes, J., S. Hainzl, S.-K. Reamer, F. Scherbaum, and K.-G. Hinzen (2005). "Moment release in the Lower Rhine Embayment, Germany: Seismological perspective of the deformation process". In: *Geophys. J. Int.* Pp. 901–909.

Scholz, C. H. (2002). *The mechanism of earthquakes and faulting*. 2nd ed. Cambridge.

Segall, P. (1989). "Earthquakes triggered by fluid extraction". In: *Geology* 17.10, pp. 942–946.

Segall, P. and S.D. Fitzgerald (1998). "A note on induced stress changes in hydrocarbon and geothermal reservoirs". In: *Tectonophysics* 289, pp. 117–128.

Segall, P. and S. Lu (2015). "Injection-induced seismicity: poroelastic and earthquake nucleation effects". In: *Journal of Geophysical Research: Solid Earth* 120.7, pp. 5082–5103.

SGS Horizon B.V. (2016). *Independent review of Groningen subsurface modelling update for winningsplan 2016 with opinion letter*. Tech. rep. SGS Horizon B.V.

Shapiro, S. A. (2018). "Seismogenic index of underground fluid injections and productions". In: *J. Geophys. Res.* Pp. 7983–7997.

Shapiro, S. A., C. Dinske, C. Langenbruch, and F. Wenzel (2010). "Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations". In: *Lead. Edge*, pp. 304–309.

Spiers, C.J., S.J.T. Hangx, and A.R. Niemeijer (2017). "New approaches in experimental research on rock and fault behaviour in the Groningen gas field". In: *Netherlands Journal of Geosciences* 96.5, s55–s69.

Stein, R. S. (1999). "The role of stress transfer in earthquake occurrence". In: *Nature* 402.6762, pp. 605–609.

Toda, S., R.S. Stein, and T. Sagiya (2002). "Evidence from the AD 2000 Izu islands earthquake swarm that stressing rate governs seismicity". In: *Nature* 419, pp. 58–61.

Uenishi, K. and J. R. Rice (2003). "Universal nucleation length for slip-weakening rupture instability under nonuniform fault loading". In: *J. Geophys. Res.* P. 2042. DOI: 10.1029/2001JB00168.

Utsu, T., Y. Ogata, and R. S. Matsu'ura (1995). "The centenary of the Omori formula for a decay law of aftershock activity". In: *J. Phys. Earth* 43, pp. 1–33.

van den Bogert, P. A. J. (2018). *Depletion-induced fault slip and seismic rupture 2D Geomechanical models for the Groningen field, The Netherlands*. Tech. rep. NAM.

van der Elst, N.J., M. T. Page, D. A. Weiser, Th. H.W. Goebel, and S. M. Hosseini (2016). "Induced earthquake magnitudes are as large as (statistically) expected". In: *J. Geophys. Res. Solid Earth*, pp. 4575–4590. DOI: 10.1002/2016JB012818.

van Stiphout, T., J. Zhuang, and D. Marsan (2012). "Seismicity declustering". In: *Community Online Resource for Statistical Seismicity Analysis*. DOI: 10.5078/corssa-52382934.

van Wees, J.D., L. Buijze, K. van Thienen-Visser, M. Nepveu, B.B.T. Wassing, B. Orlic, and P.A. Fokker (2014). "Geomechanics response and induced seismicity during gas field depletion in the Netherlands". In: *Geothermics* 52, pp. 206–219.

van Wees, J.-D., P.A. Fokker, K. Van Thienen-Visser, B.B.T. Wassing, S. Osinga, B. Orlic, S.A. Ghouri, L. Buijze, and M. Pluymaekers (2017a). "Geomechanical models for induced seismicity in the Netherlands: inferences from simplified analytical, finite element and rupture model approaches". In: *Netherlands Journal of Geosciences* 96.5, s183–s202.

van Wees, J.-D., S. Osinga, K. Van Thienen-Visser, and P.A. Fokker (2017b). "Reservoir creep and induced seismicity: inferences from geomechanical modeling of gas depletion in the Groningen field". In: *Geophysical Journal International* 212.3, pp. 1487–1497.

van Wees, J.-D., M. Pluymaekers, S. Osinga, P. Fokker, K. van Thienen-Visser, B. Orlic, B. Wassing, D. Hegen, and T. Candela (2019). "3D mechanical analysis of complex reservoirs: a novel mesh-free approach". In: *Geophysical Journal International* 219, pp. 1118–1130.

Vidale, J. E. and P. M. Shearer (2006). "A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers". In: *J. Geophys. Res.* B05312. DOI: 10.1029/2005JB004034.

Vlek, C. (2019). "Rise and reduction of induced earthquakes in the Groningen gas field, 1991–2018: statistical trends, social impacts, and policy change". In: *Environmental Earth Sciences*. DOI: 10.1007/s12665-019-8051-4.

Wentinck, H.M. (2017). *Kinematic modelling of large tremors in the Groningen field using extended seismic sources: Huizinge earthquake part 1*. Tech. rep. Nederlandse Aardolie Maatschappij BV.

Wentinck, H.M. (2018a). *Dynamic modelling of large tremors in the Groningen field using extended seismic sources*. Tech. rep. Nederlandse Aardolie Maatschappij BV.

Wentinck, H.M. (2018b). *Kinematic modelling of large tremors in the Groningen field using extended seismic sources: First results related to the Zerijp tremor M_L 3.4*. Tech. rep. Nederlandse Aardolie Maatschappij BV.

Wentinck, H.M. (2018c). *Kinematic modelling of large tremors in the Groningen field using extended seismic sources: Huizinge earthquake part 2*. Tech. rep. Nederlandse Aardolie Maatschappij BV.

Wentinck, R. (2015). *Induced seismicity in the Groningen field: A statistical assessment of tremors along faults in a compacting reservoir*. Tech. rep. NAM.

Willacy, C., E. van Dedem, S. Minisini, J. Li, J.W. Blokland, I. Das, and A. Droujinine (2018). "Application of full-waveform event location and moment-tensor inversion for Groningen induced seismicity". In: *The Leading Edge* 37.2, pp. 92–99.

Willacy, C., E. van Dedem, S. Minisini, J. Li, J.-W. Blokland, I. Das, and A. Droujinine (2019). "Full-waveform event location and moment tensor inversion for induced seismicity". In: *Geophysics* 84.2, KS39–KS57.

WP1 (2019). "Review of seismogenic source models for the Groningen gas reservoir: Review of the existing NAM seismogenic source model". In:

Zaliapin, I. and Y. Ben-Zion (2016). "Discriminating characteristics of tectonic and human-induced seismicity". In: *Bull. Seismol. Soc. Am.* 106, pp. 846–859. DOI: 10.1785/0120150211.

Zaliapin, I., A. Gabrielov, V. Keilis-Borok, and H. Wong (2008). "Clustering analysis of seismicity and aftershock identification". In: *Physical Review Letters* 101, p. 018501. DOI: 10.1103/PhysRevLett.101.018501.

Zbinden, D., A. P. Rinaldi, L. Urpi, and S. Wiemer (2017). "On the physics-based processes behind production-induced seismicity in natural gas fields". In: *J. Geophys. Res.* Pp. 3792–3812.

Zhuang, J., D. Harte, M. J. Werner, S. Hainzl, and S. Zhou (2012). "Basic models of seismicity: temporal models". In: *Community Online Resource for Statistical Seismicity Analysis*. DOI: 10.5078/corssa-79905851.

Ziv, A. and A.M. Rubin (2003). "Implications of rate-and-state friction for properties of aftershock sequences: Quasi-static inherently discete simulations". In: *J. Geophys. Res.* P. 2051. DOI: 10.1029/2001JB001219.

Zoback, M.D. (2010). *Reservoir Geomechanics*. Cambridge University Press.

Zöller, G. and M. Holschneider (2016a). "The earthquake history in a fault zone tells us almost nothing about m_{max}". In: *Seismol. Res. Lett.* DOI: 10.1785/0220150176.

Zöller, G. and M. Holschneider (2016b). "The maximum possible and the maximum expected earthquake magnitude for production-induced earthquakes at the gas field in Groningen, The Netherlands". In: *Bull. Seismol. Soc. Am.* Pp. 2917–2921.

Report Number: 19-010	Confidential: X Unlimited:	External: X Internal:	NORSAR Project No.: 10178
Title:	Review of seismogenic source models for the Groningen gas reservoir (WP2: Review of existing alternative approaches and possible knowledge gaps)		
Client:	Staatstoezicht op de Mijnen, Netherlands		
Project manager:	D. Kühn		
Authors/prepared by (alphabetical order):	T. Dahm (GFZ), S. Hainzl (GFZ), D. Kühn (NORSAR), V. Oye (NORSAR), G. Richter (GFZ), I. Vera Rodriguez (NORSAR)		
Submitted to:	Staatstoezicht op de Mijnen, Netherlands		
Contract reference:			
Archive reference:			
Approved by: Department head:	Name: V. Oye	Signature: *Volker Oye*	Date: 24.04.2020