A. G. Muntendam-Bos[1,2], and N. Grobbe[2]

[1]Department of Geoscience and Engineering, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, the Netherlands.

[2]Dutch State Supervision of Mines, P.O. Box 24037, 2490 AA The Hague, the Netherlands.

Corresponding author: Annemarie G. Muntendam-Bos (a.g.muntendam-bos@tudelft.nl)

Key Points:

- The earthquake-size distribution of Groningen induced earthquakes varies spatially with indications of a temporal decrease.
- This spatiotemporal pattern seems consistent with a stress-dependent slope of the earthquake-size distribution.
- The earthquake-size distribution is most likely tapered at a single, regional corner magnitude.

Abstract

The scaling of event sizes is one of the critical parameters controlling seismic hazard and risk. For induced seismicity, the non-stationary, heterogeneous character of subsurface stress perturbations can be a source of spatiotemporal variations in the scaling of event sizes. We analyse the spatiotemporal variations in the earthquake-size distribution of the Groningen induced seismicity catalogue using a method that circumvents arbitrary choices requiring a priori knowledge of these variations while systematically exploring the effect of possible bias in the derivation. Our results show that the b-value is spatially variable with indications of a temporal decrease. This spatiotemporal pattern may be explained by a stress-dependent b-value. At the same time, we postulate that the event-size distribution is tapered at a single, regional corner magnitude. Our results imply that the current risk assessment models overestimate the probability of larger magnitude events (M 3.5) in the Groningen gas field and thus the risk posed.

Plain Language Summary

One of the most important parameters determining the outcome of a seismic hazard and risk analysis is the scaling of earthquakes of different magnitudes. For induced seismicity, the scaling distribution can be affected by the non-stationary, heterogeneous character of the subsurface operation. In this paper, we have assessed to what extend the scaling distribution of the Groningen gas field induced seismicity catalogue is indeed affected by the operation. Our results show that the scaling is spatially variable with indications of a temporally increasing probability of larger magnitude events. This pattern may be explained by the spatiotemporally variable stresses induced by the gas extraction. Our results imply that the current risk assessment models overestimate the probability of

larger magnitude events (M 3.5) in the Groningen gas field and thus the risk posed.

1 Introduction

The scaling of earthquake sizes, the amount and temporal occurrence of events and the maximum possible magnitude are critical parameters that control seismic hazard and risk. In case of induced seismicity, the time-varying increases in stress due to gas extraction that are sufficient to destabilize previously inactive faults cause the seismicity to be a transient, non-stationary process which should be accounted for in the assessment of induced seismicity hazard and risk. Several models of forecasting induced earthquake occurrences under these non-stationary, heterogeneous conditions have been proposed (e.g. Bourne et al., 2014; 2018; Shapiro et al., 2010; Mignan et al., 2017; Richter et al., 2020). However, in most hazard and risk assessments the earthquake-size distribution is considered a stationary pure power-law distribution of seismic moments or magnitudes (Langenbruch & Zoback, 2016; Petersen et al., 2018; Shapiro, 2018).

At the same time, spatial and temporal variations in the scaling of event sizes have been reported and attributed a physical meaning (e.g. Bachmann et al., 2012; Gulia et al., 2016; Hiemer & Kamer, 2016; Muntendam-Bos et al., 2017; Schorlemmer et al. 2005). The impact of these spatio-temporal variations on the hazard and risk assessment can be quite significant (Gulia et al., 2016; Muntendam-Bos et al., 2017). Hiemer and Kamer (2016) showed that the performance of the Californian forecast models could be significantly improved when including the large-scale spatial variations in the slope of the power-law distribution (the b-value). However, notwithstanding the vast literature on spatiotemporal variations, care should be taken as potential bias due to evaluation of a finite data set may lead to non-physical variations caused by the correlation between the b-value and the maximum magnitude of the dataset (Marzocchi et al., 2020). In addition, the classical mapping technique (Wiemer and Wyss, 1997; 2002) depends heavily on external parameters for which the choice requires a priori knowledge of the spatial or temporal event-size distribution that one wishes to resolve in the first place (Kamer, 2014).

In this paper we assess whether spatiotemporal variations in the earthquake-size distribution exist in the Groningen induced seismicity catalogue. For the large Groningen gas field, the seismic risk assessment informs decision making. This emphasizes the importance of a proper derivation of potential spatiotemporal variations. First, we analyze possible spatial variations in the b-value. To circumvent the arbitrary choices in external mapping parameters, we use the method introduced by Kamer and Hiemer (2015). Second, we include the temporal dimension in the analysis and systematically explore the effect of possible sources of bias on the derived spatiotemporal variations.

2 Data

Natural gas has been produced from the Groningen gas field since 1963. At present about 70% of the estimated 2800 x 10^9 m^3 initial gas in place has

been produced, dropping the initial mean pore pressure by up to 25 MPa. The field is located in the sandstones of the Rotliegend formation, which is overlain by a thick layer of Zechstein halite and anhydrite salt deposits (De Jager and Visser, 2017). The reservoir is highly faulted with over 1100 mapped normal (extensional) faults (Figure 1a). The earthquakes in the Groningen gas field are induced by gas extraction at a depth of approximately 3 km and have relatively small magnitudes ($M_l \leq$ 3.6) (Figure 1a; Dost et al., 2017). A local geophone network with a detection threshold of local magnitude $M_l = 1.5$ was installed in 1995 (Dost et al., 2017). In 2015, the geophone network was significantly extended to increase the detection of small earthquakes (Dost et al., 2017). In total over 1400 events were detected between December 1991 and Januari 1st, 2021, ranging from local magnitudes M_l = -0.5 to 3.6 with thirteen events with a $M_l \geq 3$ (e.g. Dost et al., 2017). In this paper, we use the catalogue reported by the Royal Dutch Meteorological Institute (KNMI; www.knmi.nl) from which we select all events within the outline of the Groningen gas field (Figure 1a).

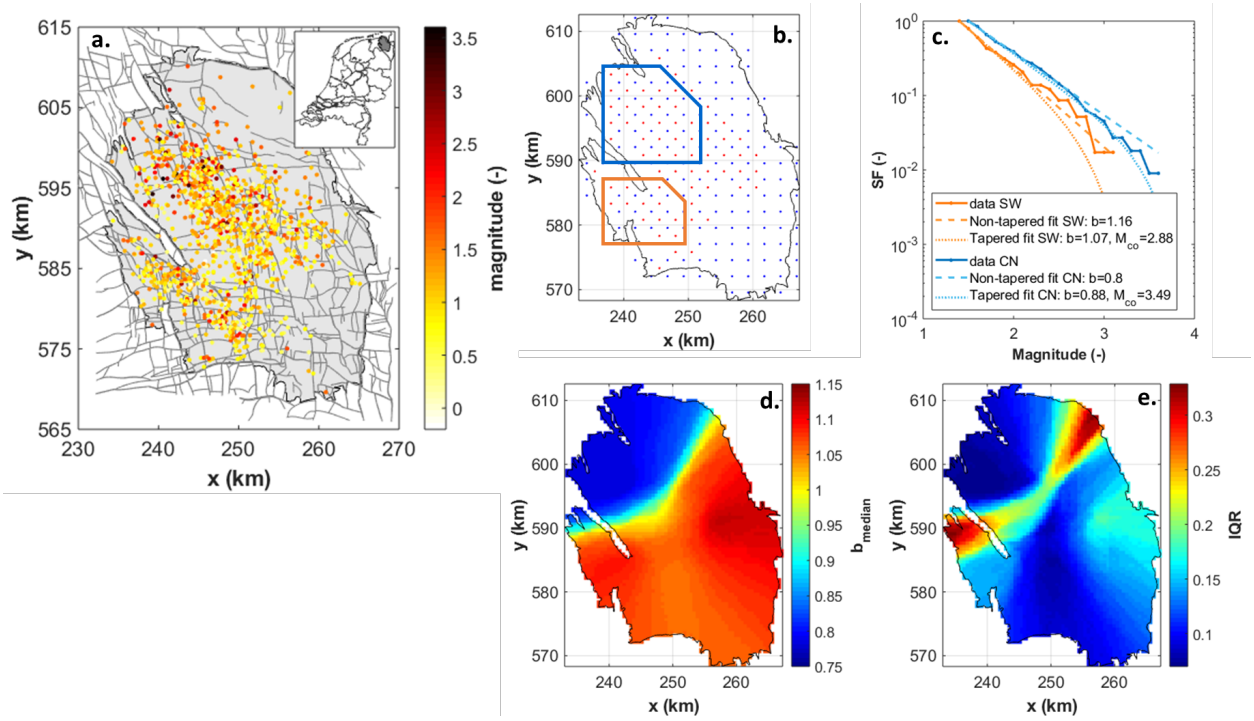

Figure 1. a. Epicentre locations of the Groningen induced seismicity. b. Overview of the 2.5 x 2.5 km grid node locations (blue) and the Voronoi node locations (red) at the centre of the grid cells containing at least 2 events $M \geq M_c$. The blue and orange contours indicate the central-north (CN) and southwestern (SW) region, respectively. c. The observed event-size distributions for the SW (orange) and CN (blue) regions and the non-tapered and tapered model fit. d. Ensemble median b-value and e. the corresponding ensemble interquartile range

of the best 1000 solutions.

3 Earthquake-size distribution

The relation between the cumulative number of earthquakes (N) and magnitude follows a power law distribution expressed as $\log_{10} N = a - b(M - M_c)$ (Gutenberg and Richter, 1944), where M_c is the magnitude of completeness and a and b (so-called b-value) are constants that describe the productivity and the relative size distribution, respectively.

In the analysis presented in this paper, the b-value is determined with the maximum likelihood method (Aki, 1965; Utsu,1965). In order to avoid bias in the b-value estimates (Marzocchi et al. 2020), we implement a correction for magnitude binning (Marzocchi et al. 2020) and small sample sizes (Ogata & Yamashina, 1986).

The (regional) magnitude of completeness is calculated with the maximum curvature method (MCM; Woesner and Wiemer, 2005). The advantage of the MCM is that results can be obtained fast and reliably even for small sample sizes. On the other hand, the method tends to underestimate M_c, especially for gradually-curved frequency-magnitude distributions. This disadvantage can be overcome by using a correction factor ($M_c = M_c\,(\mathrm{MCM}) + M_c$) in combination with the bootstrap approach (Woesner and Wiemer, 2005). After careful assessment of the (regional) MCM results while increasing the correction factor (M_c), a correction factor of M_c =0.3 was adopted. We use the bootstrap approach also to reliably estimate the uncertainties in both the b-values and M_c (Schorlemmer et al., 2003; Amorese et al. 2010; Marzocchi et al. 2020).

4 Spatial variations in earthquake-size distribution

To assess the spatial variations of the b-values, we implement the penalized likelihood-based method of Kamer and Hiemer (2015). This method is based on optimal partitioning using Voronoi tessellation, penalized likelihood, and the wisdom of the crowd philosophy. This approach allows for a flexible, non-overlapping partioning in space and the computation of the overall log-likelihood of each random tessellation (Kamer and Hiemer, 2015). The Voronoi selection space is dicretized in accordance with the event location uncertainties in 2.5x2.5 km cells (Figure 1b). Only cells with more than one event $M \geq M_c$ are considered a potential Voronoi node location (Figure 1b). The number of tessellation nodes considered in an analysis is increased from 1 to 15 and each node configuration is spatially randomized 2000 times. For each Voronoi cell in each configuration both the M_c and the b-value are estimated. If no M_c can be derived due to too small a sample size, the M_c of the null-hypothesis ($M_c = 1.2$) is adopted with an additional consistency check for larger M_c. All models are ranked by their penalized likelihood using the Bayesian Information Criterium (Schwarz, 1978). The ensemble model is calculated using the best 1000 solutions, which performs better than the simplest (prior) model of no spatiotemporal variations in the b-value for the Groningen gas field ($b = 0.92 \pm 0.04$). The resulting median b-values range from 0.78 to 1.12 (Figure 1d). We observe a very sys-

tematic division of low b-values in the north-northwest of the field and higher b-values in the south. The highest b-values are observed in the west-southwest and east of the field. The transition between the low and high b-value region is less well defined and associated with a larger interquartile range (Figure 1e). All but one larger magnitude event ($M \geq 3.0$; LME) occurred within the low b-value area or transition zone. We ascertained whether the spatial pattern in b-values may be a consequence of the presence of these LMEs (by excluding these from the analysis). However, even though the b-value in the low b-value region increased as the largest magnitude in the sample decreased (consistent with the observations in Marzocchi et al., 2020), the spatial pattern obtained remained identical.

To further assess the statistical significance of the spatial pattern, we focus on the frequency-magnitude distributions of the low b-value or central-north (CN) and high b-value west-southwest (SW) regions (as indicated in Figure 1b). For both regions we computed the regional b-value and M_c based on all events contained within each region. We obtain regional b-values of $b = 0.80 \pm 0.07$ and $b = 1.16 \pm 019$, respectively (Figure 1c; Table 1). The b-value in the SW is slightly larger than the range obtained in the Voronoi analysis. This is consistent with observations of Kamer and Hiemer (2015) that b-values in a high b-value area may be underestimated by the Voronoi approach.

Table 1. Overview of the parameter estimates of the (tapered) earthquake-size distributions and the AICc test statistics.

Region	Period		Non-tapered	Tapered; two parameter estimation.			
		$M_{\max}^{\text{obs}}$	b	AICc	b	M_{co}	AICc
SW	full	2.8	1.16±0.19	2710.8	1.07	2.88	2712
	Pre 1 Jan. 2013	2.6	1.39±0.39	1111.8	1.54	2.94	1112
	Post 1 Jan. 2013	2.8	0.95±0.19	2531.5	1.23	2.87	2532
CN	full	3.6	0.80±0.07	12791	0.88	3.49	1278
	Pre 1 July 2003	2.7	1.03±0.18	1181.1	0.70	2.49	1180
	Post 1 July 2003	3.6	0.77±0.06	12137	0.83	3.48	1213
	Pre 1 Jan. 2013	3.4	0.85±0.08	4833.6	0.78	3.47	4834
	Post 1 Jan. 2013	3.6	0.77±0.08	6195.8	0.77	3.35	6194

We use the two sample, left-tailed t-test or Welch's test (Boslaugh, 2012) to assess whether the two b-value distributions may be samples of the same distribution. We find that the two regional b-values are statistically different at the 97.5% confidence level. Compared to the null hypothesis of a constant b-value for the full catalog ($b = 0.92 \pm 0.04$), the b-value of the SW-region is statistically different at the 95% confidence level and the CN b-value at the 92.5% confidence level.

5 Spatiotemporal variations in earthquake-size distribution

We now assess the systematic behavior of the seismicity further by analyzing the temporal dimension as well. Due to the heterogeneous spatiotemporal development of the Groningen seismicity, the Voronoi analysis carried out on the full Groningen field extended in the temporal dimension showed large interquartile ranges and did not render meaningful results. Therefore, we assessed whether any indication of a temporal dependence is present in the two regions identified in the spatial analysis. For each region, we follow Gulia et al. (2016) to obtain a single, regional continuous temporal b-value series (Figure 2). To derive the *b*-values a minimum sample size of 20 events is adopted, which reasonably balances the uncertainty in the derived *b*-value estimate versus the temporal resolution. Nonetheless, similar results were obtained over a wide range of sample sizes indicating robustness of the results.

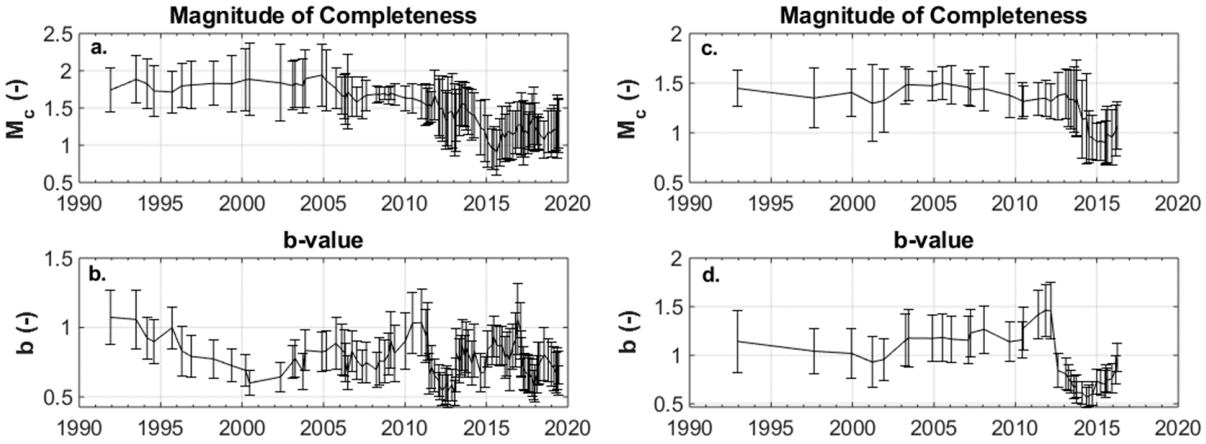

Figure 2. Temporal variations in M_c and *b*-value for the CN- (a,b) and the SW-region (c,d). Note that the b-values derived are plotted at the start of their respective analysis window.

The dataset of each window is derived by moving the onset of the window through the catalogue with a 5 event shift, selecting the following 20 events in the catalog, computing M_c, and deriving the catalogue of events $N(M \quad M_c)$. If the catalogue contains insufficient events ($<20 \quad M_c$), the window is extended event-by-event. Following each extension, M_c is recomputed and the length of the catalogue $N(M \quad M_c)$ reassessed. This process is repeated until 50 events exceed the magnitude of completeness of all the events in the window. The 5 event shift (instead of the standard 1 event) was necessary to ensure no identical datasets were obtained due to the addition of only a $M<M_c$ event.

The resulting time series for M_c and *b*-value are shown in Figure 2. In the CN-region, we observe a clear decrease in M_c following the network extension in 2014. The *b*-value series shows a decrease of the b-value between 1995 and 2000 and

fluctuations with time after 2005, which are most likely related to incompleteness of the short datasets assessed. In fact, the strong decreases between 2011 and mid 2013 and 2017/2018 correspond to the inclusion of the large $M_L = 3.6$ Huizinge and $M_L = 3.4$ Zeerijp events, respectively. The decreasing trend between 1995 and 2000 shows no fluctuation despite the inclusion of the first magnitude 3.0 events in 2003 (first included in the window starting 1-4-1996) or 3.5 event in 2006 (first included in the window starting 1-1-2003). The Welch test shows that the decrease in b-value from $b = 1.03 \pm 0.18$ for the data in this region before July 1st, 2003, to $b = 0.77 \pm 0.06$ for the data in this region after July 1st, 2003 (Table 1), is significant at the 85% confidence level. The division at July 1st, 2003, was chosen as the last analysis window before the observed decrease in Figure 2 for the range 1-4-1996 till 1-7-2003 (Figure 2).

The temporal *b*-values in the SW region show a distinct decrease as soon as only events after January 1st, 2013 are included in the assessment window. Analysis of the data in two periods, before and after the strong decrease (division made at January 1st, 2013), shows a clear decrease of the *b*-value from $b = 1.39 \pm 0.39$ to $b = 0.95 \pm 0.19$. Despite this apparently large decrease in *b*-value, given the very small sample size and hence the large standard error in the *b*-value for the early period, the Welch test renders this difference also only significant at the 85% confidence level.

6 Bias introduced by tapering or truncation of the earthquake-size distribution?

The limited capability to accumulate seismic energy in one specific region or on a single fault requires the earthquake-size distribution to decay stronger above a particular magnitude called the corner magnitude M_{co} or even be truncated at a hard upper magnitude cutoff M_{up}. This tapering or truncation of the distribution may introduce a bias upwards in the estimation of the *b*-value for $M_{up/co} - M_c < 3$ (Marzocchi et al.,2020; Bourne and Oates, 2020). As the Groningen catalogue is limited in magnitude range with M_c ranging from 0.8 to 1.7 (Figure 2) and a maximum observed magnitude of $M_L = 3.6$, our derived spatiotemporal *b*-value estimates may be influenced. Here, we explore this possible bias by performing a maximum likelihood analysis for the tapered distribution to derive estimates of both the *b*-value and corner magnitude M_{co} (Kagan, 2002).

For the CN region, the derived corner magnitudes (Table 1) are comparable to the corner magnitude estimate of the full Groningen catalogue ($M_{\mathrm{co}} = 3.43$) except for the early period, which shows a much lower corner magnitude of ~2.5, close to the maximum magnitude observed (M=2.7). The corner magnitudes for the SW region are lower and even higher than the maximum magnitude observed in each period (Table 1). This raises the question whether the implementation of a taper is warranted by the data. We use the corrected Akaike Information Criterion (AICc; Cavanaugh, 1997) to assess and compare the fit of the tapered and non-tapered models to the data. All AICc-values are given in Table 1. For the SW-region, the non-tapered model is the most likely; the relative likelihood of the tapered model is only 17%. The opposite is true for the CN-region for

both the full period and after July 1st, 2003. Here, the tapered model is the most likely model.

Based on global natural seismicity, Kagan (1999) strongly favours models with a constant b-value and an exponential taper. This is further supported by recent developments in statistical fracture and earthquake mechanics theories for highly-disordered media (Bourne and Oates, 2020). In fact, estimations with decreasing b-values could be an artifact of an actual increase in corner magnitude (Bourne and Oates, 2020). At the same time, the maximum likelihood estimates of the corner magnitude can be heavily biased for small sample sizes. A reliable estimate requires that a catalog includes a few earthquakes close in magnitude to the corner magnitude (Kagan, 2002). The fact that four out of six of our analyses derive corner magnitudes comparable to the largest magnitude observed indicates that this is most likely not the case. Only the full CN-catalogue and post January 1st, 2003 data of the CN-region appear consistent with the presence of a possible taper, as indicated by the relative likelihood of the non-tapered models of 8% and 13.5%, respectively.

7 Discussion

Our results suggest that there are indications of spatiotemporal variations of the earthquake-size distribution in the induced seismicity sequence of Groningen. For the SW-region, the non-tapered model with a relatively high, temporally decreasing b-values is the best performing model. In contrast, the best performing model for the CN-region is the tapered model. However, the limited dataset of the Groningen catalogue leads to large standard errors in the estimations lowering the confidence levels at which conclusions on the spatiotemporal distribution can be drawn, especially in the temporal domain.

Our results raise the question whether indeed there exits an intrinsic limit on the maximum size of the induced earthquakes. Following Van der Elst et al. (2016), we computed the statistically expected maximum magnitude for each of the subcatalogues (Figure 3). For the CN-region, we find that in all but the pre July 1st, 2003 periods the observed maximum magnitude is significantly smaller than statistically expected value, and falls outside the 90% confidence range of the expected distribution. For the SW-region, the observed magnitudes are as large as can be statistically expected. Based on these results, we postulate that the event-size distribution is tapered at a single, regional corner magnitude between 3.5 and 4.5.

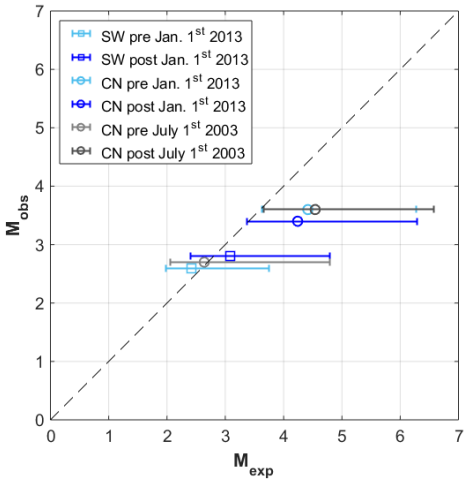

Figure 3. Observed versus expected maximum magnitude. The horizontal bars show the 90% confidence ranges associated with the b-value uncertainty for each region and period. For reference, the diagonal dotted line indicates where the observed and expected maximum magnitudes are equal.

In our opinion, a fundamental mechanism is required to explain the observed, highly systematic spatiotemporal distribution of the b-values in Groningen. In the spatial domain, pre-existing structural heterogeneities such as fault roughness or elastic moduli variations might controle the observed variations. However, these heterogeneities cannot control the observed temporal decrease of the regional *b*-values. The inverse relation between differential stress and b-value variations is a fundamental mechanism which has been established across a large range of scales and successfully applied to injection-induced seismicity (Bachmann et al., 2012). For extraction-induced seismicity the removal of the natural gas causes reservoir depletion and stress changes on the offset faults.

An inverse stress-dependence would also imply a decreasing *b*-value with time as with ongoing depletion stress keeps accumulating on the reservoir faults. Our results show consistent indications of a decreasing *b*-value in the CN-region as soon as events after July 1st, 2003 are included; for the SW-region as soon as events after January 1st, 2013 are included. When analyzing possible stress-dependence of the *b*-value, Bourne & Oates (2020) showed indications of mode switching of the *b*-value with an apparent step-like decrease. A similar stepwise decrease can be seen in Figure 2 for the SW-region between 2010 and 2015. In the CN-region a more gradual decrease is observed between 1995 and 2000. However, considering the very low level of seismicity recorded prior to 2003 and the rather high M_c value up to the window starting in 2005 (Figure 2a) this decrease may still be consistent with a step-like decrease.

Considering our hypothesized fundamental mechanism of stress dependence, the difference in timing of the decrease should then be induced by the differ-

ence in stress history between the CN and SW-regions. Candela et al. (2019) showed that due to the specific fault orientations and offset of the reservoir layer Coulomb stress induced by reservoir depletion indeed cumulated slower on faults in the SW-region than on faults in the CN-region. Considering the inverse relation between stress change and b-value, a stress-dependence of the size distribution could well explain the observed spatial variation in the observed b-values.

In addition to the analysis of the b-values, Bourne & Oates (2020) presented a model forecasting Groningen induced earthquake magnitudes. This model comprised five alternative classes of the earthquake-size distribution, each consisting of a variation between a stress-dependent pure power-law distribution, where the power-law exponent varied with Coulomb stress, and/or a power-law subject to an exponential taper, where the characteristic taper varied with Coulomb stress. From their pseudo-prospective forecast for the earthquake-size distribution, they concluded that the best performing models all included a stress-dependent taper. However, a model including a stress-dependent b-value (i.e. spatiotemporal variations in the b-value) and a stress-invariant exponential taper (i.e. a taper with a single, regional corner magnitude) was not evaluated. Our statistical analysis of the Groningen catalogue shows that the latter could well be the preferred model for the earthquake-size distribution in Groningen.

Finally, we note that in our analysis we reach the limits of the information that can be extracted from the data. Especially for the early periods, the number of events available is very limited and thus the derived b-values are prone to large uncertainties and bias. We have taken great care to minimize the bias in the derivation of the b-values as much as possible, but cannot exclude that some bias due to the small sample sizes remains.

8. Conclusions

- Our results show statistically significant spatiotemporal variations of the earthquake-size distribution in the induced seismicity sequence of Groningen.
- This spatiotemporal pattern seems consistent with a stress-dependent slope of the earthquake-size distribution.
- Our analysis of the expected maximum magnitude shows that the earthquake-size distribution is most likely tapered at a single, regional corner magnitude between 3.5 and 4.5.

Our results imply that the current risk assessment models overestimate the probability of larger magnitude events (M 3.5) in the Groningen gas field and thus the risk posed.

Acknowledgments

This paper benefited from lengthly discussions with Wouter van der Zee and Jort Vermeer.

The earthquake catalogue used in the study is freely available at the KNMI via https://www.knmi.nl/kennis-en-datacentrum/dataset/aardbevingscatalogus (Only available in Dutch).

References

Aki, K. (1965), Maximum likelihood estimate of b in the formula log N=a-bM and its confidence limits. *Bulletin Earthquake Research Institute (Tokyo), 43,* 237–239.

Amorese, D., Grasso, J.-R., & Rydelek, P.A. (2010). On varying b-values with depth: results from computer-intensive tests for Southern California. *Geophysical Journal International, 180(1)*, 347–360.

Bachmann, C.E., Wiemer, S., Goertz-Allman, B.P., & Woessner, J. (2012), Influence of pore-pressure on the event-size distribution of induced earthquakes. *Geophysical Research Letters, 39*, doi:10.1029/2012GL051480.

Boslaugh, S. (2012), Statistics in a Nutshell. O'Reilly Media, Incorporated

Bourne, S. J., & Oates, S. J. (2020), Stress-dependent magnitudes of induced earthquakes in the Groningen gas field. *Journal of Geophysical Research: Solid Earth, 125*, e2020JB020013, doi:10.1029/2020JB020013.

Bourne, S. J., Oates, S. J., & Elk, J. V. (2018), The exponential rise of induced seismicity with increasing stress levels in the Groningen gas field and its implications for controlling seismic risk. *Geophysical Journal International, 213*, 1693–1700, doi: 10.1093/gji/ggy084.

Bourne, S. J., Oates, S. J., van Elk, J., & Doornhof, D. (2014), A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir. *Journal of Geophysical Research: Solid Earth, 119*, 8991–9015, doi:10.1002/2014JB011663.

Candela, T., Osinga, S., Ampuero, J.-P., Wassing, B., Pluymaekers, M., Fokker, P. A., Van Wees, J.-D., De Waal, J.A., & Muntendam-Bos, A.G. (2019), Depletion-induced seismicity at the Groningen gas field:Coulomb rate-and-state models including differential compaction effect. *Journal of Geophysical Research: Solid Earth, 124*, 7081–7104, doi:10.1029/2018JB016670

Cavanaugh, J. E. (1997), Unifying the derivations of the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters, 31 (2)*, 201–208, doi:10.1016/s0167-7152(96)00128-9.

De Jager, J., & Visser, C. (2017), Geology of the Groningen field – an overview. *Netherlands Journal of Geosciences, 95(5)*, s3–s15, doi:10.1017/njg.2017.22.

Dost, B., Ruigrok, E., & Spetzler, J. (2017), Development of seismicity and probabilistic hazard assessment for the Groningen gas field. *Netherlands Journal of Geosciences, 95(5)*, s235–s245, doi:10.1017/njg.2017.20.

Gulia, L., Tormann, T., Wiemer, S., Hermann, M., & Seif, S. (2016), Short-term probabilistic earthquake risk assessment considering time-dependent b values. *Geophys. Res. Lett., 43*, 1100–1108, doi:10.1002/2015GL066686.

Gutenberg, B., & Richter, C.F. (1944), Frequency of earthquakes in California. *Bulletin Seismological Society of America, 34(8)*, 185–188.

Hiemer, S., & Kamer, Y. (2016), Improved seismicity forecast with spatially varying magnitude distribution. *Seismological Research Letters, 87(2A)*, doi:10.1785/0220150182.

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

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

Kamer, Y. (2014). Comment on "Systematic survey of high-resolution b-value imaging along Californian faults: Inference on asperities" by Tormann et al. (2014). *Journal Geophysical Research 119(3)*, 2029–2054.

Kamer, Y., & S. Hiemer (2015), Data-driven spatial b-value estimation with applications to California seismicity: To b or not to b. *Journal Geophysical Research, 120(7)*, 2191–5214.

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

Marzocchi, W., Spassiani, I., Stallone, A., & Taroni, M. (2020), How to be fooled searching for significant variations of the *b*-value. *Geophysical Journal International, 220*, 1845-1856, doi: 10.1093/gji/ggz541.

Mignan, A., Broccardo, M., Wiemer, S., & Giardini, D. (2017), Induced seismicity closed-form traffic light system fro actuarial decision-making during deep fluid injections. *Scientific Reports, 7:13607*, doi:10.1038/s41598-017-13585-9.

Muntendam-Bos, A.G. (2020), Clustering chracteristics of gas-extraction induced seismicity in the Groningen gas field. Geophysical Journal International, 221, 879-892, doi:10.1093/gji/ggaa038.

Muntendam-Bos, A.G., Roest, J.P.A., & De Waal, J.A. (2017), The effect of imposed production measures on gas extraction induced seismic risk. *Netherlands Journal of Geosciences, 95(5)*, s271-s278, doi:10.1017/njg.2017.29.

Ogata, Y., & Yamashina, K. (1986), Unbiased estimate for b-value of magnitude frequency. *Journal Physics Earth, 34*, 187–194.

Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Rukstales, K. S., McNamara, D. E., et al. (2018), 2018 one-year seismic hazard forecast for the central and eastern United States from induced and natural earthquakes. *Seismological Research Letters, 89*, 1049–1061.

Richter, G., Hainzl, S., Dahm, T., & Zöller, G. (2020), Stress-based, statistical modeling of the induced seismicity at the Groningen gas field, The Netherlands. *Environmental Earth Sciences, 79:252*, doi: 10.1007/s12665-020-08941-4.

Schorlemmer, D., Neri, G.,Wiemer, S., & Mostaccio, A. (2003). Stability and significance tests for b-value anomalies: example from the Tyrrhenian Sea. *Geophysical Research Letters, 30(16)*, doi:10.1029/2003GL017335.

Schorlemmer, D., Wiemer, S., & Wyss, M. (2005), Variations in earthquake-size distribution across different stress regimes. *Nature, 437*, 539-542, doi: 10.1038/nature04094.

Schwarz, G. E. (1978), Estimating the dimension of a model. *Annals of Statistics, 6 (2),* 461–464, doi:10.1214/aos/1176344136.

Shapiro, S. A. (2018), Seismogenic index of underground fluid injections and productions. *Journal of Geophysical Research: Solid Earth, 123*,7983–7997, doi:10.1029/2018JB015850

Shapiro, S., Dinske, C., & Langenbruch, C. (2010), Seismogenic index and magnitude probability of earthquakes. *The Leading Edge*, March.

Utsu, T. (1965). A method for determining the value of "b" in a formula log n= a-bm showing the magnitude-frequency relation for earthquakes. *Geophys. Bull. Hokkaido Univ., 13*, 99–103.

Van der Elst, N. J., M. T. Page, D. A. Weiser, T. H. W. Goebel, & S. M. Hosseini (2016), Induced earthquake magnitudes are as large as (statistically) expected. *Journal Geophysical Research Solid Earth, 121*, 4575–4590, doi:10.1002/2016JB012818.

Wiemer, S., and M. Wyss (1997), Mapping the frequency-magnitude distribution in asperities: An improved technique to calculate recurrence times? *Journal Geophysical Research Solid Earth,* 102(B7), 15,115–15,128, doi:10.1029/97JB00726.

Wiemer, S., and M. Wyss (2002), Mapping spatial variability of the frequency-magnitude distribution of earthquakes, *Advances Geophysics*, 45, 259–302.

Woessner, J., & Wiemer, S. (2005). Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty. *Bulletin Seismological Society of America, 95(2)*, 684–698.