Scientific Paper

Weibull fitting of earthquakes in Groningen

Frank P. Pijpers

May 2018

Contents

Nederlands

Deze rapportage behelst een voortzetting van onderzoek dat is uitgevoerd sinds midden 2014 in het kader een onderzoeksproject door het CBS in opdracht van Staatstoezicht op de Mijnen (SodM). Het betreft onderzoek ter ondersteuning van het meet- en regelprotocol van de gaswinning in Groningen.

In dit rapport ligt de aandacht op een analyse van de tijdsintervallen tussen aardbevingen binnen dezelfde gebieden die in de overige CBS rapportages ook gebruikt worden. De gebruikte dataset is de aardbevingscatalogus gepubliceerd door het KNMI, op basis van de metingen gedaan met het netwerk van seismometers, bijgewerkt tot 1 Mei 2018. Voor de meest recente periodieke CBS rapportage zie Pijpers and van Straalen (2018).

Het uitgangspunt voor deze analyse is om de aardbevingen niet als een onderling onafhankelijk (Poisson) proces te beschouwen, waarvoor de distributiefunctie van tijdsintervallen tussen opeenvolgende aardbevingen zou voldoen aan een exponentiële verdelingsfunctie. In plaats daarvan wordt voor de tijdsintervallen een twee-parameter Weibull verdeling verondersteld, waarbinnen geen onderscheid gemaakt wordt tussen bevingen en naschokken : deze behoren allen tot hetzelfde proces.

Deze Weibull verdeling lijkt goed te voldoen om de waargenomen verdelingsfuncties van tijdsintervallen te beschrijven, en de parameters van deze verdeling blijken te verschillen van gebied tot gebied een ook vóór en na Jan. 2014, resp. 2015: de momenten waarop substantiële wijzigingen in gasproductiestrategie zijn geïmplementeerd.

English

This report is a continuation of research, commenced in 2014, which is part of a research project being carried out by Statistics Netherlands and commissioned by State Supervision of Mines (SodM). This research is part of the underpinning of the statistical methods employed to support the protocol for measurement and regulation of the production of natural gas in the province of Groningen.

In this report, the focus is on an analysis of the time intervals between earthquakes occurring within the same zones as also used in the periodic CBS reporting updates. The dataset used is the earthquake catalogue published by the Royal Netherlands Meteorological Institute (KNMI), based on their processing of the network of seismometers that they manage, updated up to May 1st 2018.

Starting point for the present analysis is to regard the earthquakes not as independent (Poisson process) events, for which the distribution function for time intervals would be an exponential distribution. Instead, a two-parameter Weibull distribution for the time intervals is presumed to apply. Within this framework there is no distinction between those earthquakes caused directly and so-called aftershocks, induced by recent (nearby) preceding earthquakes.

It appears that the Weibull distribution is appropriate as a descriptor of the observed distribution functions of time intervals for the several regions, and for the different epochs, and that the parameters of the distribution function are different between different zones and epochs. The epochs distinguished is this report are before or after Jan. 2014 or Jan. 2015 respectively: the moments at which substantial changes in gas production strategy were implemented.

1 Introduction

For some decades earthquakes of modest magnitudes have occurred in the Groningen gas field. It is recognized that these events are induced by the production of gas from the field. Following an $M_L = 3.6$ event near Huizinge, and the public concern that this raised, an extensive study program has started into the understanding of the hazard and risk due to gas production-induced earthquakes.

A protocol needs to be established with the aim of mitigating these hazards and risks by adjusting the production strategy in time and space. In order to implement this regulation protocol and adaptively control production it is necessary also to measure the effects on subsidence and earthquakes in order to provide the necessary feedback.

The causality of the earthquakes induced by gas production is likely to be through the interaction of compaction of the reservoir rock with existing faults and differentiated geology of the subsurface layers. The ground subsidence occurs because with the extraction of gas, pressure support decreases in the layer from which the gas is extracted. The weight of overlying layers then compacts that extraction layer until a new pressure equilibrium can be established, cf. Dake (1978); Doornhof et al. (2006). The technical addendum to the winningsplan Groningen 2013 "Subsidence, Induced Earthquakes and Seismic Hazard Analysis in the Groningen Field" (Nederlandse Aardolie Maatschappij BV, 2013) discusses all of these aspects in the context of the Groningen reservoir in much more detail.

The seismic network of the KNMI has been in operation for a few decades, and detailed reporting on and (complete) data for earthquakes in the Groningen region are available from 1991 onwards. The locations of all earthquakes in the region are shown in fig. 2.1, together with the locations of the gas production clusters. Also indicated are the boundaries of the regions for which the earthquake rates are determined in this report, which are the same as in the previous semi-annual updates (cf. Pijpers (2014, 2015a,b, 2016a,b); Pijpers and van Straalen (2017))

In this technical report, a Weibull distribution function is fitted to the available earthquake data are examined for a signature of changes in rates. Contrary to the periodic reports, for the present analysis the individual time intervals are used between successive events, satisfying selection criteria detailed below. For this analysis the requirements for completeness of the catalogue are therefore more stringent. It is for this reason that the selection criteria include a lower limit for the earthquake magnitude that is slightly higher than is usual for the other CBS reports: a lower limit $M \geq 1.2$ is used throughout this report. Furthermore, only earthquakes having occurred on or after Jan. 1^{st} 2003 are taken into account.

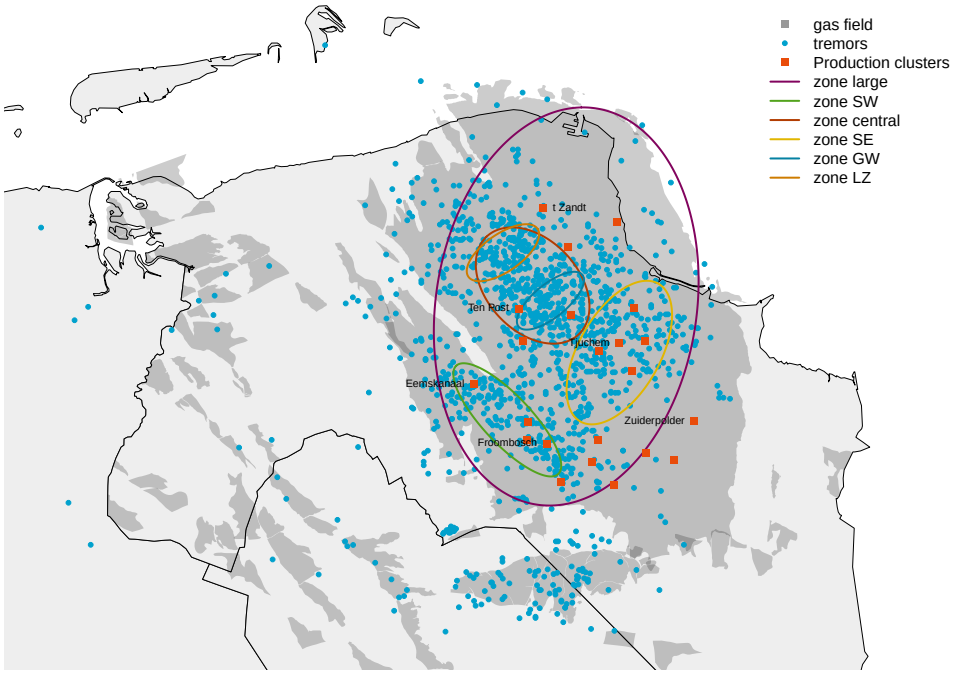
2 Background

2.1 The earthquake data

The available earthquake dataset contains in total 1502 events recorded after 1 Jan. 1991 up to 1 May 2018. Of these, there are 431 that have magnitudes $M \geq 1.2$, occurred from Jan. 1 20003

onwards, and are located within the zone indicated as `zone large' in fig. 2.1. An earthquake magnitude and time of event as well as the KNMI's present best estimate of the longitude-latitude position is available for each of the earthquakes. The KNMI has indicated that the network of seismometers was designed in the '90s to be complete in terms of both detection and localisation of earthquakes in the Groningen region above magnitudes of 1.5. Above magnitude 1.2 the detection likelihood is $> 99\%$, but the localisation may sometimes be more problematic. The elliptical contour of the localisation uncertainty progressively increases in size towards lower magnitudes and also is not uniform in orientation or size for different subregions, depending on the distances to the nearest seismic stations.

Figure 2.1 The locations of earthquakes as reported by the KNMI. The red squares are locations of the production clusters, several of which are identified by name. The purple ellipse `zone large' demarks the reference area for earthquake rates. The red and green smaller ellipses (central and SW respectively) mark the two regions of interest also reported on in previous reports. The yellow ellipse (SE) is an additional region first considered in the report of Nov 2015. The production field is also shown in dark gray, overplotted on a map of the region



Starting towards the end of 2014 an upgrade to the network of seismometers has been implemented which has pushed down these limiting magnitudes for completeness and localisation. With the selections used, this should have minimal impact on the present analysis

2.2 Weibull distributions

The catalog of quake events is likely also to contain events that are aftershocks. This means that some fraction of events has not occurred completely independently from preceding ones, which implies that it is inappropriate to assume Poissonian statistics. One possible way to approach this problem is to use an algorithm to identify which entries in the catalog are likely to be aftershocks rather than being caused directly by the compaction, and remove these shocks before carrying through any analysis. Unfortunately such algorithms always carry a small risk that they remove some shocks from the analysis which should not have been removed, and also that they retain

others that should have been removed. If even the direct process is in fact not purely Poissonian, this further complicates such aftershock removal.

For this reason it has been proposed by Post (2017) that instead of attempting to, possibly artificially, distinguish direct shocks and aftershocks, all shocks should be regarded as originating from a single statistical process that takes into account that the events are not fully independent. One distribution function that is widely used for such purposes is the two-parameter Weibull distribution function $f_W(\Delta)$ for the time intervals Δ between events; earthquakes in the case at hand:

$$f_W(\Delta) = \frac{\beta}{\tau}\left(\frac{\Delta}{\tau}\right)^{\beta-1} \exp\left(-\left(\frac{\Delta}{\tau}\right)^{\beta}\right) \quad \forall \Delta > 0 \tag{1}$$

The parameter τ is a typical timescale for the process: the expectation value $\overline{\Delta}$ for the time interval, given a finite observation time T is proportional to it:

$$\overline{\Delta} = \int_0^T \Delta f_W(\Delta) d\Delta = \tau\gamma\,(1 + 1/\beta,\ x)\ /\ [1 - \exp(-x)] \tag{2}$$

in which

$$x \equiv \left(\frac{T}{\tau}\right)^{\beta} \tag{3}$$

and where $\gamma(a, x)$ is the incomplete gamma function. In the limit that $T \to \infty$, this function becomes equal to $\Gamma(a)$. The parameter β controls the likelihood of events clustering together in time. The special case $\beta = 1$ is the exponential distribution associated with uncorrelated (Poissonian) event statistics. For $\beta < 1$ the likelihood of very short time intervals is enhanced compared to this and the distribution function decreases monotonically with Δ. For $\beta > 1$ the distribution function has a single maximum at :

$$\Delta_{max} = \tau\,(1 - 1/\beta)^{1/\beta} \quad \beta > 1 \tag{4}$$

This Weibull distribution function is used widely in engineering applications to model the failure of built structures and equipment, where failure of a particular component can lead to a cascade of failures of other components, ultimately leading to a breakdown of the entire structure or equipment. Therefore it appears a reasonable analog to a situation where the displacement of stresses along a fracture after an earthquake triggers further earthquakes.

Generalisations of this two-parameter Weibull function are also discussed in eg. Post (2017) where the power law factors in Δ in eq. (1) are replaced by a more general function h:

$$f_h(\Delta) = h(\Delta; x) \exp\left(-\int_0^{\Delta} h(u; x) du\right) \quad \forall\ \Delta > 0 \tag{5}$$

The function h needs to be positive semi-definite but has few other restrictions. A form such as this allows for instance accounting for dependencies on parameters x that may have spatial variation across the field. Such generalisations are explored in Post (2017) and are of interest for further analyses, but are not pursued in this report.

2.3 'graphical method' of Weibull parameter estimation

One widely used approach to fit a Weibull distribution of the type (1) to data is a method usually referred to as the graphical method. The starting point of this method is to use the cumulative

distribution function F_W:

$$F_W(\Delta) \equiv 1 - \exp\left(-\left(\frac{\Delta}{\tau}\right)^{\beta}\right) \tag{6}$$

which, after rearranging yields:

$$\ln\left(-\ln\left(1 - F_W(\Delta_i)\right)\right) \equiv \beta \ln(\Delta_i) - \beta \ln(\tau) \tag{7}$$

By sorting the measured intervals Δ_i in increasing order and approximating the cumulative distribution by $F_W(\Delta_i) \approx i/(N+1) \equiv Z_i$ where N is the number of measured intervals, it can be seen that eq. (7) is a linear relationship between variables X_i and Y_i:

$$\begin{aligned} X_i &\equiv \ln(\Delta_i) \\ Y_i &\equiv \ln\left(-\ln\left(1 - Z_i\right)\right) \end{aligned} \tag{8}$$

and therefore the parameters can be determined using a standard least squares method. Furthermore, plotting the data after this transformation may reveal very quickly whether the linear fit is at all an acceptable model.

To quantify the quality of the fit in terms of a χ^2-measure, and to determine error bars for the parameters, it is necessary to establish error bars for the data. The errors in the variable $\ln(\Delta_i)$ are equivalent to the relative error in the length of the time intervals. The timing with which earthquakes are reported in the catalogue is at sub-second precision. This implies that the 1σ error in this variable X is $< 10^{-6}$ for all measured intervals, except the very few that are less than one hour. The two shortest intervals in the dataset are between earthquakes near Lageland on 30/05/2010 (21 seconds) and near Appingedam on 09/08/2014 of 34 seconds. After that, the next two intervals are resp. one and two orders of magnitude larger than this. This source of error is therefore negligible compared to the error associated with the approximation that $F_W(\Delta_i) \approx i/(N+1)$.

A formal derivation of the error estimate for Y_i is beyond the scope of this paper. However, a heuristic approach provides an error estimate with an accuracy that is sufficient for the present purpose. The first step is straightforward:

$$\sigma_Y^2 = \left(\frac{\partial Y}{\partial Z}\right)^2 \sigma_Z^2 = \left(\frac{1}{(1-Z)\ln(1-Z)}\right)^2 \sigma_Z^2 \tag{9}$$

There are two contributions to consider in determining $\sigma^2(Z)$. The first contribution is that any value for Z in the range $\left[\frac{i-1/2}{N+1}, \frac{i+1/2}{N+1}\right]$ would approximate the cumulative distribution function. Under the assumption of a uniform likelihood within this range would yield a contribution $\sigma_{Z\,1}^2 = 1/\left(12 * (N+1)^2\right)$. However this does not yet take into account that a collective shift of anywhere from 0 to N of the ordering values Z would also be acceptable. A mean value per point that takes this into account produces:

$$\sigma_{Z\,1}^2 = \frac{N}{(24(N+1)^2)} \tag{10}$$

The second contribution to consider is an issue of completeness per region. The selection criterion of $M \geq 1.2$ is chosen to ensure an overall completeness of better than 99% but even a relatively small uncertainty spatially may mean that a particular earthquake is detected but located inside a region when it should have been located outside, or vice versa. This means that there is an additional contribution due to the changes in Z when a Δ_j needs to be added or

removed at position j in the ordered list:

$$\sigma_{Z\,2}^2 = \frac{1}{N}\left\{\sum_{i=1}^{j-1}\left[\frac{i}{N+1} - \frac{i}{N+2}\right]^2 + \sum_{j+1}^{N+1}\left[\frac{i}{N+1} - \frac{i+1}{N+2}\right]^2\right\} \tag{11}$$

Evaluating the summations and rearranging, and substituting the mean $\bar{j} = (N+1)/2$ for j yields:

$$\sigma_{Z\,2}^2 = \frac{N^2-1}{12(N+1)^2(N+2)^2} \tag{12}$$

For the elliptical regions considered here, and assuming a typical positional uncertainty radius of $\epsilon_r \approx 0.5$ km, this contribution needs to be modified by a factor:

$$\frac{\pi(a+\frac{1}{2}\epsilon_r)(b+\frac{1}{2}\epsilon_r) - \pi(a-\frac{1}{2}\epsilon_r)(b-\frac{1}{2}\epsilon_r)}{\pi ab}\phi\frac{N}{2} = \left(\frac{1}{a}+\frac{1}{b}\right)\epsilon_r\phi\frac{N}{2} \equiv \zeta\frac{N}{2} \tag{13}$$

In which a and b are the lengths in km of the semimajor and semiminor axes of the elliptical regions and ϕ is a filling factor ≤ 1 along the border of each of the ellipses. The modifier expresses the number of earthquakes that are expected to be involved given the area of the border region compared to the area of the interior of the ellipse. The factor ϕ expresses that for some of the ellipses the borders as they are set happen to go through regions with a slightly lower spatial density of earthquakes than the average over the entire field. For ellipses designated SW and large, this factor is set to 0.5 and 0.25 respectively, for all other ellipses this factor is set to 1. The total variance per point due to these effects is then:

$$\sigma_Z^2 = \frac{N}{24(N+1)^2}\left[1 + \zeta\frac{N^2-1}{(N+2)^2}\right] \tag{14}$$

This value in combination with eq. (9) is used to set the variances for the Y and also the weights in the weighted least squares method to determine the fit. This also provides all the information necessary to determine normalised χ^2/ν values for the fits and the error estimates for β and τ. The number of degrees of freedom $\nu = N - 3$ where N is the number of selected earthquakes per region and epoch.

2.4 maximum likelihood estimation of Weibull parameters

The starting point of a maximum likelihood (ML) method is the joint probability distribution function for the set of event intervals. The joint likelihood is given by the product of the distribution function (1) for every interval, where the parameters are presumed the same for the selected set of event intervals. The standard procedure is to maximise the logarithm of this joint likelihood for the parameters. It is equivalent to maximise the likelihood or its logarithm, and it is more convenient to maximise the logarithm, since all products reduce to summations. The log-likelihood $\ln(L)$ satisfies:

$$\ln(L) = N\left[\ln\left(\frac{\beta}{\tau}\right) - (\beta-1)\ln(\tau)\right] + (\beta-1)\sum_i^N \ln(\Delta_i) - \sum_i^N\left(\frac{\Delta_i}{\tau}\right)^\beta \tag{15}$$

The maximum is obtained by setting the derivatives with respect to the parameters τ and β to 0 and solving:

$$\begin{aligned}\frac{\partial}{\partial\tau}\ln(L) &= \frac{\beta}{\tau}\left[-N + \sum_i\left(\frac{\Delta_i}{\tau}\right)^\beta\right] = 0 \\ \frac{\partial}{\partial\beta}\ln(L) &= \frac{N}{\beta} - N\ln\tau - \left(\frac{1}{\tau}\right)^\beta\sum_i\Delta_i^\beta\ln\Delta_i + \left(\frac{1}{\tau}\right)^\beta\ln\tau\sum_i\Delta_i^\beta + \sum_i\ln\Delta_i = 0\end{aligned} \tag{16}$$

substituting the first equation of (16) into the second equation and rearranging yields:

$$\begin{aligned} \ln\tau &= \frac{1}{\beta}\ln\left[\frac{1}{N}\sum_i \Delta_i^\beta\right] \\ \frac{1}{\beta} &= \sum_i \left(w_i(\beta) - \frac{1}{N}\right)\ln\Delta_i \end{aligned} \tag{17}$$

in which $w_i(\beta)$ is defined as:

$$w_i(\beta) \equiv \frac{\Delta_i^\beta}{\sum_j \Delta_j^\beta} \tag{18}$$

The second of equations (17) is a non-linear equation in β. This can be solved using the usual Newton-Raphson approach. An alternative is iterating on the second equation of (17) by computing the $w_i(\beta)$ with the given β and substituting to provide an updated value. This method also converges, but slightly slower than the Newton-Raphson method. Using as starting values the values obtained in section 2.3 yields updates < 0.005 in β after between 6 and 16 iterations using the latter method, whereas an equal or better precision is obtained in 7 Newton-Raphson steps.

It is known that the ML estimation for β is biased. There are expressions for bias correction (cf. Ross (1996)) for β which in the present case are:

$$\beta_{corr} = \beta\frac{1}{1 + \frac{1.37}{n-2.92}\sqrt{\frac{n}{n-1}}} \tag{19}$$

The margins of uncertainty around the ML estimates for the parameters are obtained from the inverse of the Fisher information matrix, which contains the variances and covariances of the parameters as its elements:

$$\mathcal{I}^{-1} = \begin{pmatrix} \sigma^2(\tau) & cov(\beta,\tau) \\ cov(\beta,\tau) & \sigma^2(\beta) \end{pmatrix} \tag{20}$$

This Fisher information matrix $\mathcal{I}$ is:

$$\begin{aligned} \mathcal{I} &\equiv -\begin{pmatrix} \frac{\partial^2 \ln(L)}{\partial\tau^2} & \frac{\partial^2 \ln(L)}{\partial\beta\partial\tau} \\ \frac{\partial^2 \ln(L)}{\partial\beta\partial\tau} & \frac{\partial^2 \ln(L)}{\partial\beta^2} \end{pmatrix} \\ &= N\begin{pmatrix} \frac{\beta^2}{\tau^2} & -\frac{\beta}{\tau}\left[\frac{1}{\beta} - \ln\tau + \overline{\ln\Delta_i}\right] \\ -\frac{\beta}{\tau}\left[\frac{1}{\beta} - \ln\tau + \overline{\ln\Delta_i}\right] & \left[\frac{1}{\beta} - \ln\tau\right]^2 - 2\ln\tau\,\overline{\ln\Delta_i} + \sum_i w_i(\beta)\,(\ln\Delta_i)^2 \end{pmatrix} \end{aligned} \tag{21}$$

in which $\overline{\ln\Delta_i} \equiv \frac{1}{N}\sum_i \ln\Delta_i$. Wherever the parameters τ and β occur in the expression for $\mathcal{I}$, the ML estimates are to be substituted.

3 Fitting results for Weibull distributions

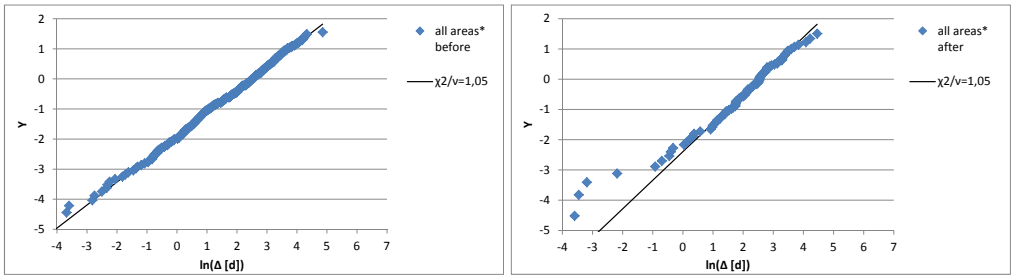
The method described in section 2.3 is used to determine weighted least squares fits for the Weibull distribution for each zone and epoch, and the results are listed in table 3.1. For the central region the cutoff date between the two epochs is set to 01/01/2014 and the period

Table 3.1 Fitted parameters using the 'graphical method' described in section 2.3

region	epoch	N $(= \nu + 3)$	χ^2/ν	β	σ_β	τ [d]	σ_τ [d]
zone large	I	341	1.05	0.7656	0.00011	11.92	0.040
	III	92	1.05	0.9454	0.00097	12.6	0.15
zone SW	I	47	0.55	0.609	0.0019	73	1.7
	III	13	0.42	0.58	0.011	97	7.1
zone central	I	124	1.29	0.8105	0.00064	31.9	0.31
	II	41	0,62	0.738	0.0029	35.1	0.94
zone SE	I	43	1.24	0.752	0.0028	70	1.8
	III	17	1.33	1.25	0.018	74	4.5
GW	I	39	2,08	0.594	0.0031	91.	2.9
	II	11	1.90	0.281	0.0095	109.	12.
LZ	I	40	2.71	0.680	0.0034	99.	3.0
	II	14	0.67	0.57	0.011	107.	7.9

before is referred to as epoch I, and after is referred to as epoch II. For the other regions the cutoff date is set to 01/01/2015 and before and after this date are referred to epochs I and III. The reason for this distinction in cutoff dates is that in the central region already in 2014 a major reduction in production was implemented for one of the clusters. In 2015 more widespread modifications in production strategy were implemented.

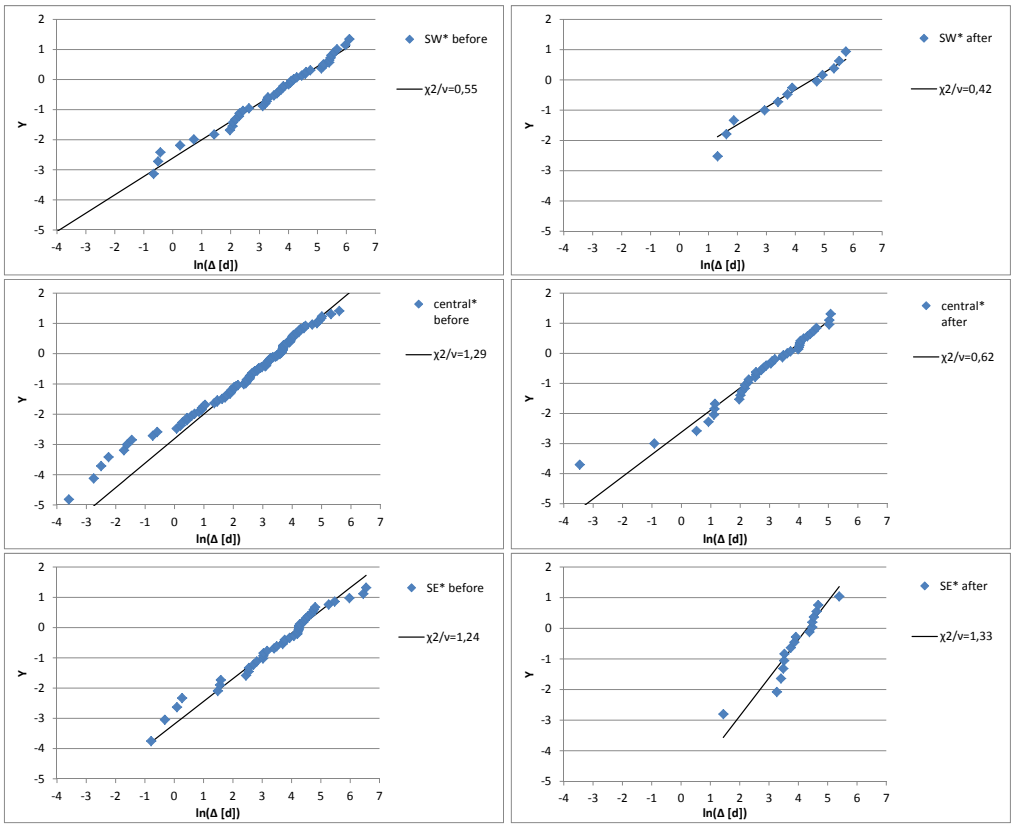
Figure 3.1 The Weibull plot, whole zone 'large'. The data as well as the fitted form for the Weibull distribution is shown, marked by the $\chi 2/\nu$ value, for the two epochs: epoch I on the left, epoch III on the right.



From table 3.1 it is clear that in all cases the parameter β is different from 1 : the difference between the result and 1 is statistically significant at a confidence level well in excess of 99% in all cases. The interval statistics are clearly not consistent with a Poisson process. For the main 4 zones the values of χ^2/ν are all close to 1. This implies that the quality of the fits in all cases is acceptable. For the two smaller zones GW and LZ (see Fig. 2.1) in epoch I the value of χ^2/ν is high implying a poor fit. From Fig. 3.3 it is clear that the shortest time intervals are the cause since they deviate from the fit lines. Earthquakes occurring within about a day and within about 1 km from each other might represent a population of events with separate statistical properties.

A comparison of the parameters between the various sub-regions and the region large, in each case for the corresponding epoch, also shows that the two parameters β and τ differ between them. These also are results that are statistically significant at a confidence level well in excess of 99%. Clearly there is spatial inhomogeneity of the process, a conclusion also drawn by Post (2017). The Weibull plot for the region central, in particular over epoch I (middle row, lefthand panel of fig. 3.2), shows that the shortest intervals, all less than ~ 8 hours, appear to displaced from the trend of the other points in the plot. The majority of the corresponding earthquake events are clustered together in a small subregion inside that zone central. For this reason, these

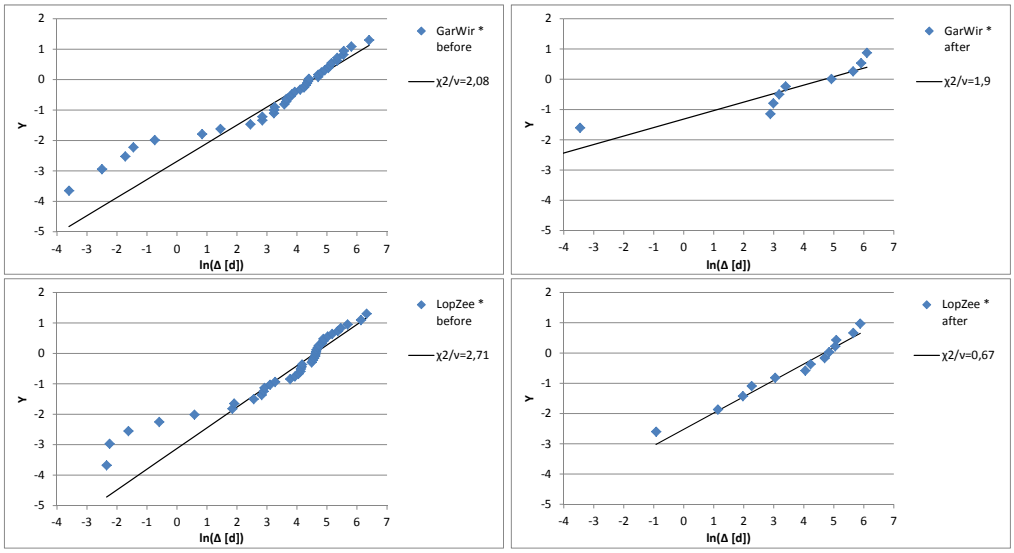
Figure 3.2 The Weibull plots, for each of the elliptical subregions in the two appropriate epochs. From top to bottom are shown zone SW, central and SE. The lefthand plots are for the earlier epoch I, the later epoch (II or III) is shown on the right. Note that for zone central in epoch I in particular the events with shorter time intervals appear to be displaced from the fit. The error bars for these points are large, but they nevertheless adversely affect the χ^2 unless they are removed altogether.



9 points are given a weight 0 when carrying out the fitting for zone central as a whole, and a separate Weibull fit is done for all earthquakes that lie inside one of two smaller ellipses (zone 'GW' and 'LZ' in Fig. 2.1) containing those events: GW is a region near the small towns of Garrelsweer and Wirdum, LZ is a region near Loppersum and Zeerijp. Figure 3.3 shows the Weibull plots for these regions, for the two epochs, and the final rows in table 3.1 shows the corresponding parameters. Clearly in this subregion the parameter β is significantly smaller than overall in zone central.

The difference in parameters for each region between the epoch before and after the cutoff date are also examined for significance. For the region SW the differences both for τ and for β are statistically significant at the 99% level. For all other regions the differences between epochs for β are statistically significant at confidence levels well in excess of the 99% level. For the regions SW and central β has decreased. For the region SE β has increased from a value below 1 to a value above 1. For the region large the value of β has also increased, becoming closer to 1. The parameter τ has increased for all regions between the earlier and later epoch. In zones SW, central and large this result is statistically significant at a confidence level of 99%. In region SE and the smaller regions GW and LZ the difference is not statistically significant. This increase in τ means that the time between earthquakes has on average increased, so that the intensity or frequency of earthquakes has decreased. The decreases in β, where they occur, do imply that triggering of secondary events has increased in some regions: producing a larger excess of short intervals. Overall over the region large, this does not appear to be the case, since both β and τ

Figure 3.3 The Weibull plots for the zones of interest near Garrelsweer & Wirdum (top row) and near Loppersum & Zeerijp (bottom row) for the epochs I (left column) and II (right column)



have increased.

The intensity of earthquakes cannot be compared directly between the regions, because larger regions will have more events and therefore lower mean time intervals between events. This area effect must be accounted for, in order to be able to compare. To this end the ratios of the inverse time scales are scaled by the area ratios, where the zone large in epoch I is used as reference. In accordance with the regular periodic reporting of CBS, the earthquake intensities in the zones of interest are enhanced compared to the zone 'large', but the ratios R_1 with respect to that zone are smaller in the more recent epochs compared to epoch I, where:

$$R_1 \equiv \frac{\tau_{large,\,I} A_{zone}}{\tau_{zone,\,epoch} A_{large,\,I}} \tag{22}$$

$$R_2 \equiv \frac{\overline{\Delta}_{large,\,I} A_{zone}}{\overline{\Delta}_{zone,\,epoch}\, A_{large,\,I}} \tag{23}$$

In addition to the intensity ratios, the ratios of quake number expectation values per unit area

Table 3.2 Events 'intensity ratios': the ratios R_1 and R_2 for all regions and epochs.

region	epoch	R_1	R_2
zone large	I	$\equiv 1$	$\equiv 1$
	III	0.95	1.08
zone SW	I	2.40	1.91
	III	1.81	1.48
zone central	I	3.45	3.50
	II	2.92	2.90
zone SE	I	1.38	1.35
	III	1.31	1.64
GW	I	3.86	4.68
	II	3.22	3.13
LZ	I	3.57	4.15
	II	3.27	3.52

and time R_2 are of interest as well. From table 3.2 it can be seen that this ratio and R_1 have similar behaviours for zones SW and central and also the zones GW and LZ inside central:

enhanced compared to zone 'large' in epoch I but less so in the more recent epoch. However the zone large R_2 value shows an increase rather than the decrease seen for R_1, and so does zone SE. For zone SE this difference is due to the substantial change in the shape parameter of the Weibull distribution, which only for this zone and epoch is > 1. Given the low number of degrees of freedom and the lower quality of the fit (fig. 3.2 bottom right panel), it is unclear whether this result by itself is meaningful. However, since also for the entire region both β and R_2 increase, this may imply that displacing production away from the regions central and SW is being followed by decreases in earthquake rates in these regions but at the same time by increases in earthquake rates outside of these regions. The higher β implies that they are less 'clustered together' in time so that the timescale parameter τ increases.

The ML method of determining the parameters β and τ, using a Newton-Raphson scheme to determine β, produces the estimates shown in table 3.3. After correction for bias using eq. (19) the parameters β as determined by the ML method differ somewhat from the values as shown in table 3.1. In terms of absolute deviation the differences are between about 0.014 and 0.092, except for zone SW in epoch III where the deviation is as large as 0.15. Since the determination of τ depends on β there are therefore also deviations between the ML values of τ and the values determined through the graphical method.

Table 3.3 Weibull parameters β and τ as determined by maximum likelihood estimation. Also given is the bias-corrected β_{corr} and the standard deviations for β and τ

region	epoch	β	β_{corr}	$\sigma(\beta)$	τ [d]	$\sigma(\tau)$ [d]
zone large	I	0.783	0.780	0.034	11.3	0.8
	III	0.890	0.877	0.073	12.9	1.6
zone SW	I	0.644	0.624	0.077	76.	18.
	III	0.809	0.70	0.19	177.	64.
zone central	I	0.672	0.664	0.043	33.3	4.7
	II	0.814	0.79	0.10	38.7	7.7
zone SE	I	0.715	0.690	0.083	84.	19.
	III	1.425	1.29	0.27	118.	21.
GW	I	0.653	0.628	0.089	90.	23.
	II*	0.386	0.32	0.11	146.	120.
LZ	I	0.724	0.697	0.068	101.	16.
	II	0.721	0.63	0.46	165.	84.

* no NR covergence; bisection used

An issue with the ML method is that all data are given equal weight in the determination of the parameters. This, combined with the weaker dependence of $\ln(L)$ on the values of τ and β, has a tendency to increase the sizes of confidence intervals for the parameters. Using the inverse of the Fisher information matrix, eq. (21), the confidence intervals for the ML determination of the parameters can be determined, which are also listed in table 3.3. Given these sizes of the confidence intervals, only one of the differences in the determinations of the parameters between the graphical method and the ML method are statistically significant. For the region central in epoch I there is a difference in β, where in the ML determination the 9 deviating points are included that are omitted in the graphical method .

4 Conclusions

From the analysis presented in this report, it can be concluded that, averaged over the period from Jan. 1 2003 to May 1 2018, there is a statistically significant spatial enhancement of the earthquake rate in all three zones, SW, central and SE, where gas production takes place, compared to the surrounding region, by factors that correspond to the factors reported previously by CBS; see (Pijpers and van Straalen, 2018) for the most recent update. Since the reduction in production levels, first at central clusters, and later also at other clusters, the present analysis also shows the earthquake rates (intensities) have reduced by modest amounts: in zones SW by a bit more than 20%, and in zone central by around 10%. Overall in zone large and in zone SE it depends on the particular measure used whether there is an increase or decrease. In two particular areas of interest, near Garrelsweer & Wirdum and near Loppersum & Zeerijp respectively, that are part of zone central, which in the past have had several 'pairs' of quakes occurring in quick succession, there is also a reduction which for the zone near Garrelsweer and Wirdum appears to be a slightly larger reduction than overall in zone central.

The present analysis differs from the periodic reporting of CBS in that the time intervals between earthquakes are fitted using a Weibull distribution function. This two-parameter family of distribution functions has as a special case an exponential distribution, which would occur if the generating process for the earthquakes were Poissonian. From the values of the parameters that are deduced in the fitting procedure it is clear that the process does not conform to a Poissonian process. This implies as well that a separation of all earthquakes in the catalogue into direct shocks and aftershocks may not be appropriate.

On the other hand, even if the statistics of the time intervals between earthquakes suggest that they are not completely independent events at any length of time interval between successive quakes, nevertheless the main conclusions of the CBS periodic reporting are corroborated in the present analysis: the reductions in production levels have been followed quite quickly by reductions in earthquake rates, i.e. the number of quakes occurring per unit time and per unit area.

References

Dake, L. (1978). *fundamentals of reservoir engineering*. Elsevier. (17^{th} impr. 1998, chapter 3).

Doornhof, D., T. Kristiansen, N. Nagel, P. Pattillo, and C. Sayers (2006). Compaction and Subsidence. Technical report, Schlumberger.

Nederlandse Aardolie Maatschappij BV (2013). A technical addendum to the winningsplan Groningen 2013 subsidence, induced earthquakes and seismic hazard analysis in the Groningen field. Technical report.

Pijpers, F. (2014). Phase 0 report 2 : significance of trend changes in tremor rates in Groningen. Technical report, Statistics Netherlands.

Pijpers, F. (2015a). Phase 1 update may 2015 : significance of trend changes in tremor rates in Groningen. Technical report, Statistics Netherlands.

Pijpers, F. (2015b). Trend changes in tremor rates in Groningen : update november 2015. Technical report, Statistics Netherlands.

Pijpers, F. (2016a). Trend changes in tremor rates in Groningen : update may 2016. Technical report, Statistics Netherlands.

Pijpers, F. (2016b). Trend changes in tremor rates in Groningen : update nov 2016. Technical report, Statistics Netherlands.

Pijpers, F. and V. van Straalen (2017). Trend changes in tremor rates in Groningen : update june 2017. Technical report, Statistics Netherlands.

Pijpers, F. and V. van Straalen (2018). Trend changes in tremor rates in Groningen : May 2018. Technical report, Statistics Netherlands.

Post, R. (2017). Statistical inference for induced seismicity in the Groningen gas field (MSc. thesis). Technical report, TU Eindhoven.

Ross, R. (1996). Bias and standard deviation due to Weibull parameter estimation for small data sets. *IEEE Trans. on Dielectrics and Electrical Insulation 3, 1*, 28--42.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information