NAM

Statistical methodology for investigating seasonal variation in rates of earthquake occurrence in the Groningen field

Shell Research

S. Bierman, R. Paleja and M. Jones

Date October 2018

Editors Jan van Elk & Dirk Doornhof

General Introduction

The seismological model (Version 5) currently used in the assessment of hazard and risk for the induced seismicity in Groningen, provides a probabilistic prediction of the seismicity dependent on the local reservoir pressure depletion associated with the gas volume produced. The seismicity is in this model not dependent on the gas production rate. The gas volume extracted determines reservoir pressure depletion, which governs the expected number and magnitude of induced earthquakes. Within the model, the expected number of events depends on the pressure depletion, but not the rate of that depletion. Theoretically, there are processes which potentially could cause the expected event number, for a given incremental volume of gas production to depend on the rate of that gas production. These could be associated with the geomechanical behaviour of faults (e.g. rate and state frictional fault behaviour) or compaction (e.g. a-seismic stress relaxation at production time scales).

However, studies carried out as part of the research program of NAM have not been able to identify whether these processes play a significant role or been able to quantify the impact of gas production rate on seismicity. In an environment of decreasing and more stable gas production rates, ignoring potential production rate dependency of the seismicity will be conservative and lead to a potential over-estimation of hazard and risk.

Given the current state of knowledge, NAM is not in a position to increase the sensitivity of the seismological model to production rate changes as this was so far found to degrade the performance of the model and accepts that as a result the assessment of hazard and risk might be conservative. The current model yields a sensitivity to seasonal depletion rate changes that is thought to be close to the upper bound of sensitivities consistent with the observed catalogue. On the other hand, based on the research to date, seasonal seismicity variations within the catalogue are lower than the detection threshold.

In the operation of the field, NAM will make every effort to reduce fluctuations in gas production. The Minister of Economic Affairs and Climate Policy has, on the advice of the regulator SodM, imposed limits to the production fluctuations. NAM will report on any excursions from these set limits.

In recent years, NAM has carried out several studies into the dependency of the induced seismicity in Groningen on the gas production rate from the field. This included studies into reservoir behaviour (Ref. 1), modelling of the various mechanisms that could induce production rate dependency (Ref. 3 and 3) and analysis of field data using machine-learning (Ref. 5) and statistical techniques (Ref. 2, and 4).

This document contains two reports describing the results of statistical studies to test the hypothesis that there is a seasonal trend in the event rate for earthquakes with the annual cycle in production rates in the Groningen field.

References

1. Geurtsen, L., P. Valvatne and A. Mar-Or, Optimisation of the Production Distribution over the Groningen field to reduce Seismicity, NAM, December 2017
2. Bourne, Stephen and Steve Oates, The influence of stress rates on induced seismicity rates within the Groningen field, Shell Research, August 2018.
3. DeDontney, Nora, and Suvrat Lele, Impact of Production Fluctuations on Groningen Seismicity – Part 1, Geomechanical Modelling using Rate of State friction, ExxonMobil Upstream Research Company, 2018.
4. Burch D. and B. Symington, Impact of Production Fluctuations on Groningen Seismicity – Part 2, Data Analytics, ExxonMobil Upstream Research Company, 2018.
5. Park T., H. Jamali-Rad, W. Oosterbosch, J. Limbeck, F. Lanz, C. Harris, E. Barbaro, K. Bisdom & K. Nevenzeel, Seasonality analysis for induced seismicity event rate time series within the Groningen Field, Shell Research and IBM, August 2018.
6. Bourne, S.J., Oates, S.J., 2017. Extreme threshold failures within a heterogeneous elastic thin-sheet and the spatial-temporal development of induced seismicity within the Groningen gas _eld. Journal of Geophysical Research: Solid Earth 122, 10,299{10,320.
7. Bourne, S.J., Oates, S.J., Elk, J.V., 2018. The exponential rise of induced seismicity with increasing stress levels in the Groningen gasfield and its implications for controlling seismic risk. Geophysical Journal International 213, 1693-1700.

NAM

Title	**The influence of stress rates on induced seismicity rates within the Groningen gas field - Seismological Model**		**Date**	October 2018
			Initiator	NAM
Autor(s)	**Stephen Bourne and Steve Oates**	**Editors**	Jan van Elk and Dirk Doornhof	
Organisation	Shell Research	**Organisation**	NAM	
Place in the Study and Data Acquisition Plan	Study Theme: Impact Production Fluctuations Comment: The seismological model (Version 5) currently used in the assessment of hazard and risk for the induced seismicity in Groningen, provides a probabilistic prediction of the seismicity dependent on the local reservoir pressure depletion associated with the gas volume produced. The seismicity is in this model not dependent on the gas production rate. The gas volume extracted determines reservoir pressure depletion which governs the expected number and magnitude of induced earthquakes. Within the model, the expected number of events depends on the pressure depletion but not the rate of that depletion. Theoretically, there are processes which potentially could cause the expected event number for a given incremental volume of gas production to depend on the rate of that gas production. These could be associated with the geomechanical behaviour of fault (e.g. rate and state frictional fault behaviour) or compaction (e.g. a-seismic stress relaxation at production time scales). However, studies carried out as part of the research program of NAM have not been able to identify whether these processes play a significant role or been able to quantify the impact of gas production rate on seismicity. In an environment of decreasing and more stable gas production rates, ignoring potential production rate dependency of the seismicity will be conservative and lead to a potential over-estimation of hazard and risk. Given the current state of knowledge, NAM is not in a position to increase the sensitivity of the seismological model to production rate changes as this was so far found to degrade the performance of the model and accepts that as a result to assessment of hazard and risk might be conservative. The current model yields a sensitivity to seasonal depletion rate changes that is thought to be close to the upper bound of sensitivities consistent with the observed catalogue. On the other hand, the effect is, based on the research to date, seasonal seismicity variations within the catalogue are less than the detection threshold. In the operation of the field, NAM will make every effort to reduce fluctuations in gas production. The Minister of Economic Affairs and Climate Policy has, on the advice of the			

	regulator SodM, imposed limits to the production fluctuations. NAM will report on any excursions from these set limits. In recent years, NAM has carried out several studies into the dependency of the induced seismicity in Groningen on the gas production rate from the field. This included studies into reservoir behaviour, modelling of the various mechanisms that could induce production rate dependency and analysis of field data using machine-learning and statistical techniques. This document contains two reports describing the results of statistical studies to test the hypothesis that there is a seasonal trend in the event rate for earthquakes with the annual cycle in production rates in the Groningen field.
Directliy linked research	(1) Gas Production (2) Reservoir Modelling (3) Geomechanical Modelling (4) Seismological Model
Used data	KNMI Earthquake catalogue Groningen gas production data
Associated organisation	NAM
Assurance	

Statistical methodology for investigating seasonal variation in rates of earthquake occurrence in the Groningen field

Restricted
SR.15.13132
October 2015

Restricted **SR.15.13132**

Statistical methodology for investigating seasonal variation in rates of earthquake occurrence in the Groningen field

by

S. Bierman (GSNL PTD/TASE)
R. Paleja (GSUK PTD/TASE)
M. Jones (GSUK PTD/TASE)

Shell Global Solutions International B.V., Amsterdam

Further electronic copies can be obtained from the Global Information Centre.

Executive Summary

In this report, we describe and apply statistical methodology to test for evidence of within-year variation (seasonality) in rates of occurrence of earthquake events as recorded in the Earthquake Catalogue (EC), and for evidence of a relationship between seasonal (monthly) variation in gas production and earthquake rates. For this statistical investigation we have used the data as recorded in the EC by the Royal Dutch Meteorological Institute (KNMI). The EC that was used in this study was complete up to and including the event recorded on the 23^{rd} of January 2015 near the town of Sappemeer. We pay special attention to possible differences in apparent seasonality of rates of events within different ranges of event magnitudes. With the current network of geophones, inclusion probabilities of events with associated magnitudes M<1.5 may vary both spatially and temporally. The probability that an earthquake with associated magnitude M<1.5, when it occurs within the Groningen field, is detected by the geophone network and included in the EC (the so-called "inclusion probability") cannot be assumed to be unity nor to be spatio-temporally invariant. In this report we have assumed, in line with recent advice from the Dutch Meteorological Society (KNMI), that inclusion probabilities for events with magnitudes $M \geq 1.5$ can be assumed to be 1 or close to 1 throughout the Groningen field and throughout the time-series under consideration. We therefore use the terminology "Earthquake rate" or "Earthquake count" for observed counts or estimated rates of events with magnitudes $M \geq 1.5$. We use the terminology "Catalogue Rate" or "Catalogue Count" for observed counts or estimated rates of events with magnitudes M<1.5.

Our main findings are:

- Strong evidence was found of within-year seasonal variation in rates of Catalogue Events with magnitudes $M \leq 1$. The within-year variation of Catalogue rates could be correlated with within-year variation in gas production with a delay of approximately 3 to 4 calendar months such that the peak in within-year monthly gas production (typically December, January or February) preceeded relatively high rates of Catalogue Events (typically in the months of April or May). We note that it is unclear what the underlying cause(s) of this seasonal trend are, and the interpretation of this finding is made difficult in particular because of the possibility that this is partly or wholly caused by variability in probabilities that earthquakes with small event magnitudes are detected and recorded in the catalogue. Further research is required to investigate this, and more insight may be gained if in future a denser network of geophones is in operation.
- Some evidence was found for seasonal variation in catalogue rates with magnitudes $1 < M < 1.5$. Some evidence was found that relatively high monthly Catalogue rates of events with associated magnitudes $1 < M < 1.5$ on average lagged the annual peak in monthly gas production rates by approximately 3-4 calendar months.
- No statistically meaningful evidence could be found of seasonality in earthquake rates of events with $M \geq 1.5$. Although there are some indications that numbers of earthquakes with $M \geq 1.5$ are unevenly distributed over the calendar months, this may have occurred by chance. It should also be noted that the distribution of events with $M \geq 1.5$ appears to be different from that of the events with magnitudes $M < 1.5$. The highest rates of events with $M \geq 1.5$ were in the months of January and February whereas the month with the lowest rate is December; this pattern is difficult to explain with the seasonal fluctuations in production data.

- We note that this study does not provide any evidence of a causal relationship between variation in gas production and rates of earthquakes.
- We note that in this report only temporal variation in event rates was investigated. No other covariates were considered, in particular spatial variation in event rates and production rates was not taken into account.
- The statistical methodology for testing for evidence of temporal trends in earthquakes needs to be further developed. We note that in the current study not all avenues for testing for a possible relationship between seasonal variation in gas production and catalogue rates or earthquake rates have been explored.

Amsterdam, October 2015.

Table of Contents

List of Figures

List of Tables

1. Introduction

In this report, we describe and apply statistical methodology to test for evidence of within-year variation (seasonality) in rates of earthquake events associated with the Groningen gas field, and for evidence of a relationship between seasonal (monthly) variation in gas production and event rates. For this statistical investigation we have used the data as recorded in the Earthquake Catalogue (EC) of events by the Royal Dutch Meteorological Institute (KNMI). The data were obtained from the KNMI web-pages at the following web-address: http://www.knmi.nl/seismologie/geinduceerde-bevingen-nl. The EC that was used in this study was complete up to and including the event recorded on the 23^{rd} of January 2015 near the town of Sappemeer. We pay special attention to possible differences in apparent seasonality of rates of events within different ranges of event magnitudes. With the current network of geophones, detection probabilities of events with magnitudes M<1.5 may vary both spatially and temporally. The probability that an earthquake with associated magnitude M<1.5, when it occurs within the Groningen field, is picked up by the geophone network and included in the EC cannot be assumed to be one or to be spatio-temporally invariant. In this report we have assumed, in line with recent advice from the Dutch Meteorological Society (KNMI), that inclusion probabilities for events with magnitudes $M \geq 1.5$ can be assumed to be 1 or close to 1 throughout the Groningen field and throughout the time-series under consideration (Dost et al. [2012]). We therefore use the terminology “Earthquake Rate” or “Earthquake Count” for observed counts or estimated rates of events with magnitudes $M \geq 1.5$. We use the terminology “Catalogue Rate” or “Catalogue Count” for outcomes of analyses whenever events with magnitudes M<1.5 have been used. The terminology "Catalogue Rate" or "Catalogue Count" is therefore also used when events of all magnitudes are analyzed together.

In this report, we:

- Provide visualisations of the data that provide insight into the data and apparent presence or absence of seasonality of event rates.
- Describe and apply statistical methodology that can be used to test for evidence of seasonality in the event rates, and quantification of the time-lag (and uncertainty thereof) which gives the optimal correlation between gas production and event rates.
- Investigate to what extent the apparent seasonal fluctuation in Catalogue rates are caused by fluctuations in rates of events below and above M=1.5.

We analyse trends in earthquake rates for a number of subsets of the EC:

- All events in the EC. The interpretation of the results of these analyses are complicated due to the uneven inclusion probabilities, and due to the fact that there is no lower bound on the magnitudes of events.
- Events in the EC with associated magnitudes $M \geq 1.5$. In the interpretation of these results it can be assumed that inclusion probabilities are spatio-temporally invariant.
- Events in the EC with associated magnitudes $1 < M \leq 1.5$. The interpretation of the results of these analyses are complicated due to the uneven inclusion probabilities.
- Events in the EC with associated magnitudes $M \leq 1$. The interpretation of the results of these analyses are greatly complicated due to the uneven inclusion probabilities, in particular due to the fact that there is no lower bound on the magnitudes of events.

2. Earthquake catalogue

The KNMI catalogue of induced earthquakes contains events for the whole of The Netherlands. In our analyses, we have used events only within the outline of the Groningen reservoir with an additional spatial buffer of 1000 m, as depicted in figure 2.1.

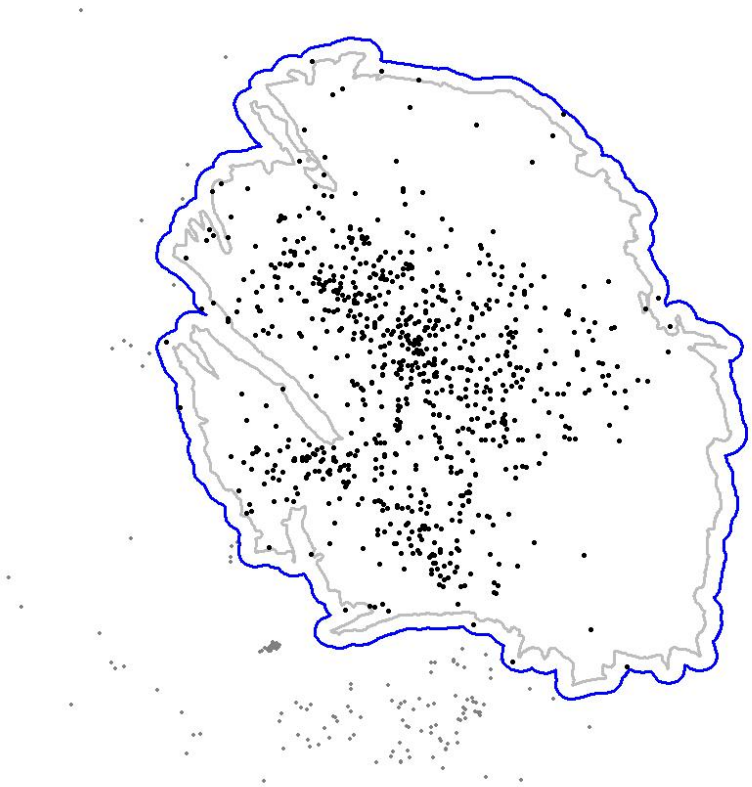

Figure 2.1.: Map of the outline of the Groningen reservoir (inner grey line) and an additional buffer with a width of 1000 meters (outer blue line).

The EC contains 795 events with epicenters inside the boundaries of the Groningen reservoir plus a buffer of 1000 m, and 281 events outside of these bounds but in the vicinity of the field (figure 2.1). Earthquakes in the Groningen field are believed to partly occur in clusters in time and space, in the form of aftershocks. In this report we perform analyses on the raw data including all counts as well as on a subset of the data (referred to hereafter as declustered) in which we have excluded events that occurred within 3 days and 2500 m of a previous event. Of all events inside the field boundary a total of 67 events (8.4%) were classed as potential aftershocks (table 2.1). Of all events outside of the field boundary a total of 63 (22.4%) events were classed as potential aftershocks (table 2.2). Thus, events outside of the field boundary occurred relatively often within a

relatively short distance and time-interval of a previous event, and any analysis of these data will be affected much by the choice of definition of aftershocks. We note that this particular choice of definition of potential aftershock is arbitrary. We attempt to, for all presented analyses, discuss to what extent potential aftershocks may have influenced the conclusions. We also refer to chapter 4 where methodology is discussed which may be used to assess the influence of potential overdispersion on inferences drawn from analyses of count data.

Table 2.1.: Numbers of earthquakes as recorded in the KNMI catalogue with epicenters inside the Groningen field boundary plus a 1000 m buffer (figure 2.1) which occurred within a certain time-interval and a certain distance of an earlier event.

	$\leq$ 100 m	$\leq$ 500 m	$\leq$ 1000 m	$\leq$ 2500 m	$\leq$ 5000 m	Total
$\leq$ 1 hour	3	6	9	16	19	53
$\leq$ 4 hours	3	7	13	26	33	82
$\leq$ 12 hours	3	8	20	36	49	117
$\leq$ 1 day	3	11	26	52	74	166
$\leq$ 2 days	3	13	33	63	99	211
$\leq$ 3 days	3	13	34	67	119	236
$\leq$ 5 days	4	14	37	81	152	288
Total	22	73	172	341	545	

Table 2.2.: Numbers of earthquakes as recorded in the KNMI catalogue with epicenters outside of the Groningen field boundary plus a 1000 m buffer (figure 2.1) which occurred within a certain time-interval and a certain distance of an earlier event.

	$\leq$ 100 m	$\leq$ 500 m	$\leq$ 1000 m	$\leq$ 2500 m	$\leq$ 5000 m	Total
$\leq$ 1 hour	10	23	26	28	30	117
$\leq$ 4 hours	12	35	39	42	44	172
$\leq$ 12 hours	13	42	45	48	51	199
$\leq$ 1 day	20	47	50	53	56	226
$\leq$ 2 days	23	49	55	59	64	250
$\leq$ 3 days	24	51	58	63	68	264
$\leq$ 5 days	28	54	59	65	71	277
Total	130	301	332	358	384	

The locations of epicenters of events within different ranges of magnitudes are depicted in figure 2.2. There are no obvious differences in the spatial extent of the estimated epicenters of the events with different magnitudes.

A peculiar aspect of events with relatively low associated magnitudes is that the rate at which they occur in the catalogue appears to vary diurnally with higher rates of events between approximately 20:00 in the evening and 04:00 in the morning (figure 2.3). Such a diurnal pattern is not immediately apparent for events with magnitudes $M > 1$. This casts some doubt on the validity of observations of events with lower magnitudes, and warrants further investigation.

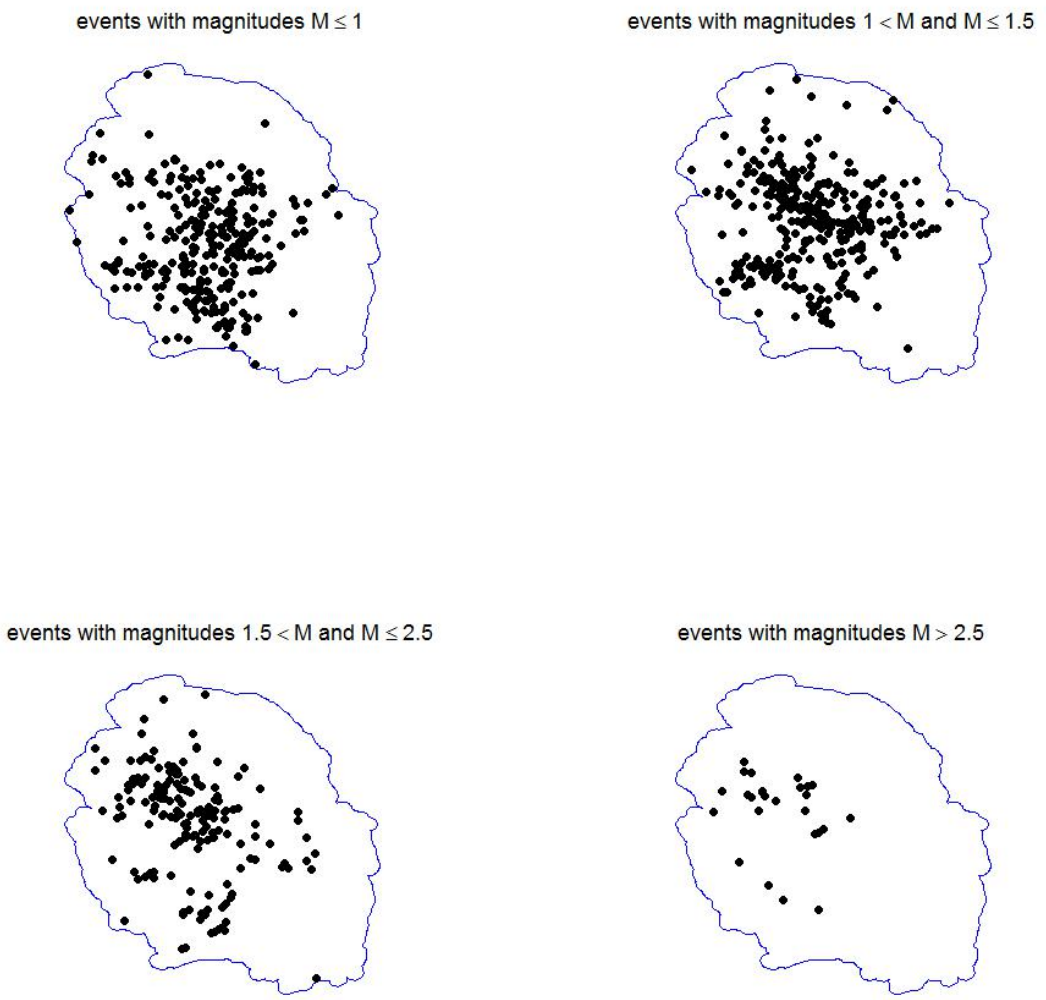

Figure 2.2.: Maps of epicenters of events in different of ranges of event magnitudes.

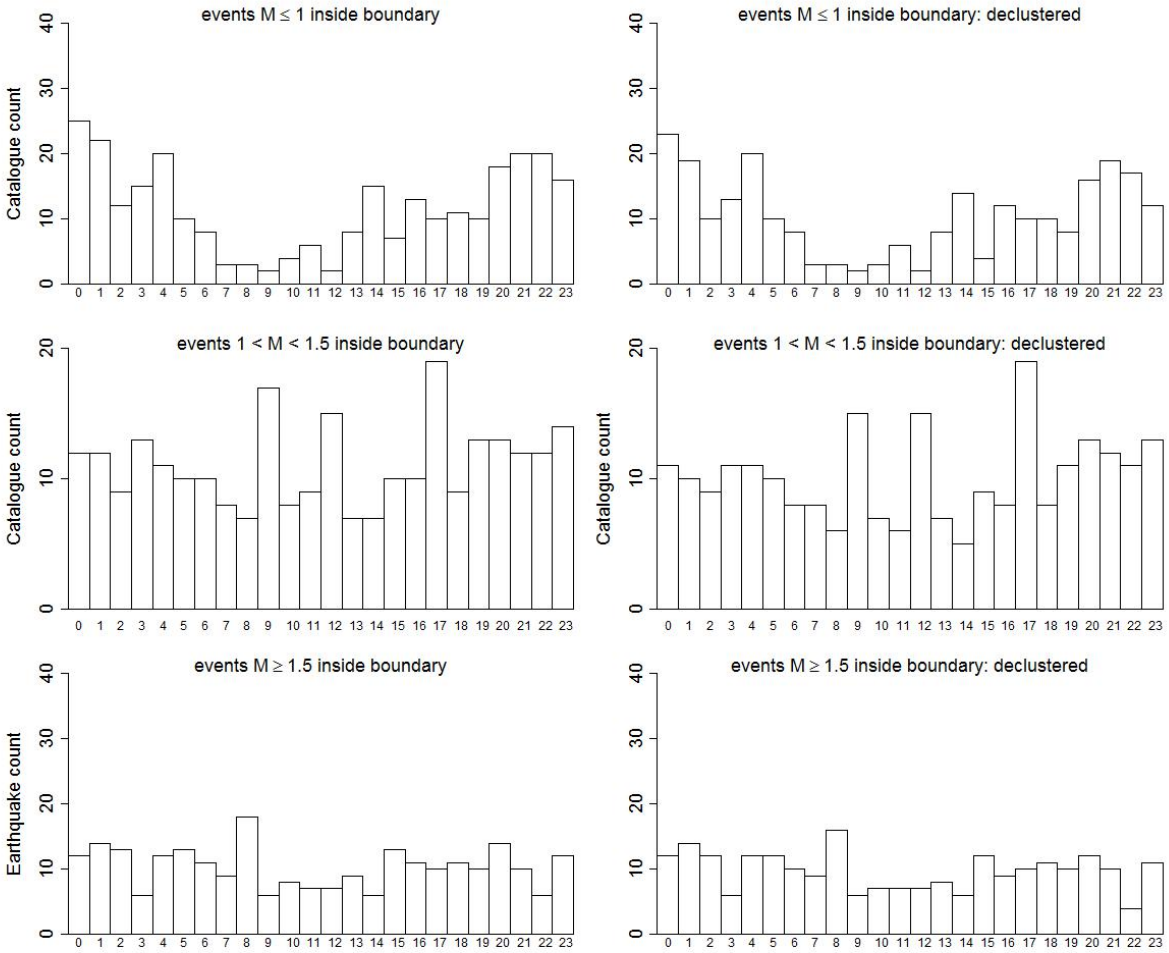

Figure 2.3.: Counts of numbers of events for each of the 24 hours within the day (00:00 - 01:00, 01:00 - 02:00,...,23:00 - 24:00) for events in different categories of associated magnitudes.

3. Seasonality in event rates: data visualisation

A visualisation of the catalogue of timings of Catalogue Events (of all magnitudes) associated with the Groningen gas field suggests that Catalogue event rates may vary seasonally and may, with some time-delay, be strongly correlated with the seasonal pattern in production rates (figure 3.1). The data visualisation is based on a moving average of counts of Catalogue Events, resulting in a temporally smooth trend in Catalogue rates. The temporally smooth trend in event rates is plotted alongside a time series of monthly gas production data (field wide). The moving average of counts of Catalogue Events is calculated using a "sliding time-window" approach, where each time-window spans three calendar months and the average of the counts of events in the three months is plotted on the graph. The time-windows are applied to each month in the time-series (incrementally).

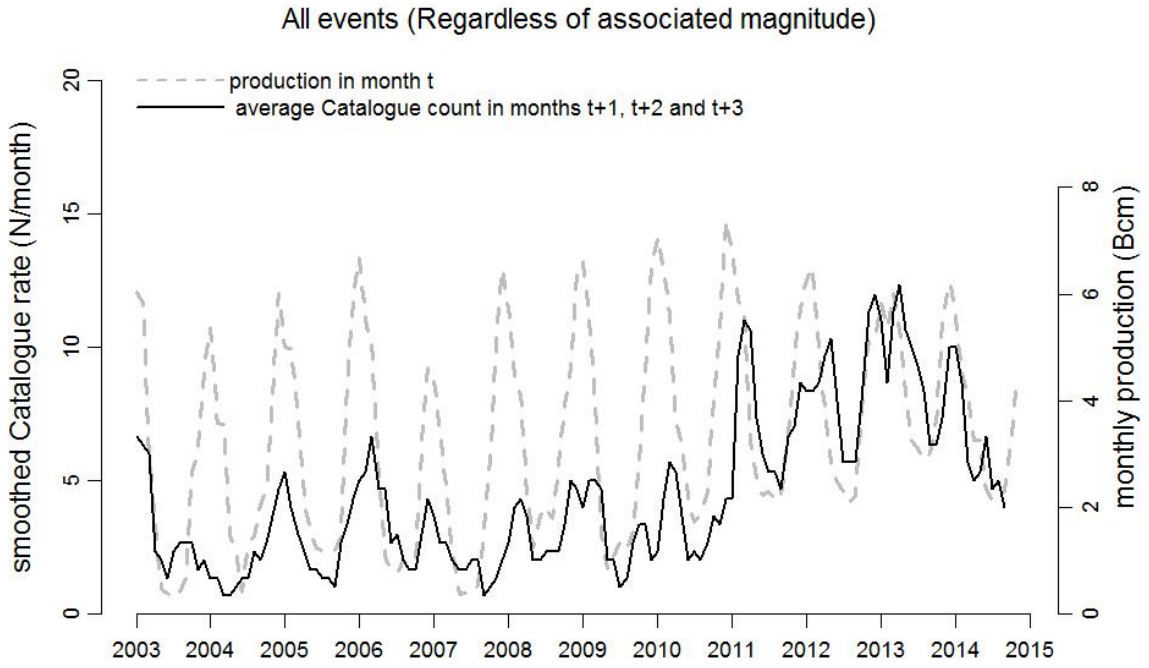

Figure 3.1.: Monthly field-wide gas production (grey dashed line) and smoothed Catalogue Events rates (black solid line). The smoothed Catalogue Event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months.

While figure 3.1 suggests that there is a correlation between gas production and Catalogue rates we note that any other variable which fluctuates seasonally within each year, such as ambient temperature, would also correlate with seasonally varying event rates. Furthermore, care is required with the interpretation of moving averages since each earthquake is used three times in the analysis (except for events in the first two and last two months in the time series).

Time series of counts of events per calendar month for all events, and events with magnitudes $M \leq 1$, $1 < M < 1.5$, and $M \geq 1.5$ are given in figure A.1. The same information is given in Ap-

pendix A in a set of four further graphs with a separate panel per calendar year (figures A.2, A.3 A.4 and A.5 for all events, events with $M \leq 1$, $1 < M < 1.5$ and $M \geq 1.5$ respectively). For events with magnitudes $M \leq 1$ inside the Groningen field boundary it appears that, for most calendar years, the highest Catalogue counts occurred in the first half of the year and were particularly high in April, May and June (figure 3.2). Counts of events with associated magnitudes $1 < M < 1.5$ were also highest in the months of April, May and June, but differences in counts between months were smaller (figure 3.3). Counts of events with associated magnitudes $M \geq 1.5$ appeared to be somewhat higher in the first six months of the year with most events occurring in January and February (figure 3.4).

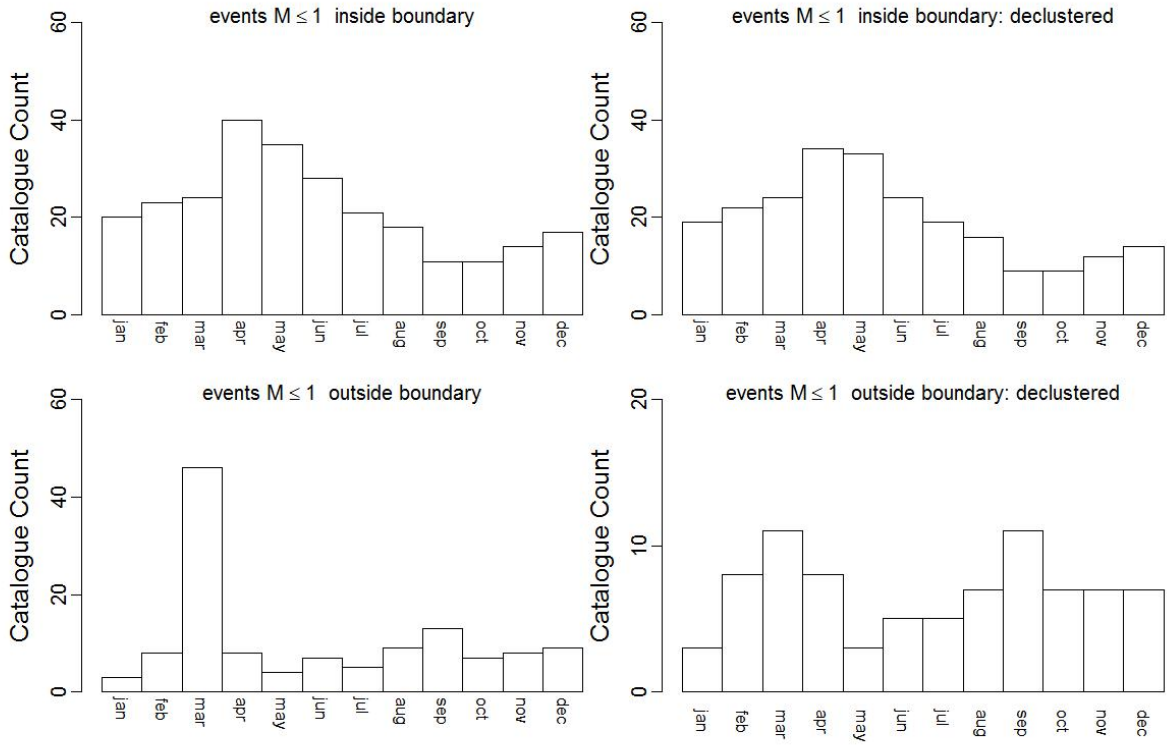

Figure 3.2.: Counts of events per calendar month, summed over all years, for events with $M \leq 1$.

Time series of catalogue counts of events per calendar month and monthly field-wide gas production are given in figure 3.5 for events $M \leq 1$, in figure 3.6 for events $1 < M < 1.5$ and in figure 3.7 for events $M \geq 1.5$. These figures provide a more direct representation of the available information than figure 3.1 because no temporal smoothing is used and each event occurs exactly once in the analysis. Visual inspection of these figures suggests that rates of events for all categories of magnitude may vary seasonally. If we apply the same smoothing as in figure 3.1, using a time-window of 3 calendar months, clear more-or-less regular seasonal fluctuations in Catalogue rates are visible for events $M \leq 1$ (figure 3.8). For events with magnitudes $1 < M < 1.5$ there is also an indication of regular seasonal fluctuations (figure 3.9), whereas there are no apparent regularly seasonal fluctuations for earthquakes with $M \geq 1.5$ (figure 3.10).

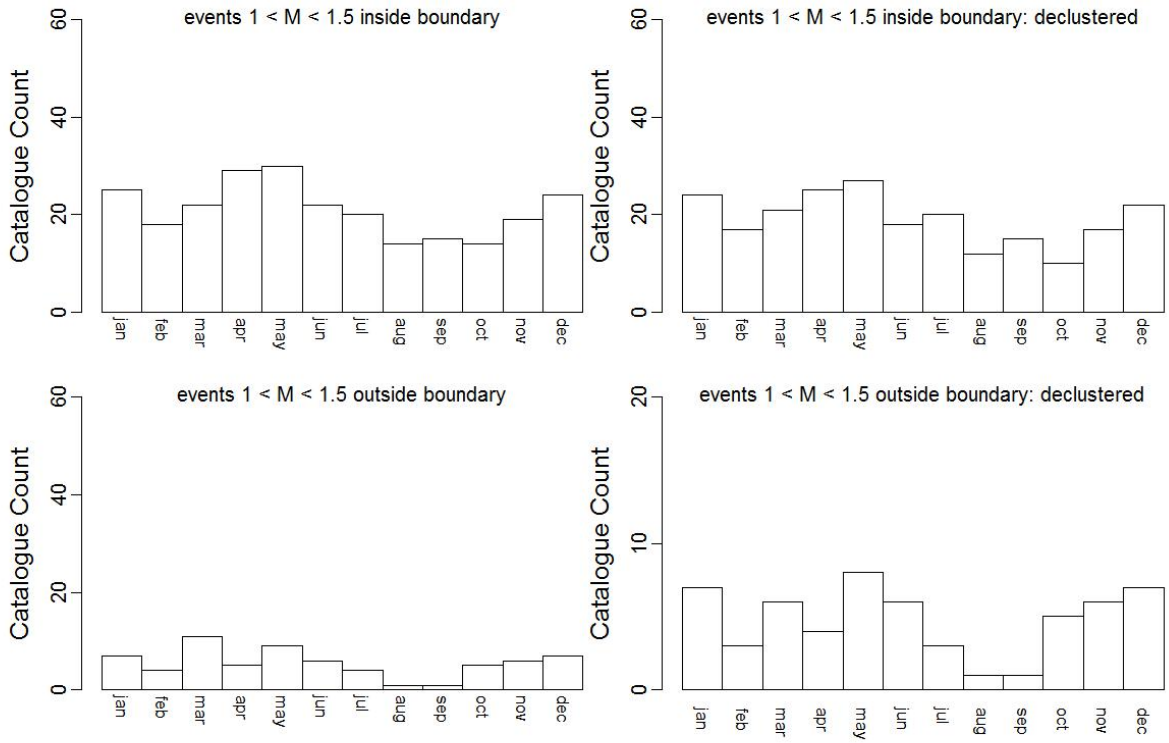

Figure 3.3.: Counts of events per calendar month, summed over all years, for events with $1 < M < 1.5$.

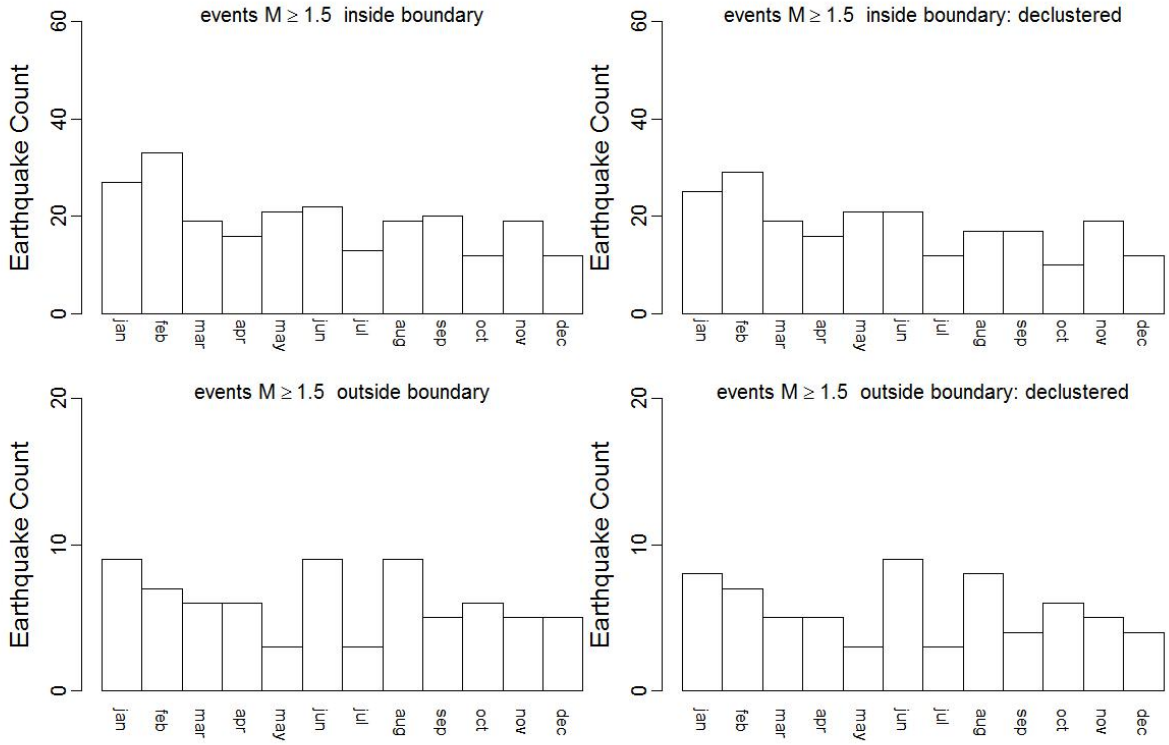

Figure 3.4.: Counts of events per calendar month, summed over all years, for events with M $\geq$ 1.5.

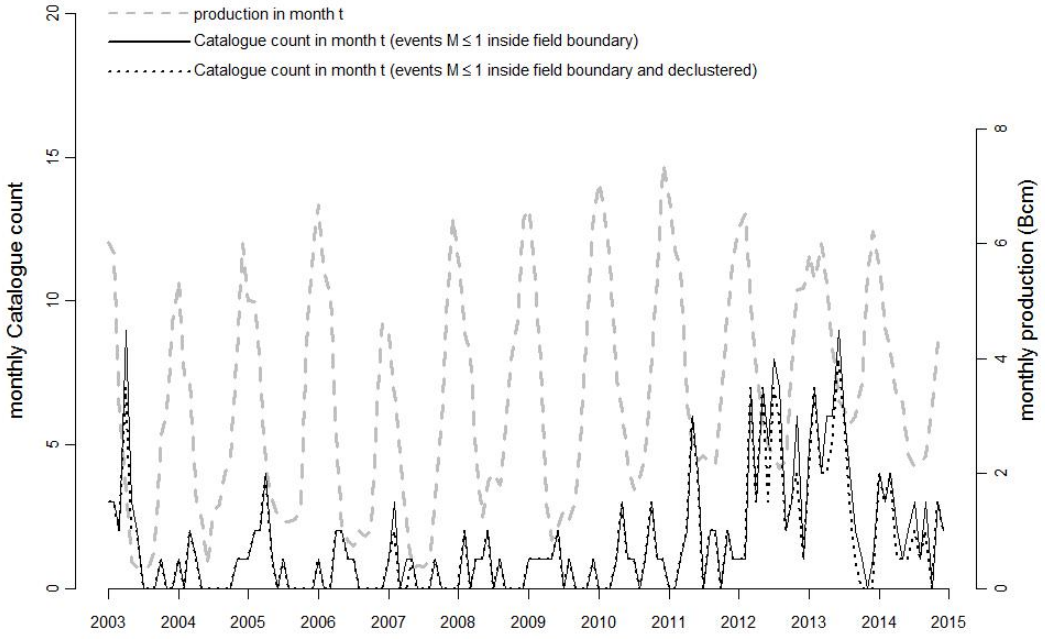

Figure 3.5.: Field wide monthly gas production and monthly catalogue counts of events inside the field boundary with associated magnitudes $M \leq 1$ (all events with $M \leq 1$ or with exclusion of events that occurred within 3 days and 2500 m of a previous event (declustered)).

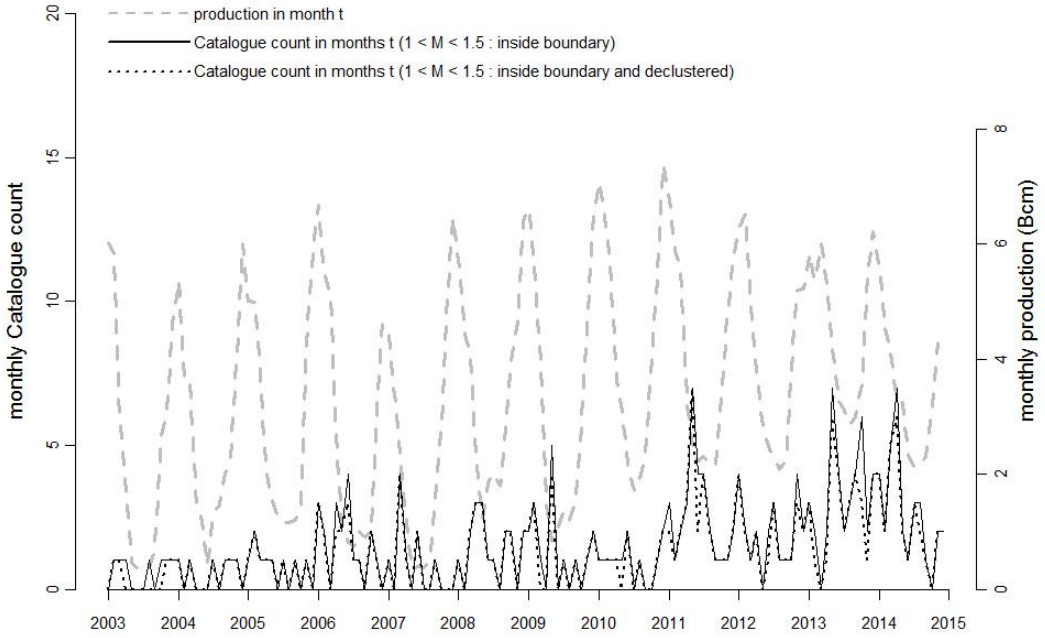

Figure 3.6.: Field wide monthly gas production and monthly counts of earthquakes inside the field boundary with associated magnitudes M $\geq$ 1.5 (all events with 1 $<$ M $<$ 1.5 or with exclusion of events that occurred within 3 days and 2500 m of a previous event (declustered)).

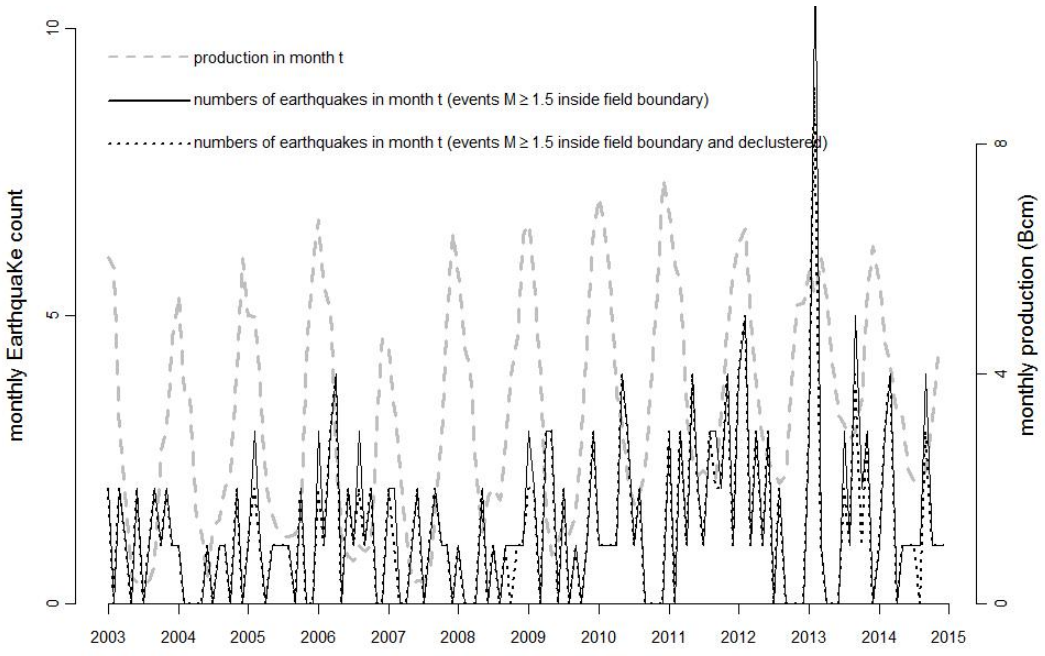

Figure 3.7.: Field wide monthly gas production and monthly counts of earthquakes inside the field boundary with associated magnitudes M $\geq$ 1.5 (all events with M $\geq$ 1.5 or with exclusion of events that occurred within 3 days and 2500 m of a previous event (declustered)).

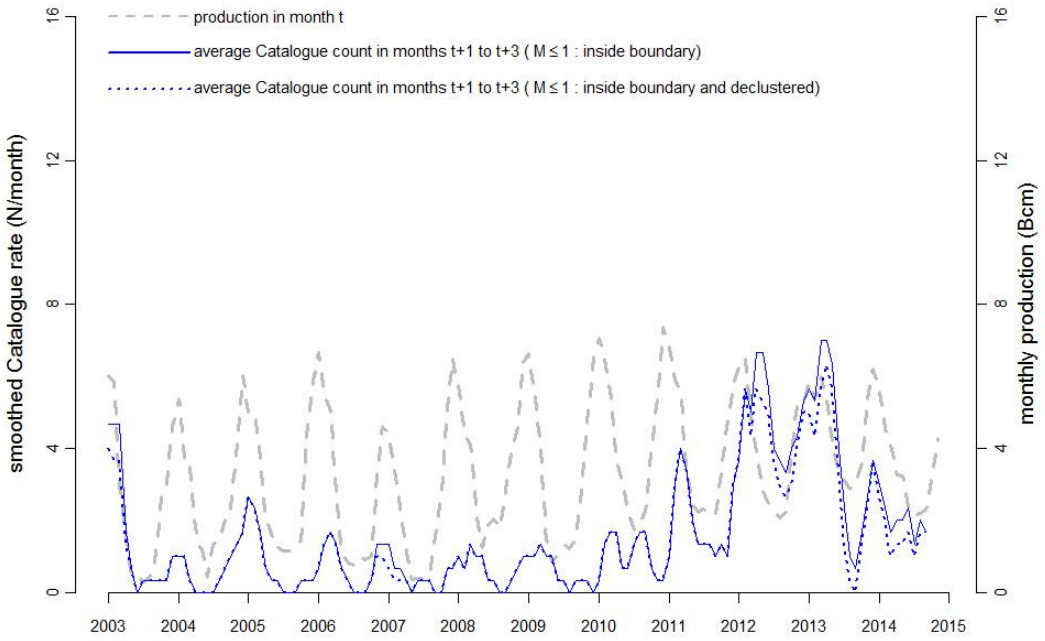

Figure 3.8.: Monthly field-wide gas production (grey dashed line) and smoothed catalogue rates for $M \leq 1$ (blue dotted line). The smoothed event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months and the averages of the counts of events in the three months are plotted on the graph and connected by lines.

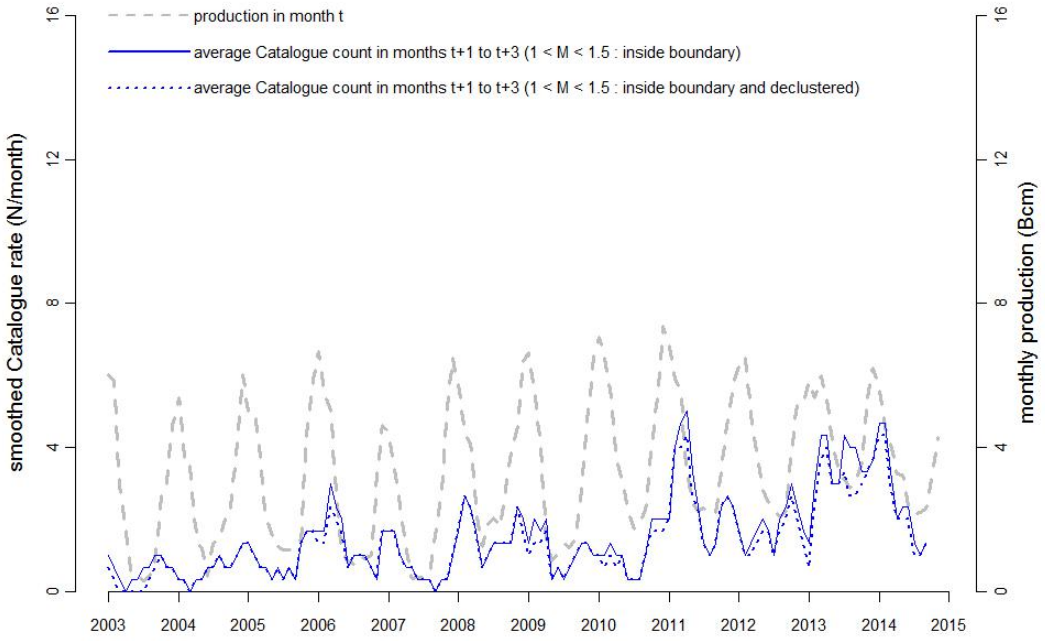

Figure 3.9.: Monthly field-wide gas production (grey dashed line) and smoothed catalogue rates for $1 < M < 1.5$ (blue dotted line). The smoothed event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months and the averages of the counts of events in the three months are plotted on the graph and connected by lines.

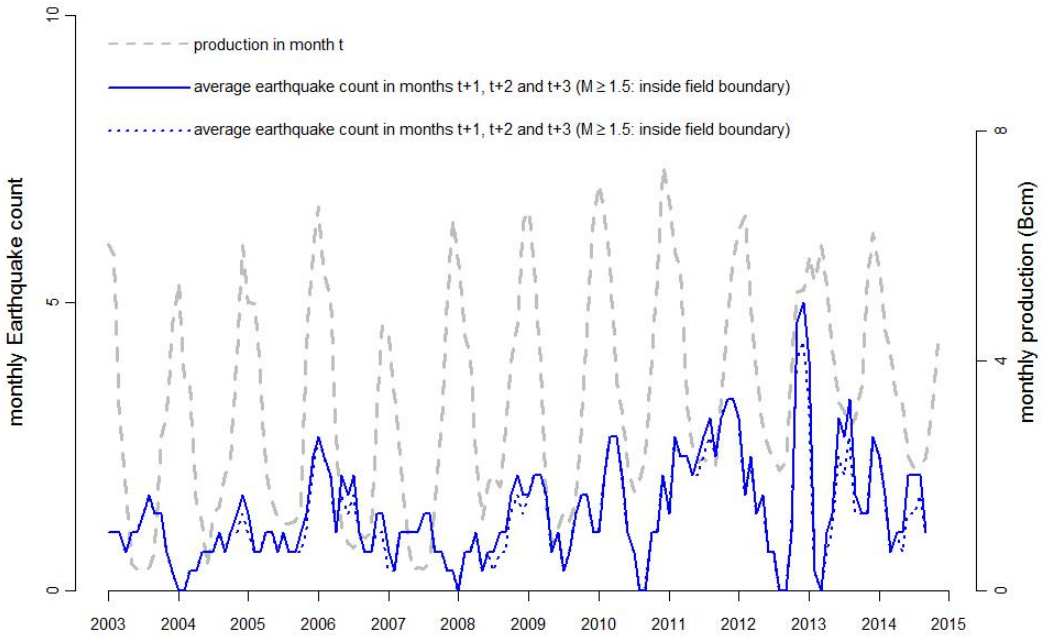

Figure 3.10.: Monthly field-wide gas production (grey dashed line) and smoothed catalogue rates for M $\geq$ 1.5 (blue dotted line). The smoothed event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months and the averages of the counts of events in the three months are plotted on the graph and connected by lines.

4. Seasonality in event rates: statistical analyses

We assume that the number of events N in a fixed time interval is Poisson distributed. This implies that we assume that events occur with some known average rate in this time interval, and that the probability of the occurrence of an event at any given time does not depend on the elapsed time since the last event. If random variate N is Poisson distributed with average rate λ ($\lambda > 0$), it takes integer values $n = 1, 2, ...$ with probability

$$\Pr\{N = n\} = \frac{e^{-\lambda}\lambda^n}{n!}. \tag{4.1}$$

The sum of independent Poisson random variables are also Poisson distributed. A practical consequence of this result is that a time series of event times can be summarised as a set of counts in disjoint time-intervals, and these counts can subsequently be analysed using a statistical model under the assumption that they are Poisson distributed.

Our main aim is to test whether there is evidence that event rates vary seasonally within a calendar year. We restrict the statistical analyses to counts of events that occurred in 2003 through to 2014, because most events occurred in this period and secondly because monthly production data was available for this period.

Let N_i be the number of events that occurred in "Julian month" i ($i = 1, 2, 3, ..., K$ with $K = 12 \times 12 = 144$ the total number of months from January 2003 up to an including December 2014). Let $\mathrm{y}(i)$ be the calendar year ($\mathrm{y}(i) \in \{2003, 2004, ..., 2014\}$) and $\mathrm{m}(i)$ the calendar month (within calendar year) of Julian month i ($\mathrm{m}(i) \in \{1, 2, ..., 12\}$).

All models for event rates are compared against a null model in which event rates in month i are modelled as an average event rate per calendar year $\alpha_{\mathrm{y}(i)}$:

$$\log_{\mathrm{e}}(\lambda_i) = \alpha_{\mathrm{y}(i)} \tag{4.2}$$

The log-linear Poisson model described above is also commonly referred to as a Generalized Linear Model (GLM) with Poisson error and log link, the latter ensuring positivity of λ_i. The parameters of the model can be estimated using iteratively reweighted least squares (McCullagh and Nelder [1989]). The GLM has been implemented using the R language for statistical computing (R Core Team [2014]).

The null model (4.2) is extended by estimating, within calendar year, deviations from the average year effect which are allowed to vary smoothly as a function of calendar month:

$$\log_{\mathrm{e}}(\lambda_i) = \alpha_{\mathrm{y}(i)} + \mathrm{s}(\mathrm{m}(i)) + \log_{\mathrm{e}}(\mathrm{O}_{\mathrm{m}(i)}) \tag{4.3}$$

where $\mathrm{O}_{\mathrm{m}(i)}$ is a correction factor ("offset") to correct for the differences in numbers of days per calendar month (within calendar year) $\mathrm{m}(i)$ (31 days for January, 28 for February, etc.), and $\mathrm{s}(\mathrm{m}(i))$ is a smooth function represented using penalized regression splines. The spline bases in this analysis are cyclic, which means that only smooth transitions are allowed between the month of December and January. The amount of smoothness is estimated using generalised cross-validation. This model is implemented using the functionality for generalised additive modeling (gam; see e.g. Hastie and Tibshirani [1990]) in R (Wood [2006], Wood [2011]). The potential flexibility of the

fitted curve for the seasonal response is governed partly by the basis dimension. The basis dimension (the number of knots at which the spline is allowed to 'flex') should be large enough to allow for enough flexibility to fit the potential shapes of the seasonal curves that may reasonably be expected. Here, we have chosen to fit the models using 12 knots, equal to the number of calendar months. The actual degrees of freedom of the models are estimated using cross-validation, and were in all cases much less than 12, which indicates that the number of knots was not restricting the shape of the seasonal curves that were fitted. In the above-mentioned implementation of gams in the R software, approximate multivariate posterior distributions of the fitted parameters are obtained by assuming a multivariate normal prior distribution and by using the derivatives of the likelihood with respect to the fitted parameters at convergence of the algorithm (see Wood [2000]).

The null model (4.2) is also extended by estimating, within calendar year, deviations from the average year effect as a log-linear function of average daily field-wide gas production per month i with some delay M ($M = 0, 1, 2, ..., 11$ calendar months), $P_{(\mathrm{m}(i)-M)}$:

$$\log_{\mathrm{e}}(\lambda_i) = \alpha_{\mathrm{y}(i)} + \beta P_{(\mathrm{m}(i)-M)} + \log_{\mathrm{e}}(\mathrm{O}_{\mathrm{m}(i)}) \tag{4.4}$$

where β is a year-invariant slope parameter (to be estimated) for the effect of monthly production on monthly event rate.

In the Poisson models it is assumed that the the variance in counts is equal to the expected count. In practice, it is commonly seen that the variability in counts is larger than the expectation, a phenomenon that is commonly referred to as overdispersion. Ignoring potential overdispersion is not likely to bias estimates of rate parameters (λ), but may lead to over-optimistic estimates of standard errors of these rate parameters. In an analysis of earthquake events, overdispersion may be expected to occur as a consequence of aftershocks. Here, we investigate the influence of the potential presence of overdispersion on our inferences in the following way:

- We perform analyses on the raw data including all counts as well as on a subset of the data (declustered) in which we have excluded events that occurred within 3 days and 2500 m of a previous event (see chapter 2).
- Apart from fitting the Poisson models as described above, we have also fitted an over-dispersed version of these models, commonly reffered to as "quasi-poisson" models, to allow for potential larger variation in counts than expected by the Poisson distribution. In the quasi-Poisson approach the variance is an unknown multiple of the mean, and this multiple,commonly known as the overdispersion parameter or scale parameter, is estimated from the data. Thus, in the quasi-poisson approach one extra parameter is estimated from the data. Estimates of rate parameters are identical between the standard Poisson and the quasi-Poisson models, but estimates of standard errors may be different depending on the presence of overdispersion.
- In situations where there is an indication of a seasonal effect, we further investigate the assumption that counts are Poisson distributed using standard statistical plots using the model residuals.

We propose (as future work) that more insight is obtained in the relationship between the mean (average rate) and variance of the rates. Other models models for the counts may be more applicable, such as for example the negative binomial which assumes a quadratic relationship between the mean and variance. The choice of mean-variance relationship may be influential because it influences the relative weights that are given to large and low counts.

To assess whether there is evidence of seasonality or a relationship with production we use a combination of the following:

- The estimated standard errors of the estimated deviations from the annual average rates (parameter(s) s(m(i)) in equation 4.3 and parameter β in equation 4.4) are used to test whether there is evidence that they are significantly different from zero.
- The estimated standard errors of the parameters and the Akaike Information Criterion (AIC) (see e.g. Burnham and Anderson [2004]) to compare the relative ability of the models to explain the data. The AIC is computed as minus twice the log-likelihood of a model plus twice the number of parameters of that model. If $\mathrm{AIC}_{\mathrm{null}}$ and $\mathrm{AIC}_{\mathrm{extended}}$ are the AIC for the null model (annual average rates only) and an extended model with within-year seasonal deviations in the rate respectively, then there is evidence of seasonal variations in rates only if $\mathrm{AIC}_{\mathrm{extended}} < \mathrm{AIC}_{\mathrm{null}}$, where the quantity $e^{((\mathrm{AIC}_{\mathrm{extended}}\mathrm{AIC}_{\mathrm{null}})/2)}$ is the relative likelihood of the extended model compared to the null model. Here, we use the convention that a difference in AIC (ΔAIC) of 2 units or more indicates that there is evidence that the extended model can explain the data better, whereas a difference of $\Delta\mathrm{AIC} \geq 10$ units or more indicates that there is strong evidence.
- For model 4.4, in addition to the estimated standard error of year-invariant parameter β, we estimate a slope β_y for each calendar year independently and construct a sampling distribution of the average over all slopes $\bar{\beta}$ by randomly resampling with replacement from the β_y.

We note that quasi-Poisson models do not have a fully specified likelihood and AIC values for this class of models can therefore not be calculated.

5. Results of statistical analyses

5.1. Smooth seasonal trend in event rates

The estimated smooth monthly deviations from the annual average event rates are depicted in figures 5.1 and 5.2 for the Poisson and quasi-Poisson models respectively. The differences between the two approaches (Poisson or quasi-Poisson) are small, and only apparent for events with $M \geq 1.5$ for which there is no evidence of seasonality given the large overlap in the estimated confidence intervals of the estimated monthly deviations. Also, the estimated smooth seasonal trends are little affected by the exclusion of events that occurred within 3 days and 2500 m of a previous event (declustering). This indicates that assumptions regarding over-dispersion are not highly influential in the analyses. For events with $M \leq 1$ there is good evidence that catalogue rates vary seasonally with higher catalogue rates in the first half of the year (approximately January - June) and lower rates in the last half of the year (approximately July - December), and highest rates around April and May (figures 5.3). For events with $1 < M < 1.5$ there is weak evidence of seasonality and the amplitude of this effect, if it is real, is likely to be small (figure 5.4). For events with $M \geq 1.5$ there is no evidence of seasonality (figure 5.5).

5.2. Relationship between monthly field-wide gas production and event rates

5.2.1. Events of all magnitudes

Evidence was found that monthly variation in production rates can explain some of the variation in the within-year differences in Catalogue rates of events of all magnitudes (table 5.1 for all events in the EC and table B.1 for all events after declustering). The optimal lag between gas production and catalogue rate was estimated to be around 3 months, although there is much uncertainty surrounding this timing and significant correlations could also be found for lags of 2,4 or 5 months. Because monthly production rates and event rates vary (approximately) periodically, an almost equally good correlation between production and event rates can be found with a lag of 9 months though parameter β swaps sign. A more detailed investigation into the modeling assumptions was done which indicated that the potential influence of overdispersion on the inferences is minor. For example, with a 3 month lag for all events the following estimates were obtained: $\beta = 3.79$ with an estimated standard error of 0.76 under the Poisson model and 0.87 under the quasi-Poisson model, and an estimated overdispersion parameter for the quasi-Poisson family of 1.31. These estimates change little if declustering is applied: $\beta = 3.86$ with an estimated standard error of 0.79 under the Poisson model and 0.86 under the quasi-Poisson model, and an estimated overdispersion parameter for the quasi-Poisson family of 1.2. The assumption that counts are Poisson distributed appears to be reasonable, based on a visual inspection of residuals of the model in a quantile-quantile plot (figure A.6).

5.2.2. Events with associated magnitudes $M \leq 1$

Evidence was found that monthly variation in production rates can explain some of the variation in the within-year differences in Catalogue rates of events with magnitudes $M \leq 1$ (figures 5.2 and B.2). The optimal lag between gas production and catalogue rate was estimated to be around

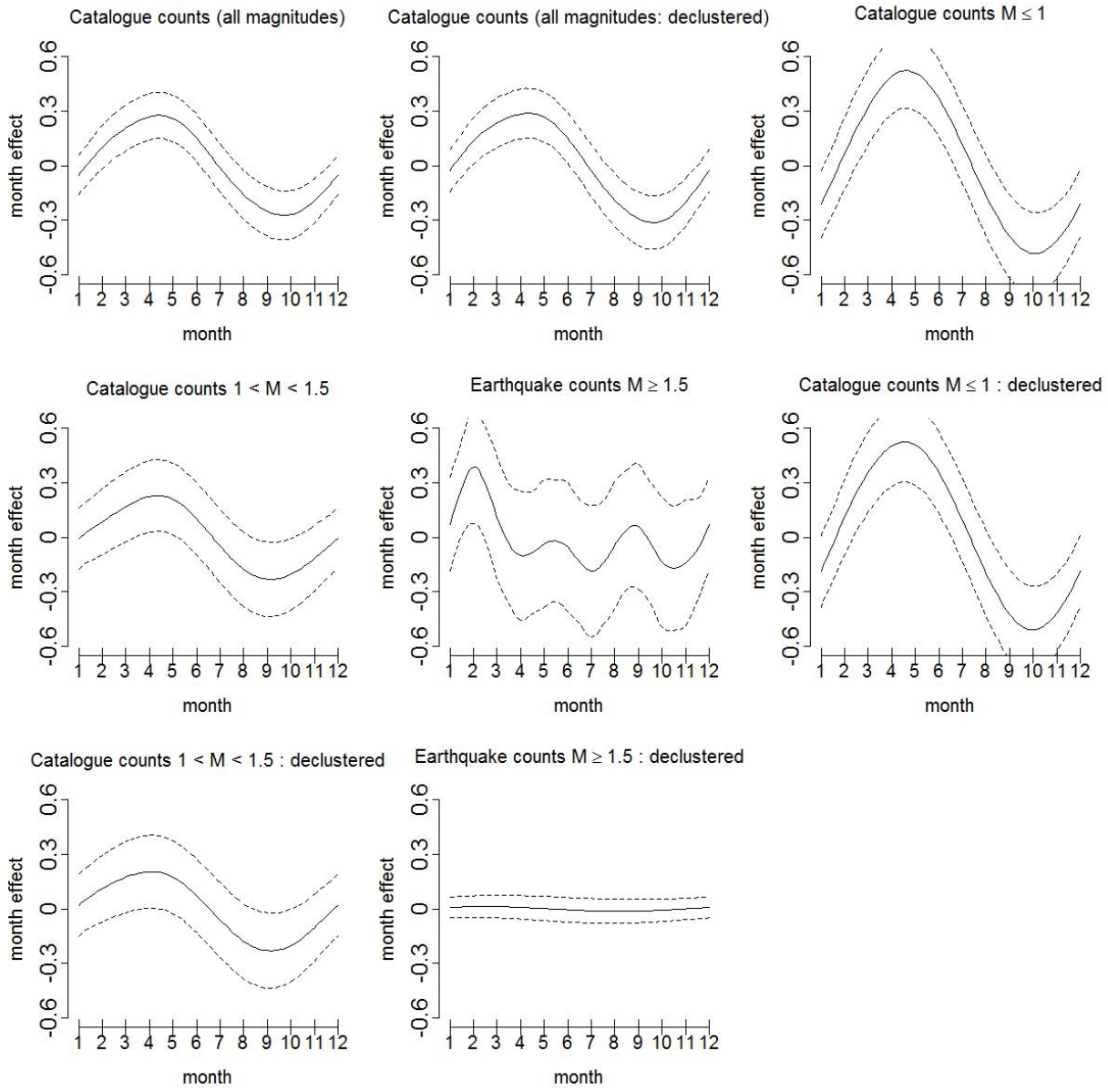

Figure 5.1.: Estimated monthly smooth deviations ("month effects") from the annual average event rate (equation 4.3), using the Poisson likelihood. The solid and dotted lines depict the estimated rates and the 95% confidence interval of the estimates respectively.

3 months, although there is much uncertainty surrounding this timing and significant correlations could also be found for lags of 2,4 or 5 month lags. Because monthly production rates and event rates vary (approximately) periodically, an almost equally good correlation between production and event rates can be found with a lag of 9 months though parameter β swaps sign. A more detailed investigation into the modeling assumptions was done which indicated that the potential influence of overdispersion on the inferences is minor. For example, with a 3 month lag for all events the following estimates were obtained: $\beta = 7.56$ with an estimated standard error of 1.38 under the Poisson model and 1.48 under the quasi-Poisson model, and an estimated overdispersion parameter for the quasi-Poisson family of 1.15. These estimates change little if declustering is applied: $\beta = 7.77$ with an estimated standard error of 1.42 under the Poisson model and 1.52 under the

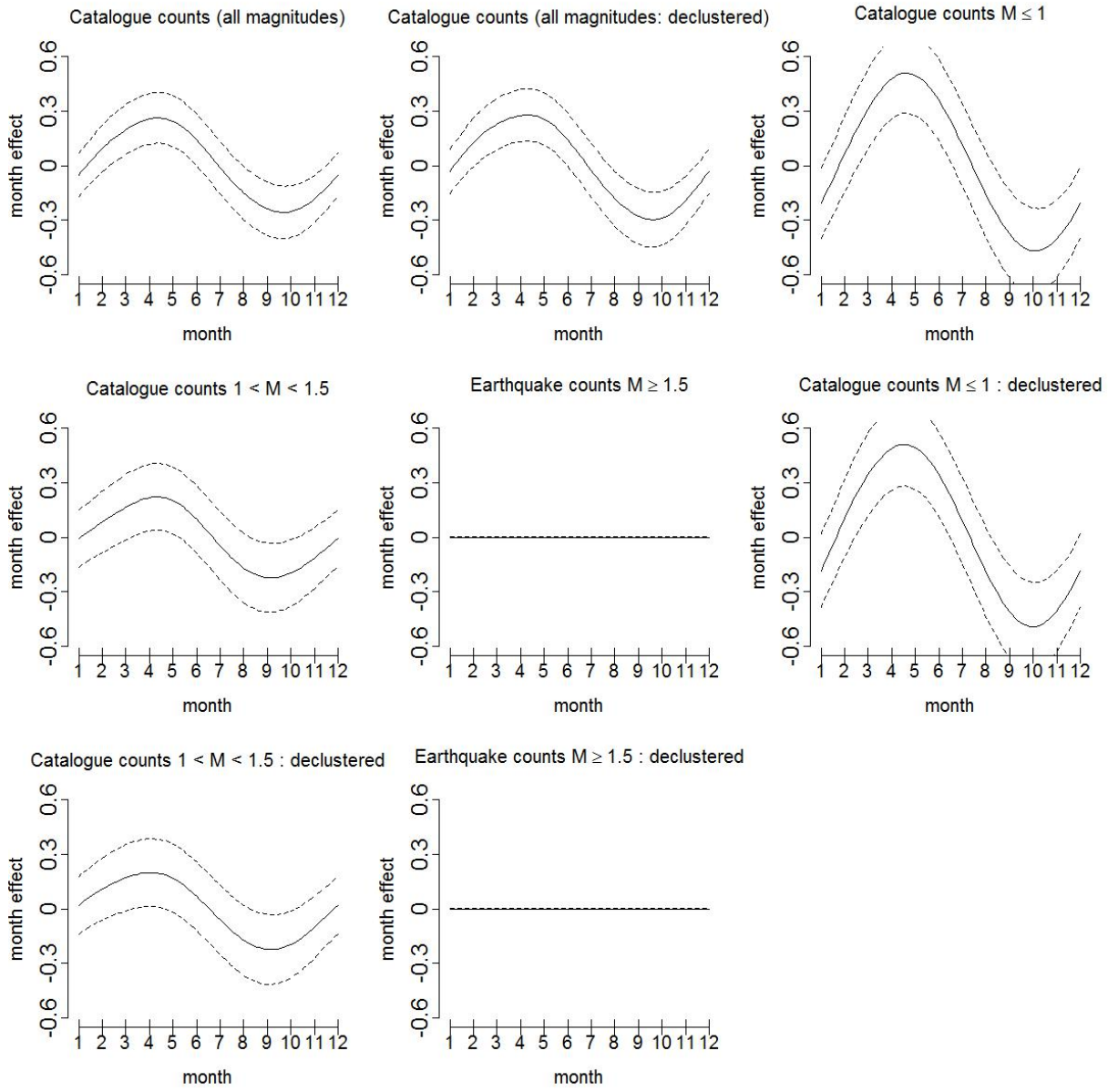

Figure 5.2.: Estimated monthly smooth deviations ("month effects") from the annual average event rate (equation 4.3), using the quasi-Poisson likelihood. The solid and dotted lines depict the estimated rates and the 95% confidence interval of the estimates respectively.

quasi-Poisson model, and an estimated overdispersion parameter for the quasi-Poisson family of 1.13. The assumption that counts are Poisson distributed is questionable (figure A.7). Despite the evidence that monthly production is correlated, with some time-lag, to monthly Catalogue counts of events with $M \leq 1$ there is clearly much uncertainty surrounding this relationship (figure 5.6).

5.2.3. Events with associated magnitudes $1 < M < 1.5$

Weak evidence was found of a relationship between monthly production and Catalogue rates of events with associated magnitudes $1 < M < 1.5$ (table 5.3 and B.3). The optimal lag between gas production and catalogue rate was estimated to be around 3 months, although there is much

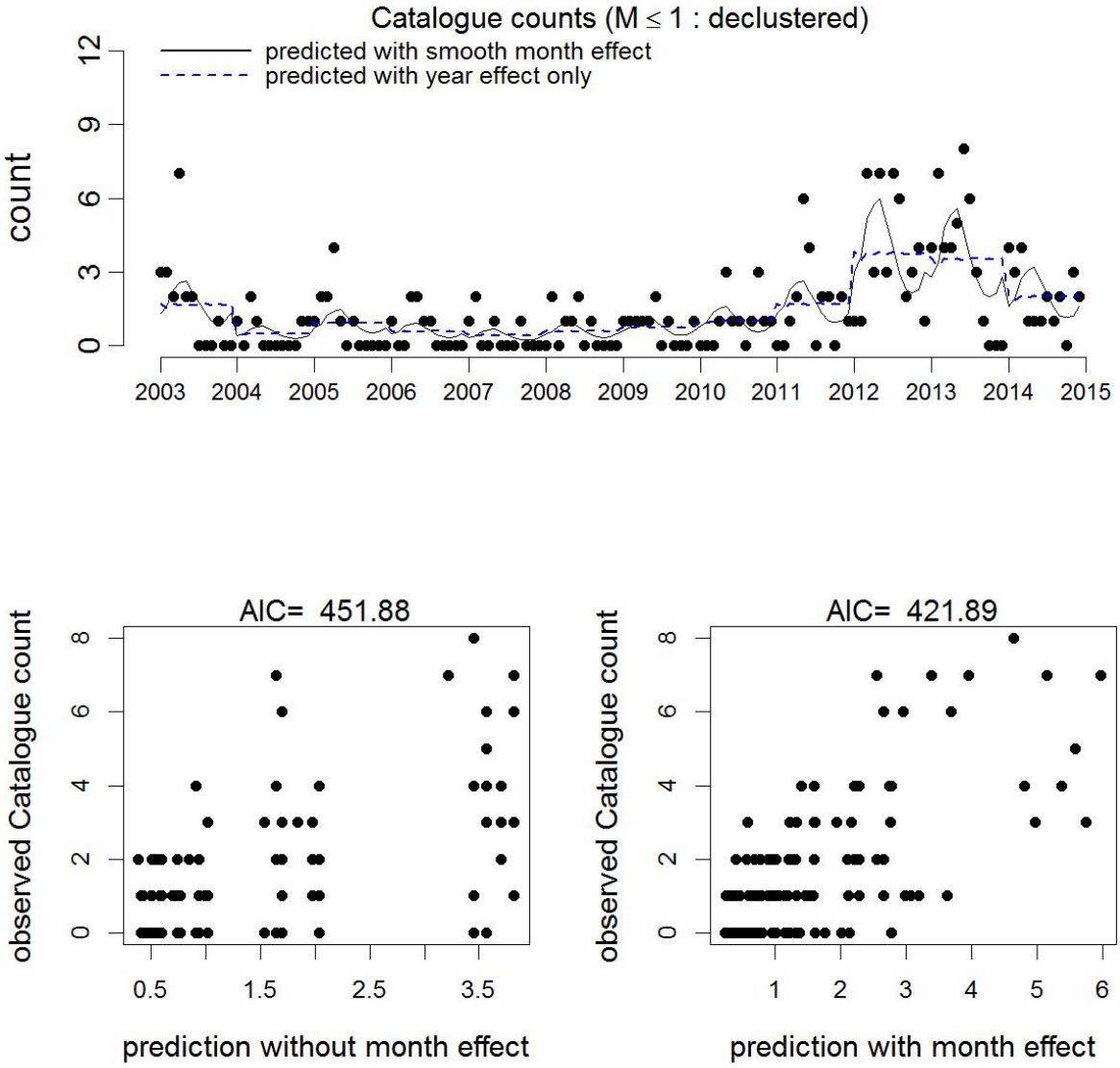

Figure 5.3.: Observed and predicted monthly counts of events with and without a smooth month effect for events with M $\leq$ 1. Top panel: time series of observed and predicted monthly counts. Bottom left panel: predicted versus observed monthly counts for the "null" model with yearly average rates only (equation 4.2). Bottom right panel: predicted versus observed monthly counts for the model with smooth month effect (equation 4.3). Also quoted is the Akaike Information Criterion (AIC) for each model. These results are from models with a Poisson likelihood.

uncertainty surrounding this timing and significant correlations could also be found for lags of 2,4 or 5 month lags. Because monthly production rates and event rates vary (approximately) periodically, an almost equally good correlation between production and event rates can be found with a lag of 9 months though parameter β swaps sign. A more detailed investigation into the modeling assumptions was done which indicated that the potential influence of overdispersion on the inferences is minor. For example, with a 3 month lag for all events the following estimates were ob-

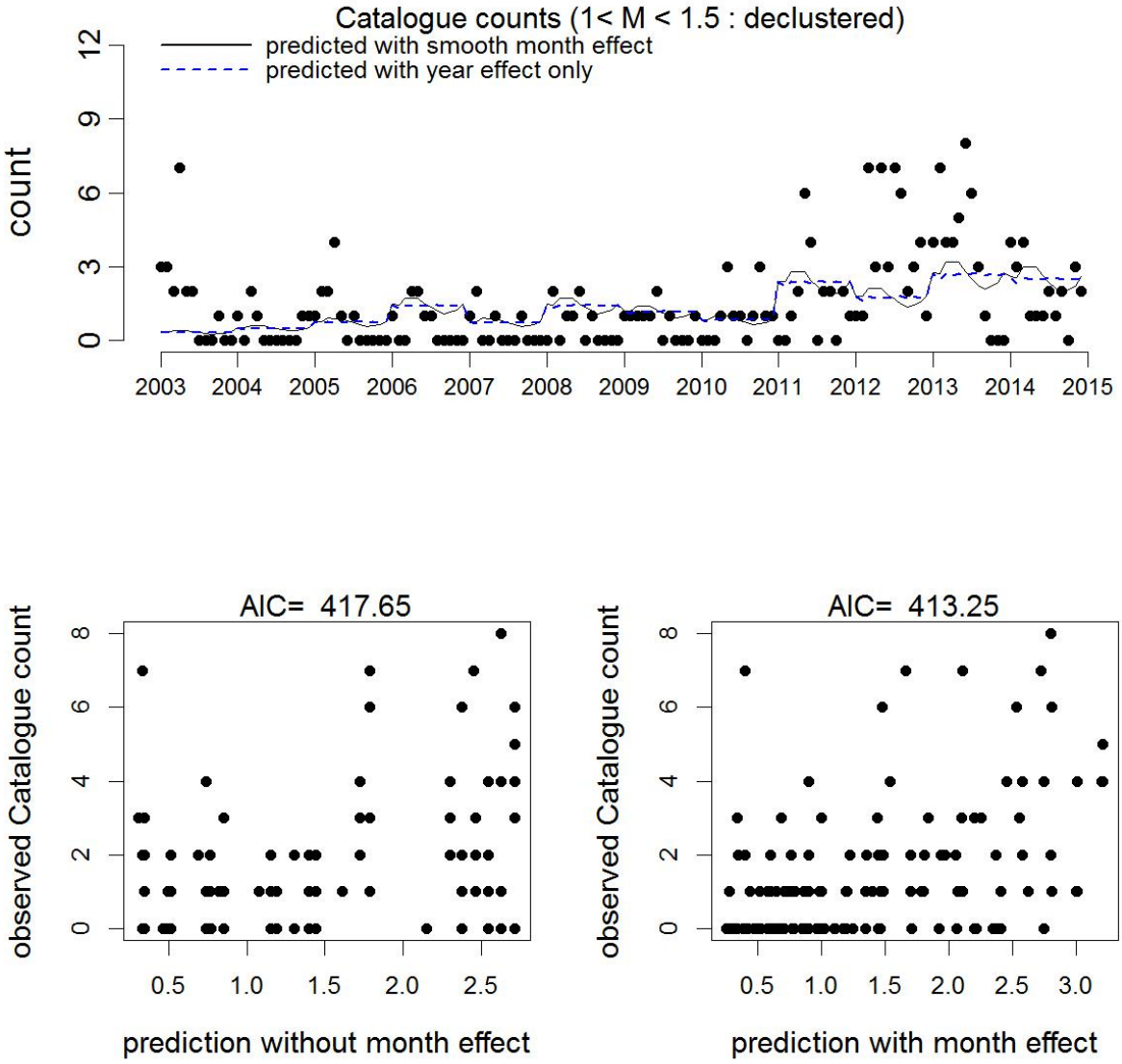

Figure 5.4.: Observed and predicted monthly counts of events with and without a smooth month effect for events with $1 < M < 1.5$. Top panel: time series of observed and predicted monthly counts. Bottom left panel: predicted versus observed monthly counts for the "null" model with yearly average rates only (equation 4.2). Bottom right panel: predicted versus observed monthly counts for the model with smooth month effect (equation 4.3). Also quoted is the Akaike Information Criterion (AIC) for each model. These results are from models with a Poisson likelihood.

tained: $\beta = 3.19$ with an estimated standard error of 1.27 under the Poisson model and 1.24 under the quasi-Poisson model, and an estimated overdispersion parameter for the quasi-Poisson family of 0.95. Different estimates are obtained if declustering is applied: $\beta = 2.66$ with an estimated standard error of 1.33 under the Poisson model and 1.29 under the quasi-Poisson model, and an estimated overdispersion parameter for the quasi-Poisson family of 0.94. This indicates that the estimates are sensitive to the application of declustering, and that these estimates are only marginally

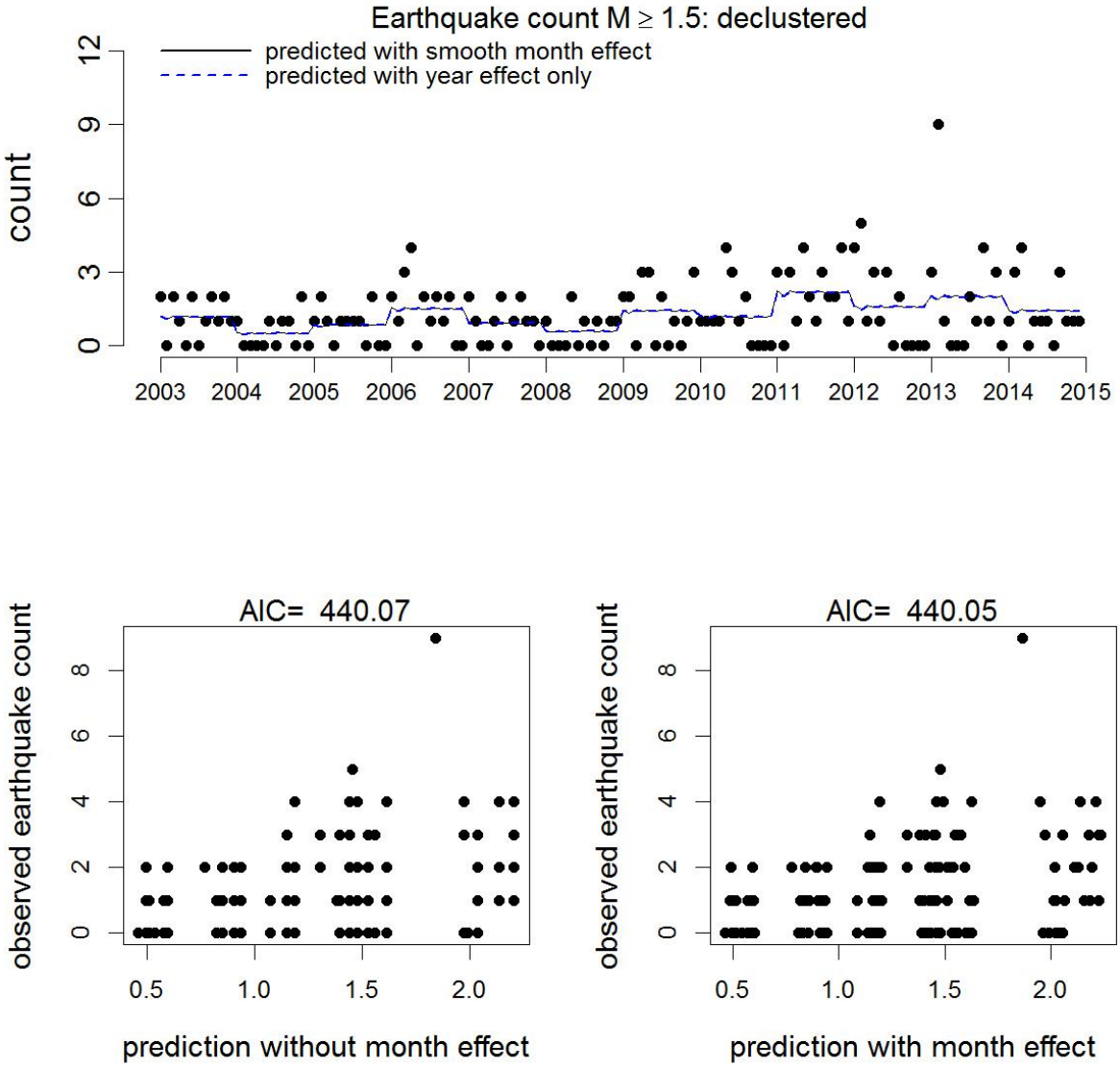

Figure 5.5.: Observed and predicted monthly counts of events with and without a smooth month effect (equation 4.3) for events with M $\geq$ 1.5. Top panel: time series of observed and predicted monthly counts. Bottom left panel: predicted versus observed monthly counts for the "null" model with yearly average rates only (equation 4.2). Bottom right panel: predicted versus observed monthly counts for the model with smooth month effect (equation 4.3). Also quoted is the Akaike Information Criterion (AIC) for each model. These results are from models with a Poisson likelihood.

significantly different from zero at the 95% confidence level. The assumption that counts are Poisson distributed appears to be reasonable (figure A.8).

Table 5.1.: Overview of model parameters and indicators for evidence that Catalogue rates of events (of all magnitudes) vary across months within year as a function of monthly production with some delay. The AIC values are from the extended model (equation 4.4 with slope parameter β) and are compared against the null model (equation 4.2), where $\Delta\text{AIC} = \text{AIC}_{\text{extended}} - \text{AIC}_{\text{null}}$. The estimate of the year-invariant parameter β and its standard error are compared against the standard normal (z) distribution to compute the quoted p-values. The quoted percentiles are computed from the resampling distributions of the average value of the annual slope parameters β_y.

delay	AIC	ΔAIC	β	SE(β)	P($>\|z\|$)	2.5%	5%	50%	95%	97.5%
0	551.63	-1.36	-0.62	0.779	0.426	-2.07	-1.87	-0.31	1.46	1.82
1	546.96	-0.13	1.05	0.762	0.170	-0.38	-0.13	1.32	2.65	2.88
2	533.97	15.88	3.22	0.759	<0.001	1.88	2.10	3.58	5.36	5.72
3	522.88	29.27	4.26	0.761	<0.001	2.26	2.62	4.21	5.36	5.56
4	526.78	23.11	3.79	0.755	<0.001	0.11	0.53	2.98	4.51	4.67
5	533.14	14.28	3.02	0.746	<0.001	-0.59	-0.11	2.00	3.67	3.97
6	546.18	-0.03	1.04	0.740	0.160	-1.77	-1.54	-0.03	1.40	1.67
7	547.06	0.38	-1.14	0.743	0.124	-3.34	-3.09	-1.96	-0.90	-0.76
8	529.22	17.15	-3.27	0.760	<0.001	-6.12	-5.78	-3.85	-2.22	-1.93
9	521.09	25.90	-4.01	0.779	<0.001	-6.35	-6.01	-4.07	-2.25	-1.71
10	526.22	18.41	-3.45	0.780	<0.001	-5.06	-4.80	-2.89	-0.80	-0.33
11	545.42	4.77	-1.94	0.754	0.010	-2.99	-2.80	-1.32	0.44	0.83

5.2.4. Events with associated magnitudes M ≥ 1.5

No evidence was found of a relationship between monthly production and earthquake rates M ≥ 1.5 (table 5.4 and B.4).

Table 5.2.: Overview of model parameters and indicators for evidence that Catalogue Rates of events with M $\leq$ 1 vary across months within year as a function of monthly production with some delay. The AIC values are from the extended model (equation 4.4 with slope parameter β) and are compared against the null model (equation 4.2), where $\Delta\text{AIC} = \text{AIC}_{\text{extended}} - \text{AIC}_{\text{null}}$. The estimate of the year-invariant parameter β and its standard error are compared against the standard normal (z) distribution to compute the quoted p-values. The quoted percentiles are computed from the resampling distributions of the average value of the annual slope parameters β_y.

delay	AIC	ΔAIC	β	SE(β)	P($>\|z\|$)	2.5%	5%	50%	95%	97.5%
0	363.53	4.20	-3.53	1.446	0.015	-6.10	-5.70	-2.41	1.38	2.26
1	369.44	-1.91	0.42	1.371	0.758	-1.75	-1.25	2.46	6.06	6.94
2	359.86	8.83	4.49	1.362	0.001	3.08	3.69	7.79	12.07	12.79
3	341.85	27.21	7.40	1.384	<0.001	6.09	6.56	9.19	11.93	12.46
4	339.38	29.06	7.56	1.375	<0.001	4.11	4.61	7.27	10.09	10.79
5	345.98	21.42	6.46	1.347	<0.001	1.39	2.13	5.06	7.91	8.23
6	364.60	2.53	2.79	1.309	0.033	-5.79	-4.88	-0.54	3.63	4.33
7	367.53	-1.24	-1.14	1.311	0.384	-12.15	-10.96	-5.41	-1.26	-0.50
8	358.47	7.63	-4.10	1.351	0.002	-12.90	-12.27	-8.18	-4.58	-3.72
9	348.02	18.68	-6.24	1.439	<0.001	-20.07	-18.75	-11.43	-5.58	-4.68
10	337.80	27.83	-7.76	1.517	<0.001	-16.03	-14.87	-9.64	-5.08	-3.95
11	350.09	16.82	-5.97	1.440	<0.001	-9.66	-9.14	-5.75	-1.60	-0.61

Table 5.3.: Overview of model parameters and indicators for evidence that Catalogue Rates of events with 1 $<$ M $<$ 1.5 vary across months within year as a function of monthly production with some delay. The AIC values are from the extended model (equation 4.4 with slope parameter β) and are compared against the null model (equation 4.2), where $\Delta\text{AIC} = \text{AIC}_{\text{extended}} - \text{AIC}_{\text{null}}$. The estimate of the year-invariant parameter β and its standard error are compared against the standard normal (z) distribution to compute the quoted p-values. The quoted percentiles are computed from the resampling distributions of the average value of the annual slope parameters β_y.

delay	AIC	ΔAIC	β	SE(β)	P($>\|z\|$)	2.5%	5%	50%	95%	97.5%
0	366.19	-2.00	-0.03	1.311	0.981	-2.15	-1.76	0.33	2.59	2.95
1	365.93	-1.21	1.15	1.284	0.372	-0.57	-0.38	1.22	2.73	2.99
2	361.96	2.84	2.82	1.278	0.027	0.48	0.92	3.18	5.57	5.96
3	361.51	4.94	3.39	1.282	0.008	-1.25	-0.50	3.03	6.25	6.93
4	360.83	4.26	3.19	1.273	0.012	-2.22	-1.43	2.03	4.86	5.42
5	361.25	2.75	2.75	1.256	0.029	-1.87	-1.45	0.97	3.15	3.46
6	364.96	-1.33	1.02	1.243	0.411	-2.60	-2.26	-0.36	1.82	2.29
7	366.11	-1.60	-0.79	1.244	0.526	-3.45	-3.25	-1.66	0.05	0.40
8	360.09	4.63	-3.24	1.279	0.011	-8.54	-7.90	-4.56	-1.50	-0.97
9	359.76	4.30	-3.17	1.288	0.014	-7.49	-6.92	-3.43	-0.27	0.35
10	363.01	1.13	-2.24	1.284	0.081	-5.16	-4.80	-2.25	0.40	0.85
11	367.28	-1.59	-0.79	1.245	0.524	-2.29	-1.92	-0.06	1.85	2.26

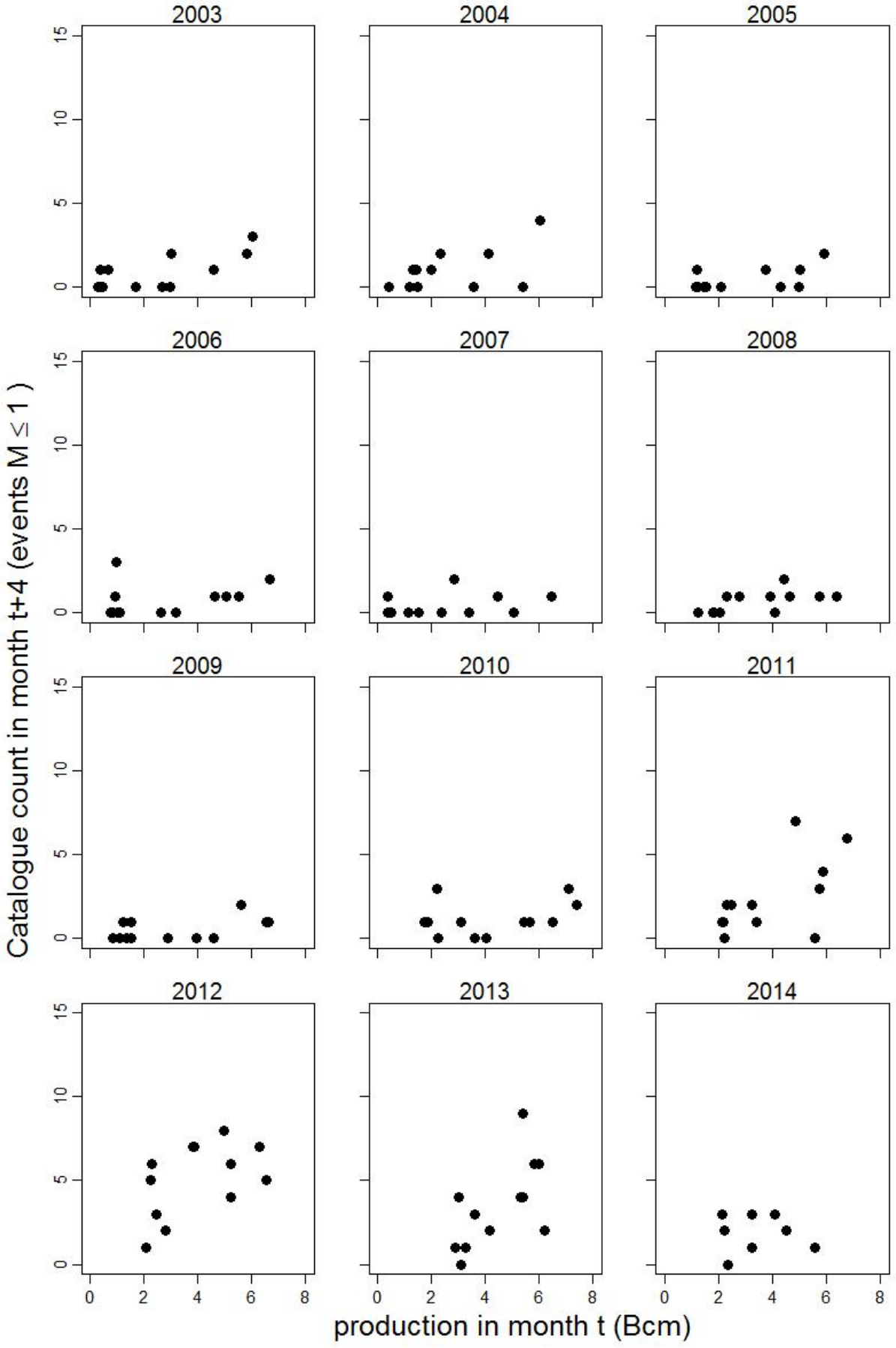

Figure 5.6.: Monthly counts of events versus monthly field-wide production with a delay of 4 calendar months (one panel per year).

Table 5.4.: Overview of model parameters and indicators for evidence that Earthquake rates of events with M $\geq$ 1.5 vary across months within year as a function of monthly production with some delay. The AIC values are from the extended model (equation 4.4 with slope parameter β) and are compared against the null model (equation 4.2), where $\Delta\text{AIC} = \text{AIC}_{\text{extended}} - \text{AIC}_{\text{null}}$. The estimate of the year-invariant parameter β and its standard error are compared against the standard normal (z) distribution to compute the quoted p-values. The quoted percentiles are computed from the resampling distributions of the average value of the annual slope parameters β_y.

delay	AIC	ΔAIC	β	SE(β)	P($>\|z\|$)	2.5%	5%	50%	95%	97.5%
0	388.78	-0.95	1.35	1.312	0.304	-1.11	-0.74	1.44	3.37	3.79
1	384.73	-0.68	1.51	1.306	0.248	-2.10	-1.67	0.47	2.84	3.32
2	384.33	1.51	2.46	1.309	0.060	-3.83	-3.23	0.58	4.16	4.78
3	385.09	1.04	2.30	1.314	0.080	-4.36	-3.63	-0.02	3.90	4.64
4	387.37	-1.52	0.92	1.319	0.487	-5.48	-4.73	-1.14	2.60	3.55
5	387.50	-2.00	0.03	1.315	0.979	-4.25	-3.59	-0.77	2.01	2.55
6	386.61	-1.74	-0.66	1.310	0.614	-2.72	-2.38	-0.46	1.31	1.66
7	386.73	-0.62	-1.53	1.311	0.242	-4.19	-3.76	-1.48	0.50	0.83
8	383.42	1.62	-2.48	1.322	0.060	-6.51	-6.04	-2.02	1.67	2.18
9	383.10	2.61	-2.83	1.339	0.035	-7.21	-6.49	-2.60	0.91	1.49
10	385.81	-1.46	-0.96	1.311	0.463	-4.89	-4.05	-0.18	3.35	3.96
11	388.97	-1.88	0.44	1.278	0.729	-3.14	-2.52	0.74	3.08	3.47

6. Conclusions

In this report, we describe and apply statistical methodology to test for evidence of within-year variation (seasonality) in rates of occurrence of earthquake events as recorded in the Earthquake Catalogue (EC), and for evidence of a relationship between seasonal (monthly) variation in gas production and earthquake rates. We pay special attention to possible differences in apparent seasonality of rates of events within different ranges of event magnitudes. With the current network of geophones, inclusion probabilities of events with magnitudes M<1.5 may vary both spatially and temporally. The probability that an earthquake with associated magnitude M<1.5, when it occurs within the Groningen field, is detected by the geophone network and included in the EC (the so-called "inclusion probability") cannot be assumed to be unity nor to be spatio-temporally invariant. In this report we have assumed, in line with recent advice from the Dutch Meteorological Society (KNMI), that inclusion probabilities for events with magnitudes $M \geq 1.5$ can be assumed to be 1 or close to 1 throughout the Groningen field and throughout the time-series under consideration. We therefore use the terminology "Earthquake rate" or "Earthquake count" for observed counts or estimated rates of events with magnitudes $M \geq 1.5$. We use the terminology "Catalogue Rate" or "Catalogue Count" for observed counts or estimated rates of events with magnitudes M<1.5.

Our main finding is that there is an indication that catalogue rates vary seasonally within year and may be correlated with some time-lag to gas production rates. The seasonal variation in catalogue rates appears to be mostly if not wholly due to seasonal variation in catalogue rates of events with magnitudes $M \leq 1$ for which there is good evidence that these rates may be seasonally varying. The interpretation of this finding is complicated due to the spatio-temporally uneven inclusion probabilities of events of these magnitudes. Equally, the apparent diurnality in rates of events with magnitudes $M \leq 1$ warrants further investigation as to the possible causes underlying the trends. There was only weak evidence of seasonal variation in events with magnitudes $1 < M < 1.5$ and no evidence for events with magnitudes $M \geq 1.5$.

Appendix A.

Further graphs

In this appendix, we provide further graphs to illustrate monthly variation in counts of events in the KNMI catalogue, for events of all magnitudes (A.2), magnitudes $M \leq 1$ (A.3), magnitudes $1 < M < 1.5$ (A.4) and magnitudes $M \geq 1.5$ (A.5).

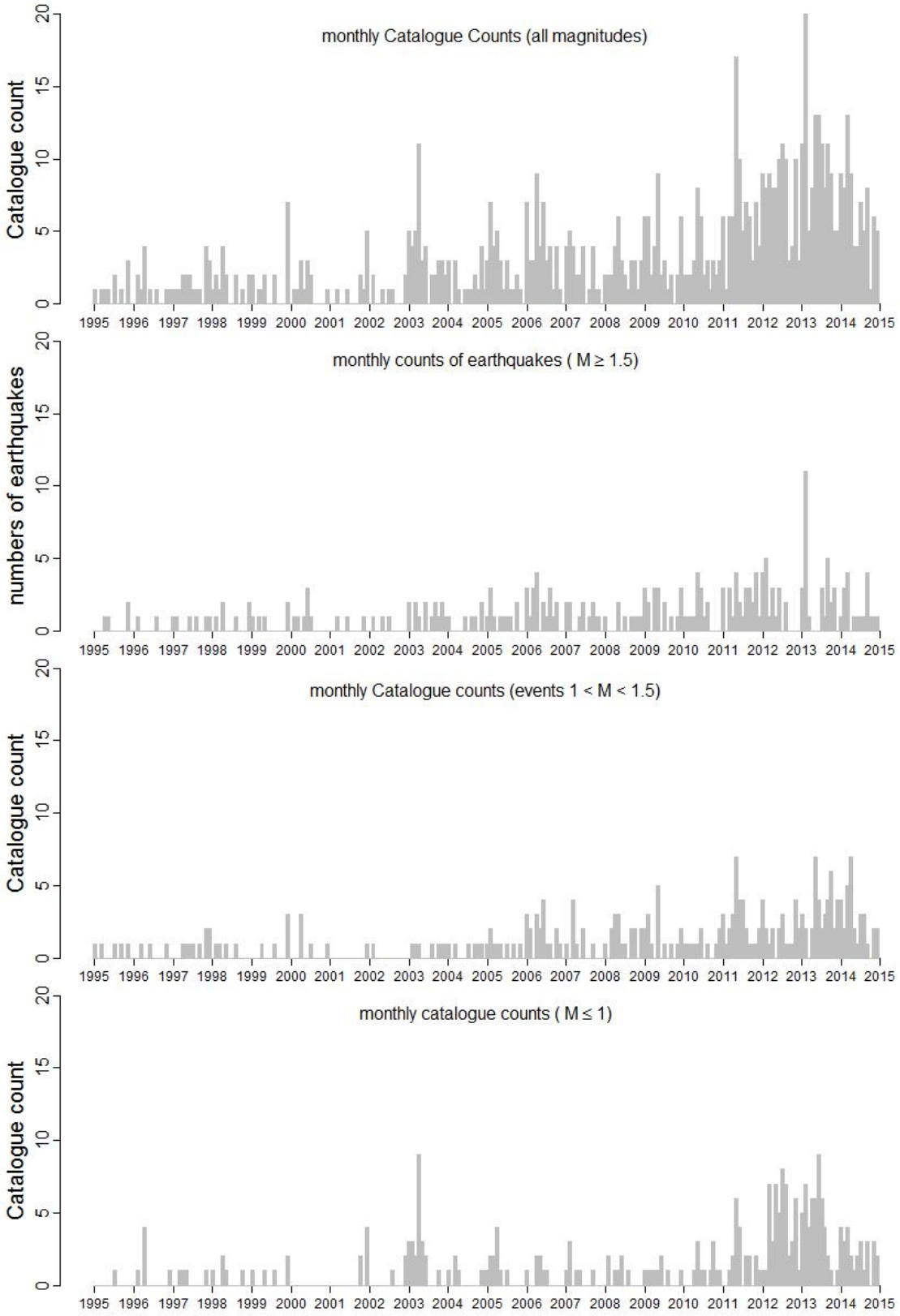

Figure A.1.: Time series of counts of events per calendar month for all events (top graph) or events in different categories of associated magnitudes.

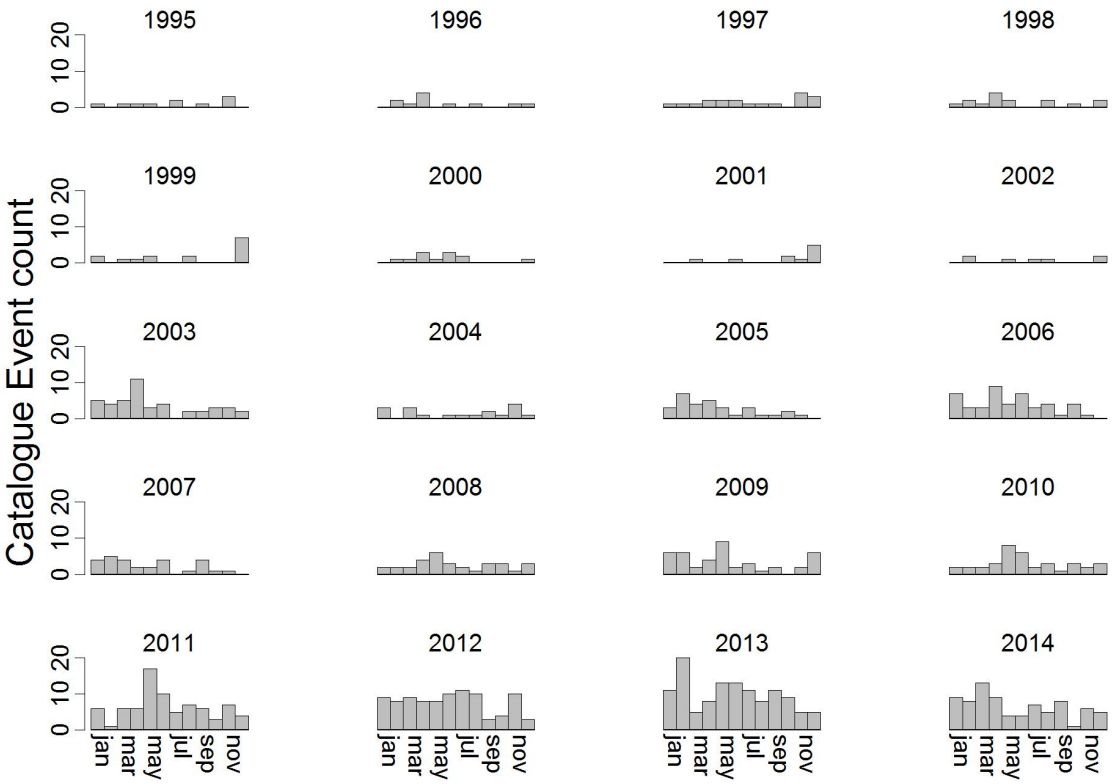

Figure A.2.: Time series of counts of events per calendar month for all events, with a panel per year.

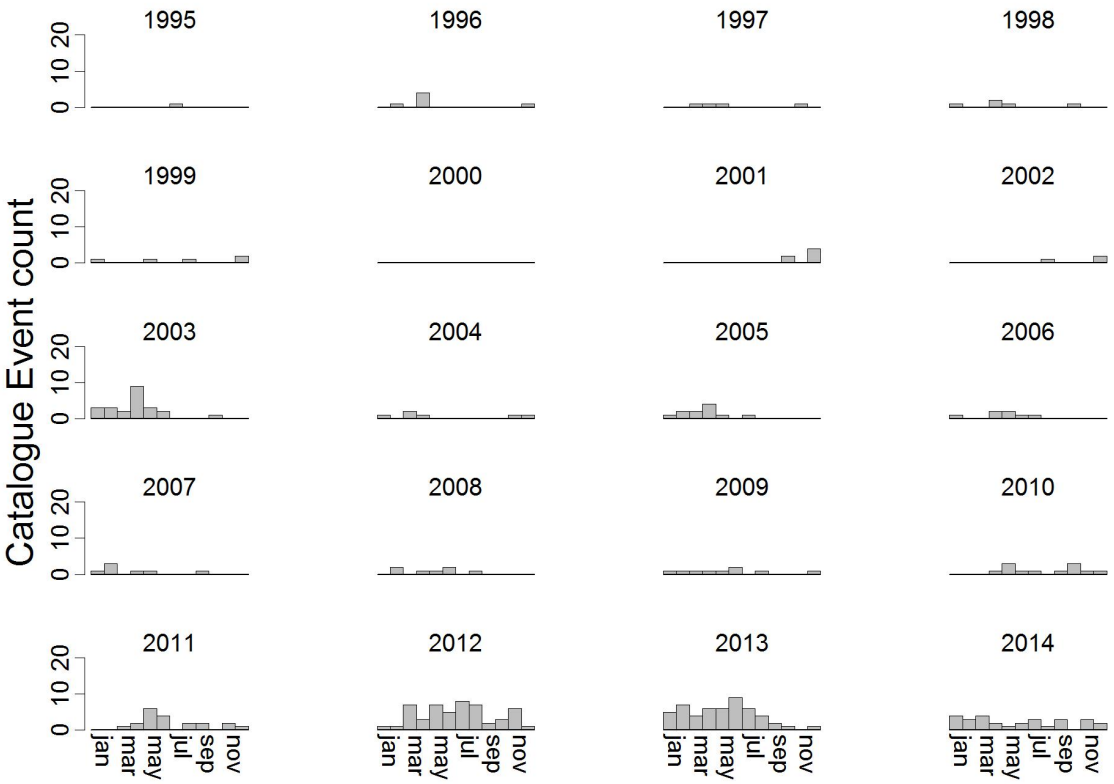

Figure A.3.: Time series of catalogue counts of events per calendar month events with magnitudes $M \leq 1$, with a panel per year.

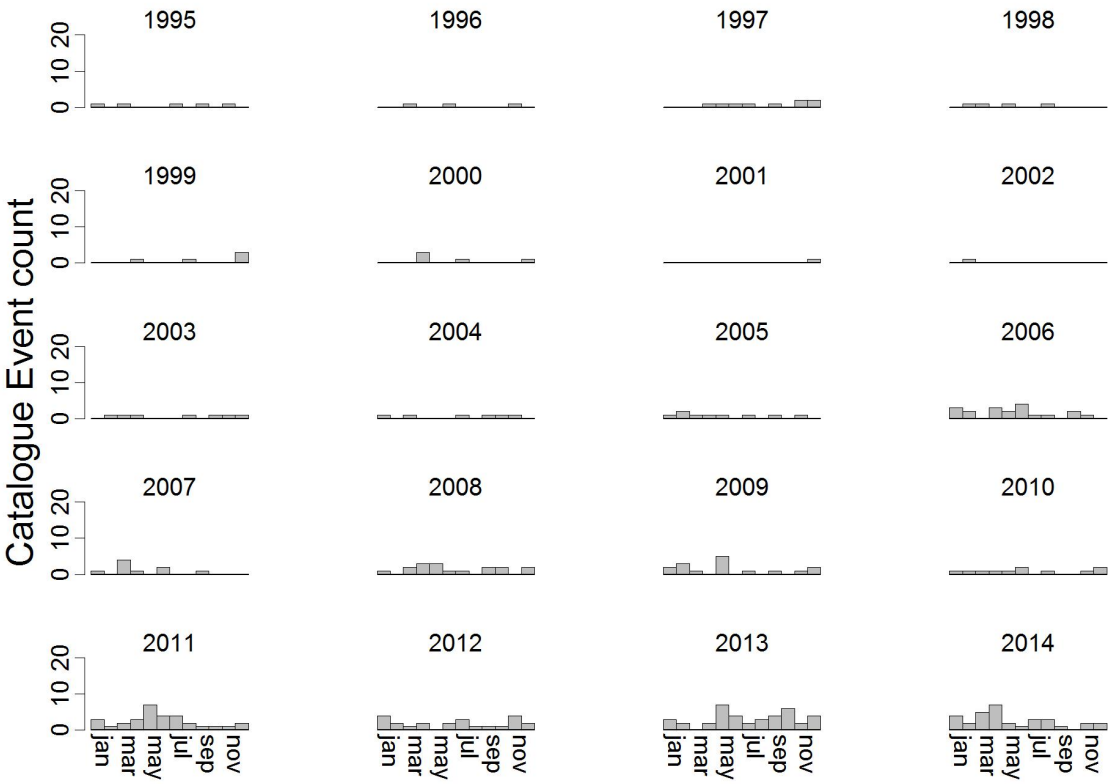

Figure A.4.: Time series of counts of earthquakes per calendar month events with magnitudes $1 < M < 1.5$, with a panel per year.

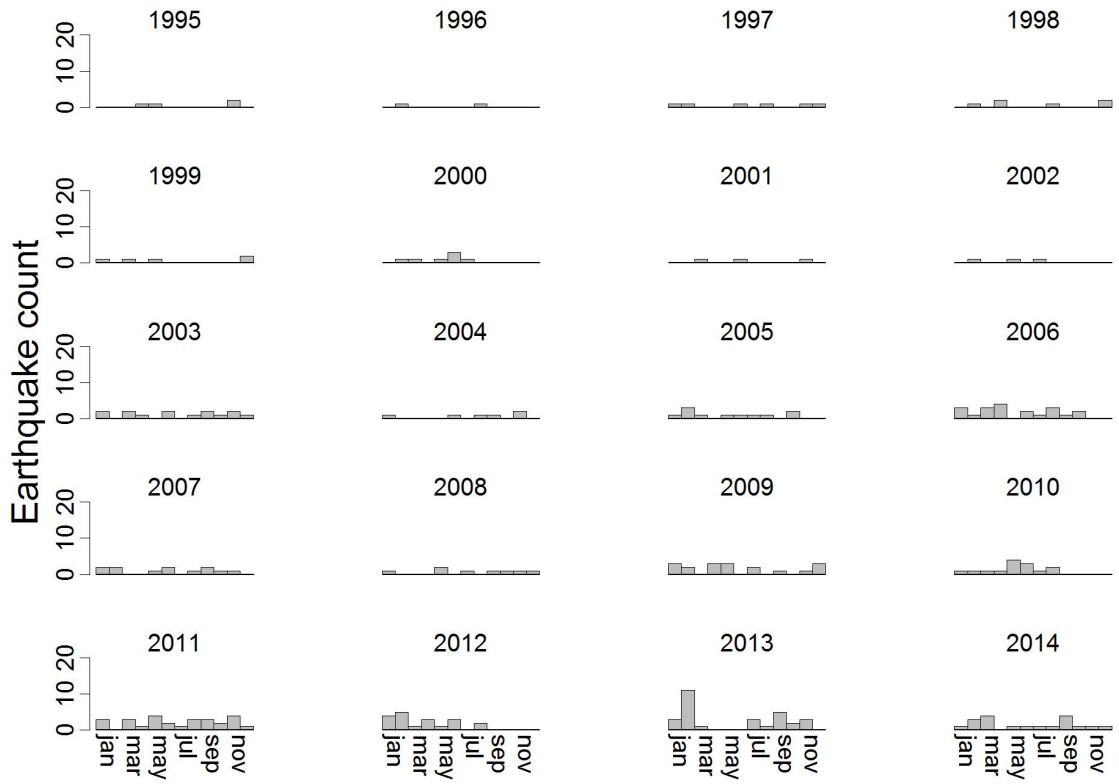

Figure A.5.: Time series of counts of earthquakes per calendar month events with magnitudes $M \geq 1.5$, with a panel per year.

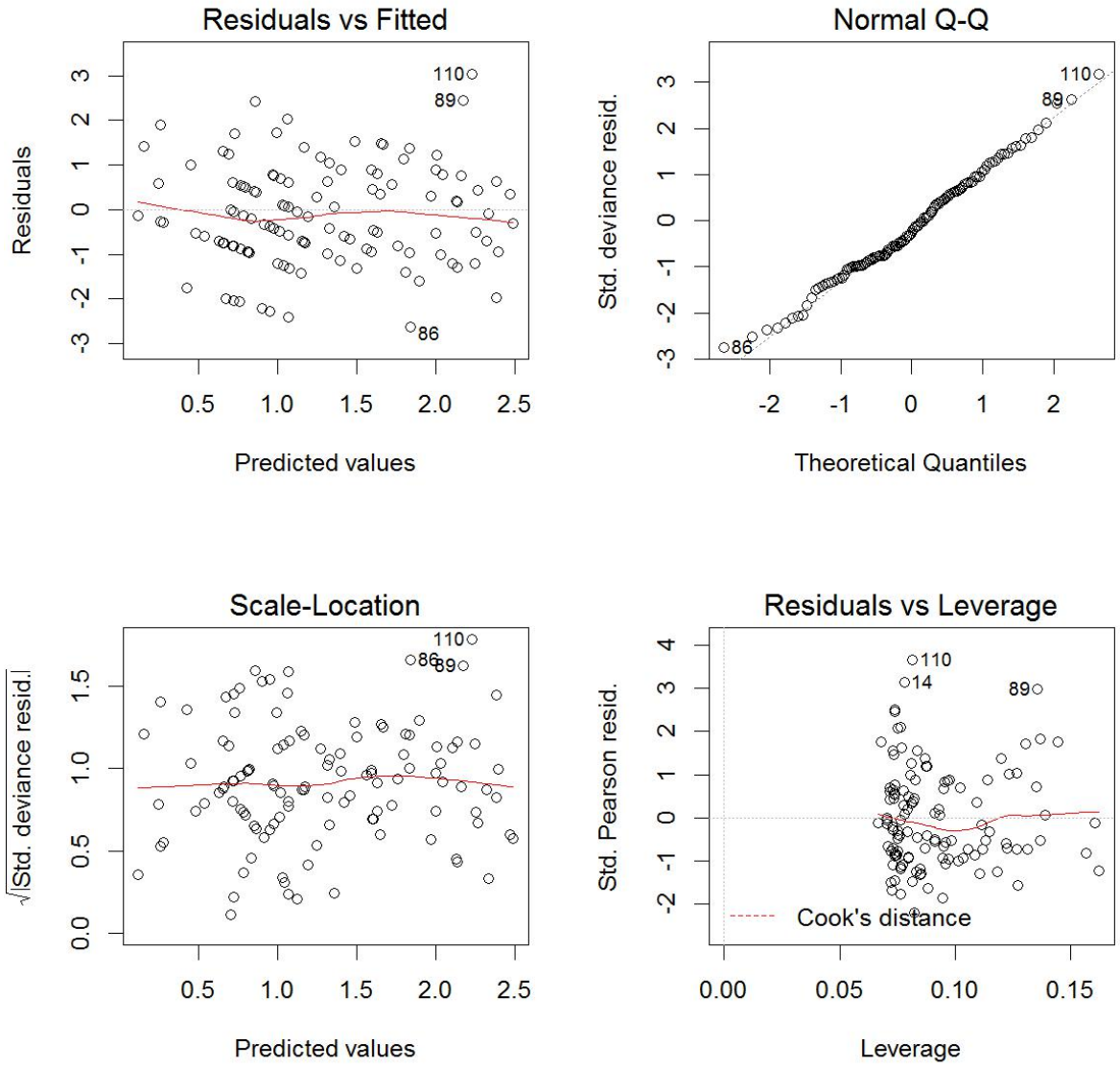

Figure A.6.: A visual assessment of the model assumptions for the Poisson model of catalogue rates of all events.

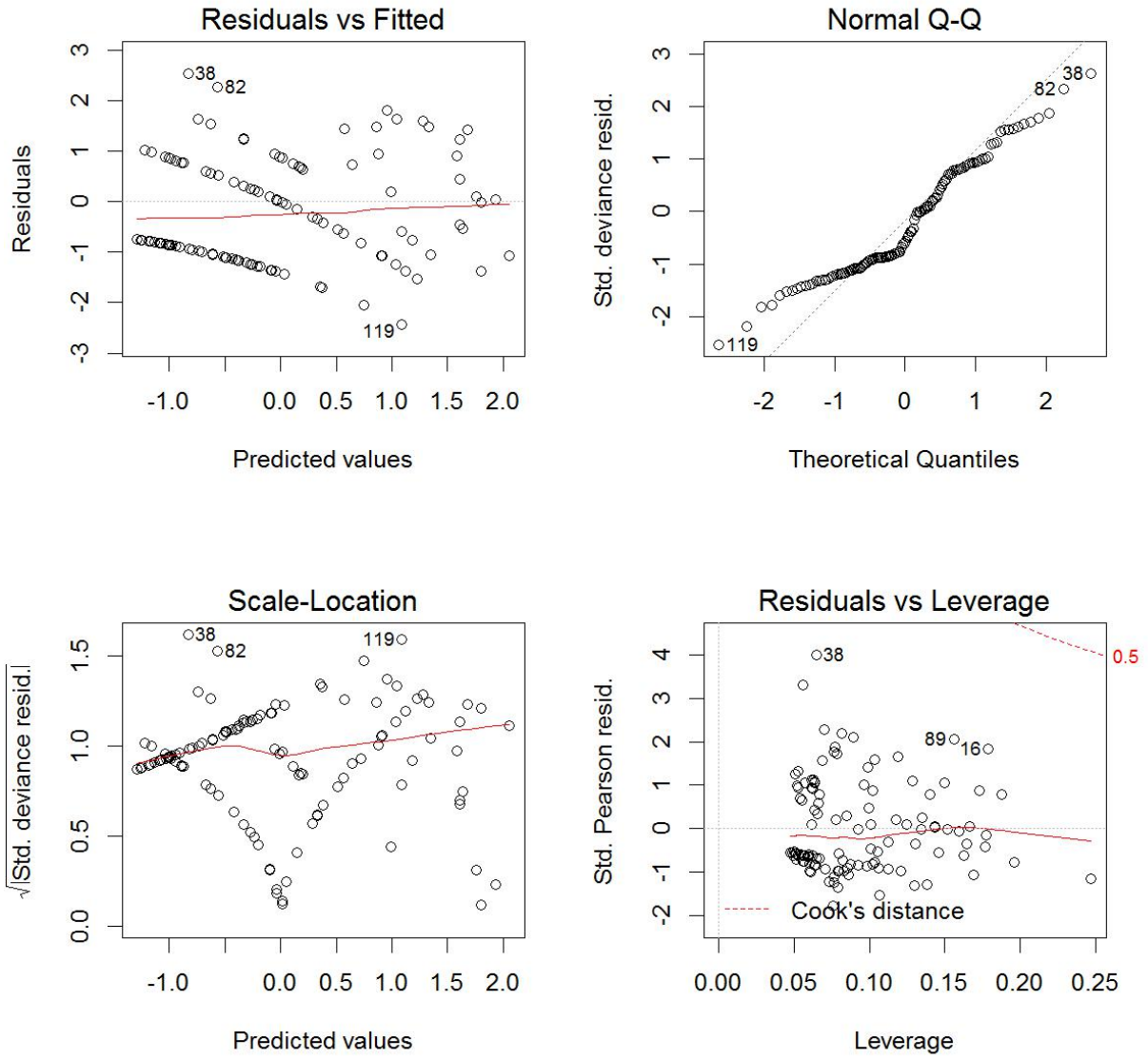

Figure A.7.: A visual assessment of the model assumptions for the Poisson model of catalogue rates of events with $M \leq 1$.

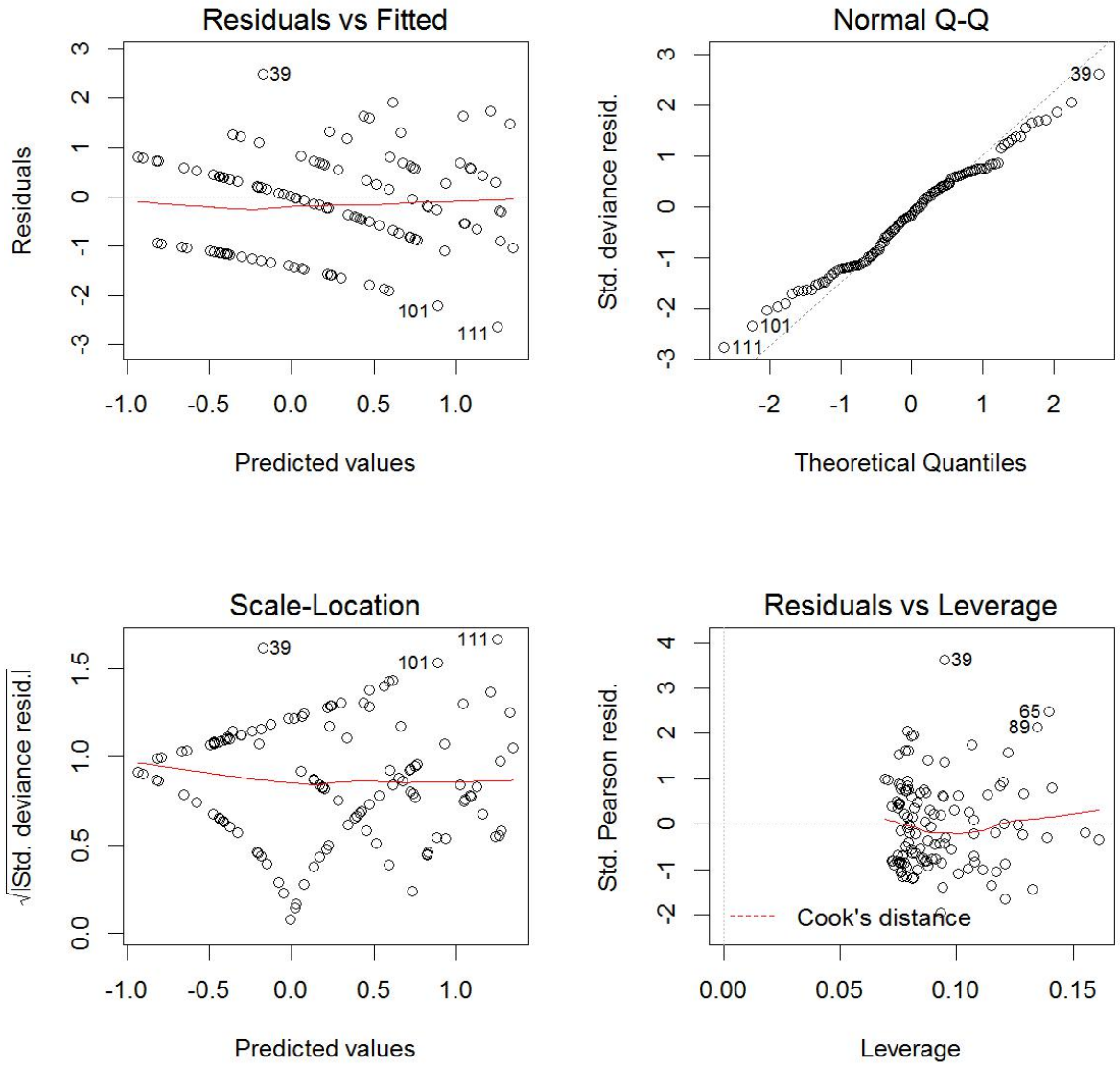

Figure A.8.: A visual assessment of the model assumptions for the Poisson model of catalogue rates of events with $1 < M < 1.5$.

Appendix B.

Further tables with results

In this appendix, in addition to the tables shown in section 5.2, we provide further tables of results of the fitting of the Poisson models to declustered data sets for events of all magnitudes (B.1), magnitudes M $\leq$ 1 (B.2), magnitudes 1 < M < 1.5 (B.3) and magnitudes M $\geq$ 1.5 (B.4).

Table B.1.: Overview of model parameters and indicators for evidence that Catalogue Rates of declustered events of all magnitudes vary across months within year as a function of monthly production with some delay. The AIC values are from the extended model (equation 4.4 with slope parameter β) and are compared against the null model (equation 4.2), where $\Delta\text{AIC} = \text{AIC}_{\text{extended}} - \text{AIC}_{\text{null}}$. The estimate of the year-invariant parameter β and its standard error are compared against the standard normal (z) distribution to compute the quoted p-values. The quoted percentiles are computed from the resampling distributions of the average value of the annual slope parameters β_y.

delay	AIC	ΔAIC	β	SE(β)	P($>\|z\|$)	2.5%	5%	50%	95%	97.5%
0	531.68	-1.48	-0.58	0.810	0.473	-2.03	-1.74	-0.37	1.37	1.70
1	528.08	0.09	1.15	0.791	0.147	-0.48	-0.14	1.32	2.67	2.84
2	513.53	16.52	3.40	0.788	<0.001	2.10	2.31	3.52	5.23	5.52
3	504.52	28.42	4.37	0.791	<0.001	2.34	2.58	4.17	5.43	5.66
4	507.90	22.05	3.86	0.785	<0.001	0.21	0.61	2.94	4.55	4.65
5	515.37	12.87	3.00	0.776	<0.001	-0.71	-0.29	2.06	3.62	3.80
6	527.88	-0.73	0.87	0.771	0.258	-1.61	-1.40	0.02	1.34	1.60
7	526.86	1.06	-1.35	0.776	0.082	-3.09	-2.97	-1.88	-0.94	-0.74
8	509.35	17.34	-3.43	0.794	<0.001	-5.57	-5.29	-3.79	-2.27	-2.02
9	501.17	26.41	-4.22	0.815	<0.001	-6.29	-6.02	-4.04	-2.24	-1.88
10	507.56	18.21	-3.57	0.814	<0.001	-5.00	-4.69	-3.01	-0.94	-0.42
11	525.67	5.19	-2.09	0.787	0.008	-3.17	-3.02	-1.60	0.24	0.58

Table B.2.: Overview of model parameters and indicators for evidence that Catalogue Rates of declustered events with $M \leq 1$ vary across months within year as a function of monthly production with some delay. The AIC values are from the extended model (equation 4.4 with slope parameter β) and are compared against the null model (equation 4.2), where $\Delta\text{AIC} = \text{AIC}_{\text{extended}} - \text{AIC}_{\text{null}}$. The estimate of the year-invariant parameter β and its standard error are compared against the standard normal (z) distribution to compute the quoted p-values. The quoted percentiles are computed from the resampling distributions of the average value of the annual slope parameters β_y.

delay	AIC	ΔAIC	β	SE(β)	P($>\|z\|$)	2.5%	5%	50%	95%	97.5%
0	357.09	3.48	-3.46	1.504	0.022	-6.09	-5.56	-2.39	1.30	2.29
1	362.69	-1.78	0.67	1.419	0.638	-1.77	-1.19	2.49	6.15	6.68
2	351.20	9.99	4.89	1.412	0.001	3.15	4.03	7.97	12.07	12.90
3	333.40	28.83	7.89	1.442	<0.001	5.77	6.35	9.39	12.64	13.28
4	332.79	28.41	7.77	1.429	<0.001	2.76	3.61	7.11	10.51	10.94
5	341.08	19.30	6.39	1.397	<0.001	0.72	1.56	4.98	7.75	8.13
6	358.54	1.62	2.59	1.359	0.056	-5.88	-5.19	-0.21	3.23	3.78
7	360.01	-0.87	-1.45	1.368	0.290	-11.82	-10.71	-5.13	-1.08	-0.69
8	350.91	8.35	-4.44	1.416	0.002	-12.25	-11.58	-8.01	-4.10	-3.39
9	338.90	20.64	-6.86	1.523	<0.001	-19.41	-17.73	-11.49	-6.09	-4.81
10	329.83	28.84	-8.29	1.606	<0.001	-15.67	-14.97	-9.67	-4.24	-3.18
11	342.86	17.22	-6.33	1.518	<0.001	-9.72	-9.01	-5.58	-1.13	-0.24

Table B.3.: Overview of model parameters and indicators for evidence that Catalogue Rates of declustered events with $1 < M < 1.5$ vary across months within year as a function of monthly production with some delay. The AIC values are from the extended model (equation 4.4 with slope parameter β) and are compared against the null model (equation 4.2), where $\Delta\text{AIC} = \text{AIC}_{\text{extended}} - \text{AIC}_{\text{null}}$. The estimate of the year-invariant parameter β and its standard error are compared against the standard normal (z) distribution to compute the quoted p-values. The quoted percentiles are computed from the resampling distributions of the average value of the annual slope parameters β_y.

delay	AIC	ΔAIC	β	SE(β)	P($>\|z\|$)	2.5%	5%	50%	95%	97.5%
0	354.14	-1.95	0.32	1.367	0.816	-2.50	-2.16	0.58	3.06	3.51
1	353.83	-0.98	1.36	1.343	0.311	-0.46	-0.16	1.36	2.98	3.20
2	350.08	2.59	2.88	1.339	0.031	0.75	1.08	3.18	5.39	5.78
3	351.88	2.62	2.90	1.343	0.031	-1.33	-0.70	2.55	5.83	6.34
4	350.72	1.95	2.66	1.334	0.046	-2.33	-1.82	1.68	4.71	5.15
5	351.36	1.05	2.31	1.317	0.080	-2.21	-1.75	0.66	2.98	3.33
6	353.66	-1.76	0.64	1.305	0.624	-3.08	-2.64	-0.72	1.43	1.93
7	354.14	-1.49	-0.93	1.306	0.477	-3.52	-3.15	-1.64	0.22	0.62
8	348.85	3.72	-3.15	1.338	0.019	-8.54	-7.73	-4.33	-1.41	-0.71
9	349.12	2.87	-2.91	1.341	0.030	-6.88	-6.24	-3.34	-0.20	0.54
10	352.61	0.06	-1.90	1.335	0.156	-4.74	-4.31	-1.75	0.78	1.38
11	355.60	-1.85	-0.50	1.297	0.701	-2.21	-1.76	0.33	2.34	2.71

Table B.4.: Overview of model parameters and indicators for evidence that Earthquake rates of declustered events with M $\geq$ 1.5 vary across months within year as a function of monthly production with some delay. The AIC values are from the extended model (equation 4.4 with slope parameter β) and are compared against the null model (equation 4.2), where $\Delta\text{AIC} = \text{AIC}_{\text{extended}} - \text{AIC}_{\text{null}}$. The estimate of the year-invariant parameter β and its standard error are compared against the standard normal (z) distribution to compute the quoted p-values. The quoted percentiles are computed from the resampling distributions of the average value of the annual slope parameters β_y.

delay	AIC	ΔAIC	β	SE(β)	P($>\|z\|$)	2.5%	5%	50%	95%	97.5%
0	370.98	-1.43	1.03	1.361	0.447	-1.56	-1.13	1.01	3.42	3.95
1	367.51	-0.98	1.37	1.352	0.311	-2.34	-1.81	0.24	2.53	2.92
2	366.73	1.58	2.57	1.352	0.057	-3.80	-3.17	0.41	3.95	4.45
3	366.50	1.84	2.67	1.356	0.049	-4.14	-3.50	0.11	3.62	4.11
4	369.00	-0.81	1.49	1.358	0.273	-4.97	-4.43	-0.64	2.84	3.52
5	369.42	-1.85	0.52	1.354	0.699	-4.20	-3.51	-0.31	2.70	3.23
6	369.03	-1.83	-0.56	1.354	0.678	-3.12	-2.50	-0.20	1.78	2.15
7	368.65	-0.41	-1.70	1.361	0.210	-4.10	-3.68	-1.25	0.70	1.02
8	364.90	2.08	-2.74	1.375	0.047	-6.47	-5.86	-2.03	1.47	2.08
9	364.69	3.49	-3.21	1.397	0.022	-6.97	-6.30	-2.71	1.21	1.76
10	367.61	-1.06	-1.31	1.363	0.335	-4.89	-4.30	-0.50	3.10	3.82
11	371.15	-2.00	-0.03	1.332	0.981	-3.59	-2.96	0.23	2.63	3.19

References

K. P. Burnham and D. R. Anderson. Multimodel inference: understanding aic and bic in model selection. *Sociological Methods and Research*, 33:261–304, 2004.

B. Dost, F. Goutbeek, D. T. van Eck, and Kraaijpoel. *Monitoring induced seismicity in the North of the Netherlands: status report 2010*. KNMI Scientific report: WR 2012-03, 2012.

Hastie and Tibshirani. *Generalized Additive Models*. Chapman and Hall, 1990.

P. McCullagh and J.A. Nelder. *Generalized Linear Modeling*. London: Chapman and Hall, 1989.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org/.

S. N. Wood. *Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC*. London: Chapman and Hall/CRC, 2006.

S.N. Wood. Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society (B)*, 62:412–428, 2000.

S.N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semi-parametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73:3–36, 2011.

Bibliographic information

Classification	Restricted
Report Number	SR.15.13132
Title	Statistical methodology for investigating seasonal variation in rates of earthquake occurrence in the Groningen field
Subtitle	
Authors	S. Bierman (GSNL PTD/TASE) R. Paleja (GSUK PTD/TASE) M. Jones (GSUK PTD/TASE)
Keywords	Seismicity, Groningen, Seasonality, Statistical analysis
Issue Date	October 2015
Period of work	April - August 2015
US Export Control	Non US — Public Domain
WBSE Code	ZZPT/015656/010125
Reviewed by	Randell, David GSUK-PTD/TASE
Approved by	Jonathan, Philip GSUK-PTD/TASE
Sponsoring Company/ Customer	Nederlandse Aardolie Maatschappij
Spons./Cust. Address	
Issuing Company	Shell Global Solutions International B.V., Amsterdam P.O. Box 38000 1030 BN Amsterdam The Netherlands

Report distribution

Electronic distribution (PDF)

Name, Company, Ref. Ind.	
PT Information Services, PTT/TIKE, PT-Information-Services@Shell.com	PDF
Jonathan Philip GSUK-PTD/TASE	PDF
Gibbs Rick, GSUSI-PTD/TASE	PDF
Bierman Stijn, GSUK-PTD/TASE	PDF
Greenall Carol, GSUK, PTD/FTCT	PDF
Maron Karel, P NAM-UIO/T/DGR	PDF
van Elk Jan, NAM-UIO/T/DGR	PDF
Bourne Stephen, GSNL-PTI/RC	PDF
Harris Christopher, GSNL-PTI/RC	PDF
Wentinck Rick, GSNL-PTI/RC	PDF
Paleja Rakesh, GSUK-PTD/TASE	PDF
Kraaijeveld Famke, GSNL-PTI/RC	PDF
Park Timothy, GSUK-PTD/TASE	PDF
Randell David, GSUK-PTD/TASE	PDF
van den Bogert Peter, AJ GSNL-PTU/E/Q	PDF
Beckers Janwillem, NAM-UIO/T/D	PDF
Wenham Maggie, SEPL PTD/TASE	PDF
Das Saptarshi, PTIN-PTI/CT	PDF
van Beest Bertwim, PTIN-PTI/CT	PDF
Reppart William, GSUSI-PTD/TASE	PDF

Principal typesetting performed with LaTeX system (MiKTeX) using the PTreport2015.cls class based on the KOMA-Script with the necessary modifications to match the Shell Projects and Technology house style.

The PTreport2015.cls class file is written and maintained by IdlTeX (www.idltex.com). All rights reserved.

Seasonal variation in rates of earthquake occurrences in the Groningen field

Unrestricted

SR.17.00811

October 2018

Unrestricted **SR.17.00811**

Seasonal variation in rates of earthquake occurrences in the Groningen field

by

S.M. Bierman (GSNL-PTI/CA)

Shell Global Solutions International B.V., Amsterdam

Further electronic copies can be obtained from the Global Information Centre.

Executive Summary

Statistical methodology is applied to test for evidence of within-year seasonal variation in rates of earthquake occurrences in the Groningen gas field, and for evidence of a relationship between seasonal variation in gas production and event rates. In earlier reports on the same topic (Bierman et al. [2015], Bierman et al. [2016]) it is concluded that there is evidence of seasonal variation in rates of occurrences of events with small magnitudes, e.g. events with magnitudes $M \leq 1$ (well below the magnitude of completeness of $M = 1.5$). It is unclear what the underlying cause(s) of these seasonal trends are. There is a possibility that this is partly or wholly caused by variability in probabilities that earthquakes with small event magnitudes were detected and recorded in the catalogue. For events with associated magnitudes above the magnitude of completeness, no evidence was found of seasonality in earthquake occurrence rates.

In this report additional methods (compared to those used in previous reports) are applied to test for seasonality. Also, the following recent developments warrant particular attention:

- Since 2014, annual gas production rates as well as seasonal fluctuations in gas production rates have been greatly reduced.
- In recent years, the detection capabilities of the geophone network have improved (see above).

Our findings are as follows:

1. There is strong evidence of diurnal and annual (seasonal) periodicity in occurrence times of events with associated magnitudes $M \leq 1$, and some evidence of annual periodicity for events $M \geq 1$.
2. There is no evidence of any type of periodicity for events with magnitudes $M \geq 1.5$.
3. If data from the post-January 2014 epoch are analysed separately, there is no longer any evidence of any periodicity for events regardless of their magnitude.

The absence of evidence of seasonality in occurrence rates in the post-January 2014 epoch for events with small associated magnitudes ($M \leq 1$) may be explained by:

- The low power of the statistical test to detect periodicities, due to the small size of the catalogue in the post-January 2014 epoch. However, because of improved capability of the geophone network similar numbers of events with $M \leq 1$ are detected in both epochs.
- A lower magnitude of completeness, due to improvements in the geophone network (under the hypothesis that the underlying cause was seasonal changes in event detection rates).
- The less pronounced monthly variations in gas production rates (under the hypothesis that the underlying cause was seasonal variation in gas production rates).
- A combination of the above.

The apparent disappearance of diurnality in occurrence rates of events with small magnitudes is most straightforwardly explained by improvements in the geophone network.

Amsterdam, October 2018.

Table of Contents

List of Figures

1. Introduction

In this report, we describe and apply statistical methodology to test for evidence of within-year variation (seasonality) in rates of earthquake events associated with the Groningen gas field, and for evidence of a relationship between seasonal (within-year; e.g. monthly) variation in gas production and event rates. This report is the latest in a series of reports on the same topic (Bierman et al. [2015], Bierman et al. [2016]). For this latest report we have used the data as recorded in the Earthquake Catalogue (EC) of events by the Royal Dutch Meteorological Institute (KNMI). The data were obtained from the KNMI web-pages at the following web-address: http://cdn.knmi.nl/knmi/map/page/seismologie/all_induced.json. The EC that was used in this study was complete up to and including the event with associated magnitude $M = 0.4$ as recorded on the 22^{nd} of July 2017 near the town of Noordbroek.

In Bierman et al. [2015] it is concluded that there is evidence that catalogue rates of events with magnitudes $M \leq 1$ vary seasonally. It is unclear what the underlying cause(s) of this seasonal trend are. In particular, there is a possibility that this is partly or wholly caused by variability in probabilities that earthquakes with small event magnitudes were detected and recorded in the catalogue. For events with $M \geq 1.5$, no evidence was found of seasonality nor of correlations between seasonal variation in gas production rates and earthquake event rates.

In this report, we:

- Describe and apply statistical methodology that can be used to test for evidence of trends in the magnitude of completeness M_c.
- Describe and apply statistical methodology that can be used to test for evidence of seasonality in the event rates, and (possibly lagged) correlations between gas production and event occurrence rates. In addition to the methods as described in the previous reports, we apply the "Schuster spectrum method" (Ader and Avouac [2013]) to the Groningen earthquake catalogue to test for evidence of the existence periodicities in event occurrence rates.

In the context of testing for evidence of seasonality of earthquake occurrence rates, the following recent developments warrant particular attention:

- Since 2014, annual gas production rates as well as seasonal fluctuations in gas production rates have been greatly reduced.
- In recent years, the detection capabilities of the geophone network have improved (see above).

We therefore pay particular attention to potential recent (following 2014) changes over time in the time-trends of event occurrences. Under the hypothesis that seasonality in occurrence rates of events $M \leq 1$ is caused by seasonality in detection capabilities of the geophone network, we may expect that seasonality would be less pronounced post 2014 due to improvements of the geophone network. However, the same pattern is expected under the hypothesis that seasonality in gas production was the main driver due to the changes in the way the field was produced. If there is evidence of seasonality in event occurrence rates post 2014, this would provide evidence against both hypotheses, and we would have to consider the possibility that there is another underlying cause.

The "magnitude of completeness", (M_c, is defined as the lowest magnitude at which 100% of the earthquakes in a space-time volume are detected (see e.g. Mignan and Woessner [2012]). In this

report we have assumed, in line with advice from the Dutch Meteorological Society (KNMI), that $M_c = 1.5$ for the time-series under consideration (Dost et al. [2012]). We use the terminology "earthquake count" for counts of events with magnitudes M $\geq$ 1.5. Instead, the term "catalogue count" is used when events with associated magnitudes M<1.5 have been used.

2. Earthquake catalogue

The KNMI catalogue of induced earthquakes contains events for the whole of The Netherlands. Events within the outline of the Groningen reservoir with an additional spatial buffer of 1000 m (figure 2.1) were defined to have occurred "within the Groningen field" in all analyses. The additional buffer of 1000 m was used to allow for errors in locating of epicenters of events and the possibility that events outside the Groningen field area were triggered by events inside the field. All events that were classified as "induced earthquakes" by KNMI with associated epicenters with the X-coordinate (easting) X > 200000 on the "Rijksdriehoeksstelsel" (RD) coordinate reference system were included in the analysis. The choice of width of the buffer zone does not influence the outcomes of the analyses much because the vast majority of events occurred within the boundaries of the Groningen field.

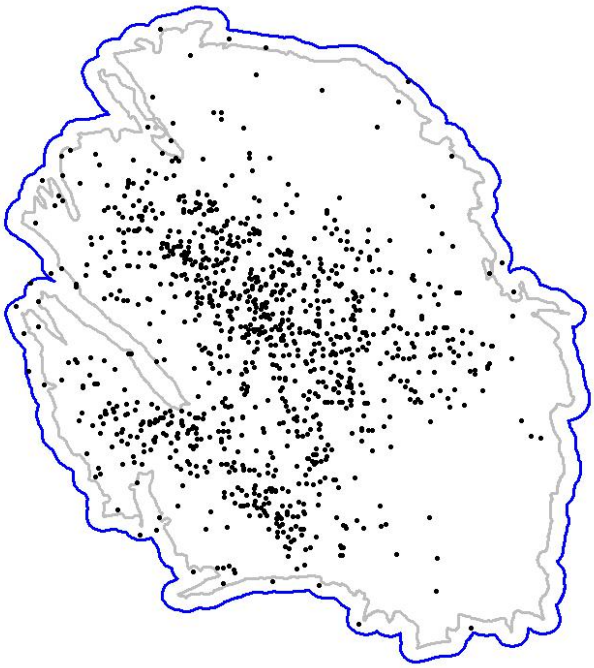

Figure 2.1.: Map of the outline of the Groningen reservoir (inner grey line) and an additional buffer with a width of 1000 meters (outer blue line).

Earthquakes in the Groningen field are believed to partly occur in clusters in time and space, in the form of aftershocks (see e.g. Touati et al. [2009]). In this report we perform analyses on the raw data including all counts as well as on a subset of the data (referred to hereafter as declustered) in which we have excluded events that occurred within 3 days and 2500 m of a previous event with magnitude $M \geq 2$. Of all events inside the field boundary a total of 21 events (1.9%) were classed as potential aftershocks. We note that this particular choice of definition of potential

aftershock is arbitrary. There may in reality have been many more aftershocks than the 21 events identified here. The choice of aftershock declustering algorithm is most influential in uncertainty analyses: the uncertainty bounds as presented in this report are likely to be "over-optimistic" (too narrow) because they do not appropriately reflect the uncertainties surrounding the identification of putative aftershocks. In particular, in some of the data visualisations presented in this report we use the Poisson discrete probability density function to estimate the expected variability in counts across calendar months given the average monthly count. In the Poisson distribution it is assumed that the variance in counts is equal to the expected count. In practice it is commonly seen that the variability in counts is larger than the expectation, a phenomenon that is commonly referred to as overdispersion. Ignoring potential overdispersion is not likely to bias estimates of rate estimates (average monthly counts), but may lead to over-optimistic estimates of standard errors of these parameters (variation in monthly counts).

The locations of epicenters of events within different ranges of magnitudes are depicted in figure 2.2. There are no obvious differences in the spatial distribution or extent of the epicenters of the events with different magnitudes.

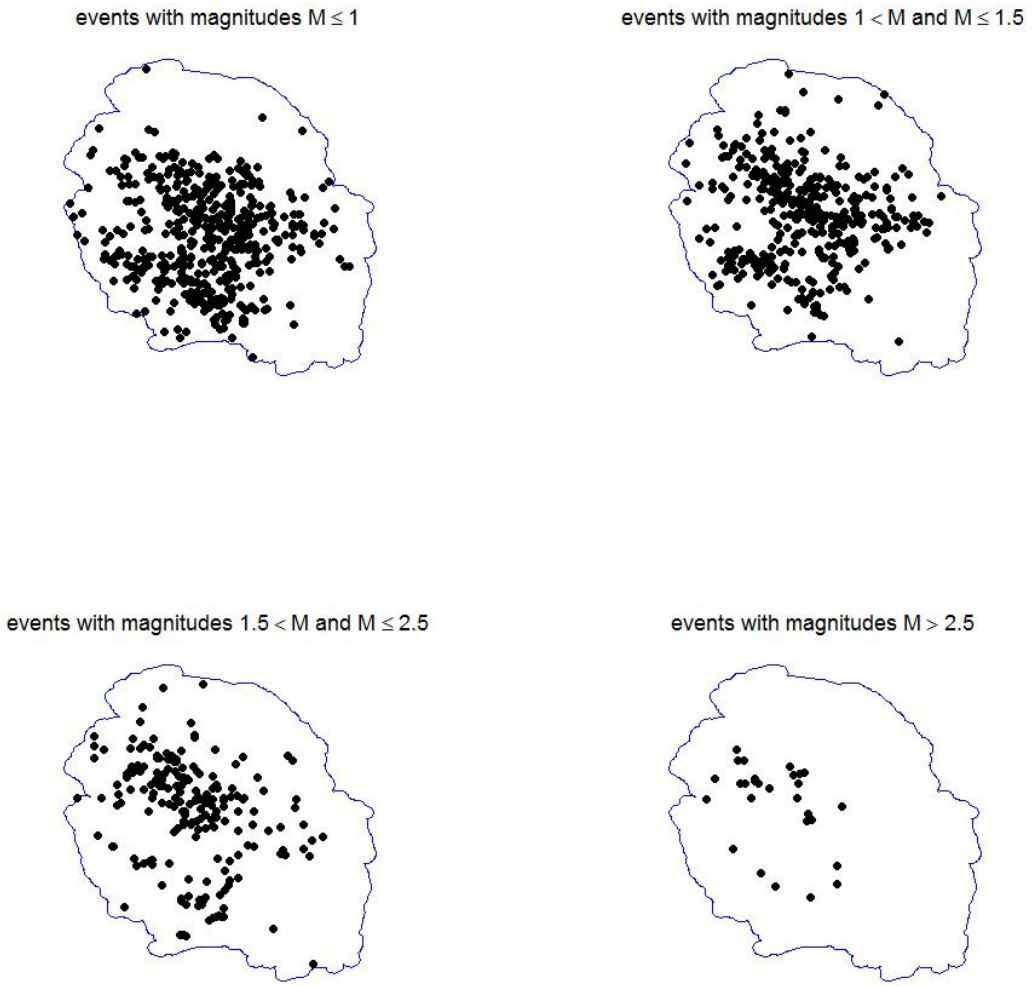

Figure 2.2.: Maps of epicenters of events in different of ranges of event magnitudes.

Time series of counts of events per calendar month for all events, and events with magnitudes $M \leq 1$, $1 < M < 1.5$, and $M \geq 1.5$ are given in figure A.1.

3. Seasonality in event rates: data visualisation

A visualisation of the catalogue of events (of all magnitudes) associated with the Groningen gas field suggests that catalogue rates may vary seasonally and may, possibly with some time-delay, be strongly correlated with the seasonal pattern in production rates (figure 3.1). The data visualisation is based on a moving average of counts of Catalogue Events, resulting in a temporally smooth trend in Catalogue rates. The temporally smooth trend in event rates is plotted alongside a time series of monthly gas production data (field wide). The moving average of counts of Catalogue Events is calculated using a "sliding time-window" approach, where each time-window spans three calendar months and the average of the counts of events in the three months is plotted on the graph. The time-windows are applied to each month in the time-series (incrementally).

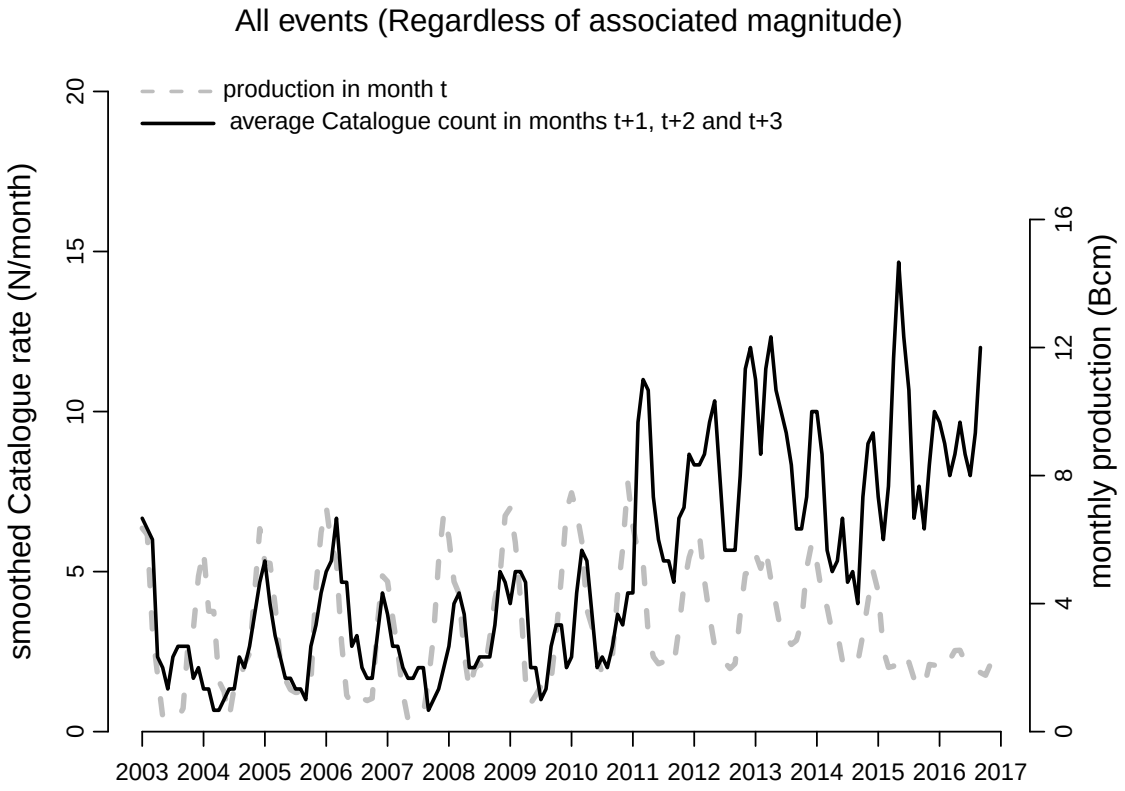

Figure 3.1.: Monthly field-wide gas production (grey dashed line) and smoothed Catalogue Events rates (black solid line). The smoothed Catalogue Event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months.

While figure 3.1 suggests that there is a correlation between gas production and catalogue rates we note that any other variable which fluctuates seasonally within each year, such as ambient temperature, would also correlate with seasonally varying event rates. Furthermore, care is required with the interpretation of moving averages since each earthquake is used three times in the analysis (except for events in the first two and last two months in the time series).

In the period January 2003 - January 2014, gas production was strongly seasonal (figure 3.2; top left panel). During this period, catalogue rates of events with magnitudes $M \leq 1$ appear to vary seasonally with, for most calendar years, particularly high counts in April, May and June (figure 3.2; top right panel). For events $M \geq 1$ there is no strong indication that event rates differ

between months. In the period January January 2014 - January 2017, gas production was only weakly seasonal (figure 3.2; bottom left panel). There was no indication that underlying rates differ between calendar months (figure 3.2; bottom row).

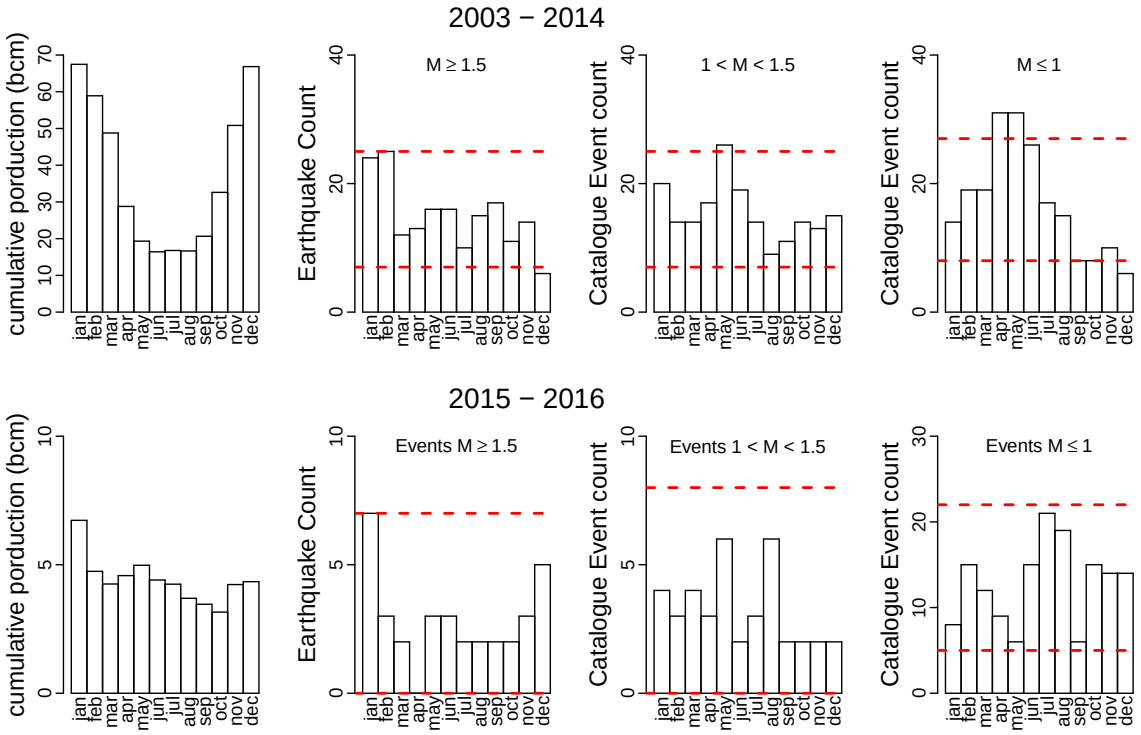

Figure 3.2.: Counts of events per calendar month, summed over all years, for events with $M \leq 1$. The red dotted lines indicate 95% prediction intervals for the counts. These prediction bounds are based on the assumption of a stationary Poisson process with mean event rate equal to the overall mean.

Time series of catalogue counts of events per calendar month and monthly field-wide gas production are given in figure 3.3 for events M $\leq$ 1, in figure 3.4 for events 1 $<$ M $<$ 1.5 and in figure 3.5 for events M $\geq$ 1.5. These figures provide a more direct representation of the available information than figure 3.1 because no temporal smoothing is used and each event occurs exactly once in the analysis. If the same procedure for smoothing using a time-window of 3 calendar months is applied, then more-or-less regular seasonal fluctuations in catalogue rates are visible for events M $\leq$ 1 (figure 3.6). For events with magnitudes 1 $<$ M $<$ 1.5 there is also an indication of regular seasonal fluctuations (figure 3.7), whereas there are no apparent regularly seasonal fluctuations for earthquakes with M $\geq$ 1.5 (figure 3.8).

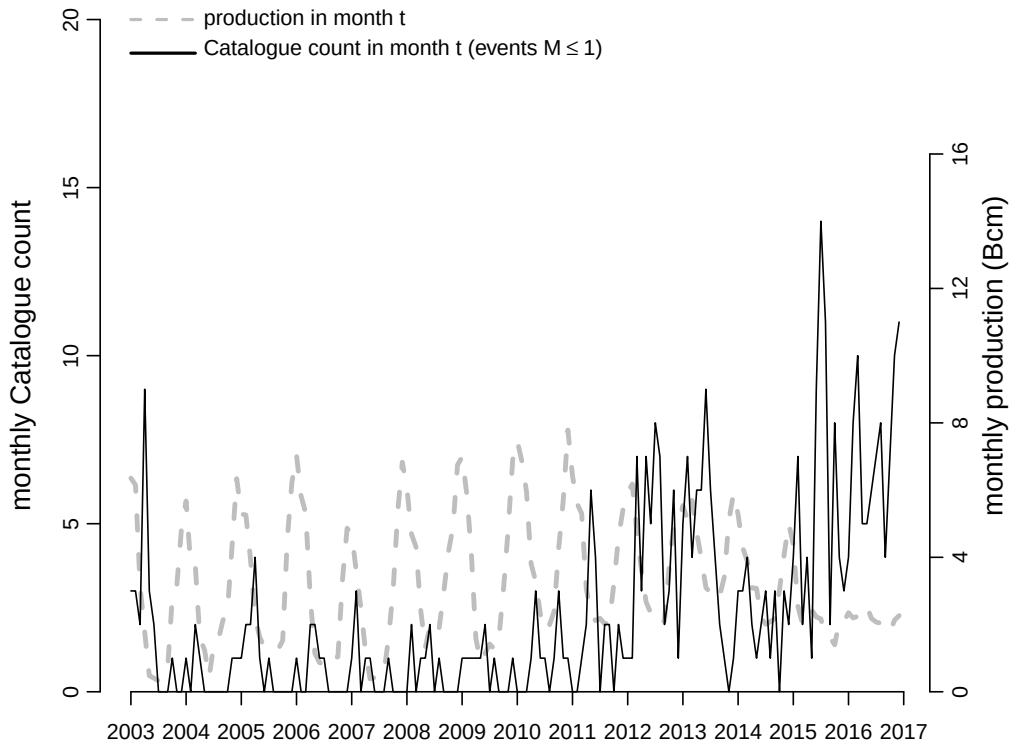

Figure 3.3.: Field wide monthly gas production and monthly catalogue counts of events inside the field boundary with associated magnitudes $M \leq 1$ (all events with $M \leq 1$ or with exclusion of events that occurred within 3 days and 2500 m of a previous event (declustered)).

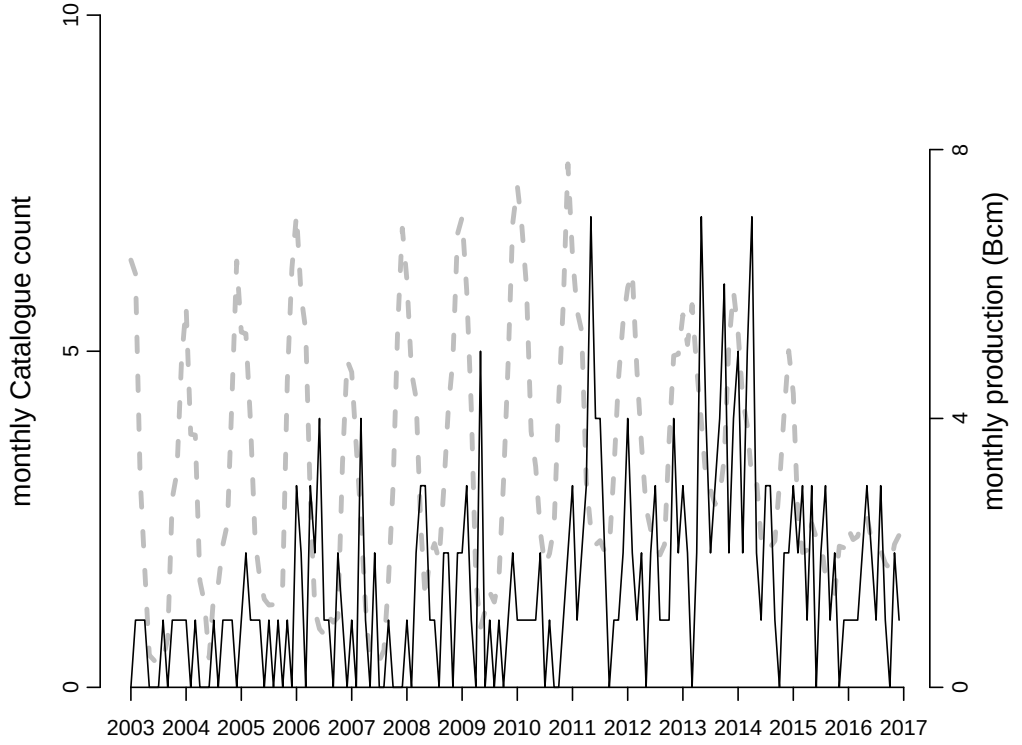

Figure 3.4.: Field wide monthly gas production and monthly counts of earthquakes inside the field boundary with associated magnitudes $M \geq 1.5$ (all events with $1 < M < 1.5$ or with exclusion of events that occurred within 3 days and 2500 m of a previous event (declustered)).

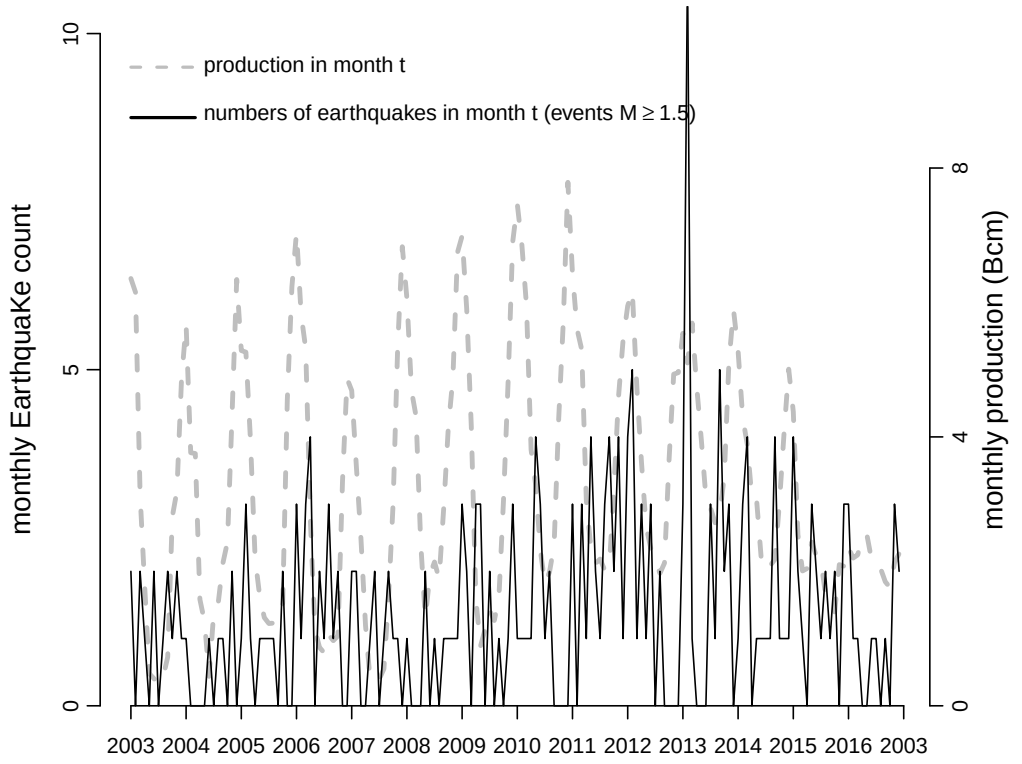

Figure 3.5.: Field wide monthly gas production and monthly counts of earthquakes inside the field boundary with associated magnitudes M ≥ 1.5 (all events with M ≥ 1.5 or with exclusion of events that occurred within 3 days and 2500 m of a previous event (declustered)).

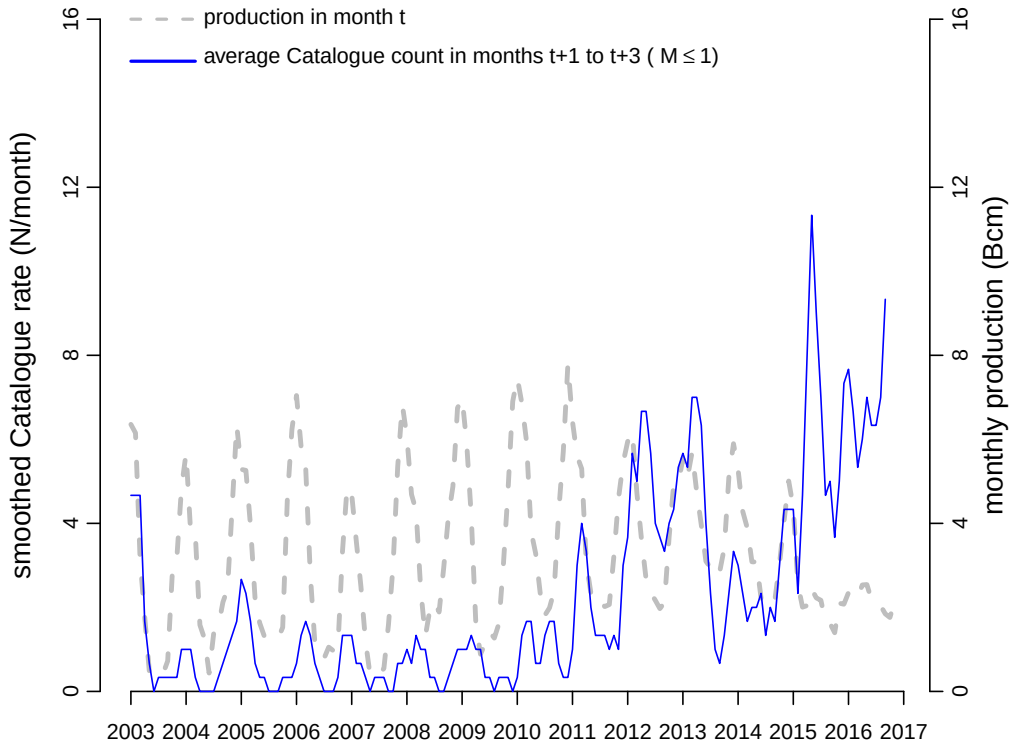

Figure 3.6.: Monthly field-wide gas production (grey dashed line) and smoothed catalogue rates for M ≤ 1 (blue line). The smoothed event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months and the averages of the counts of events in the three months are plotted on the graph and connected by lines.

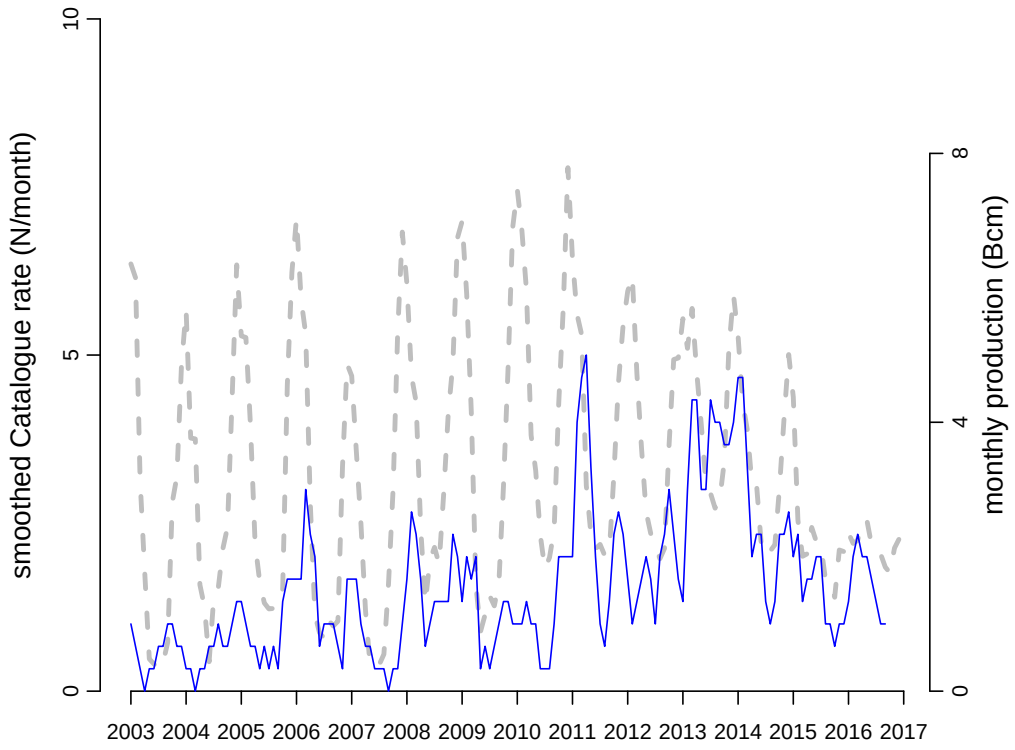

Figure 3.7.: Monthly field-wide gas production (grey dashed line) and smoothed catalogue rates for $1 < M < 1.5$ (blue line). The smoothed event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months and the averages of the counts of events in the three months are plotted on the graph and connected by lines.

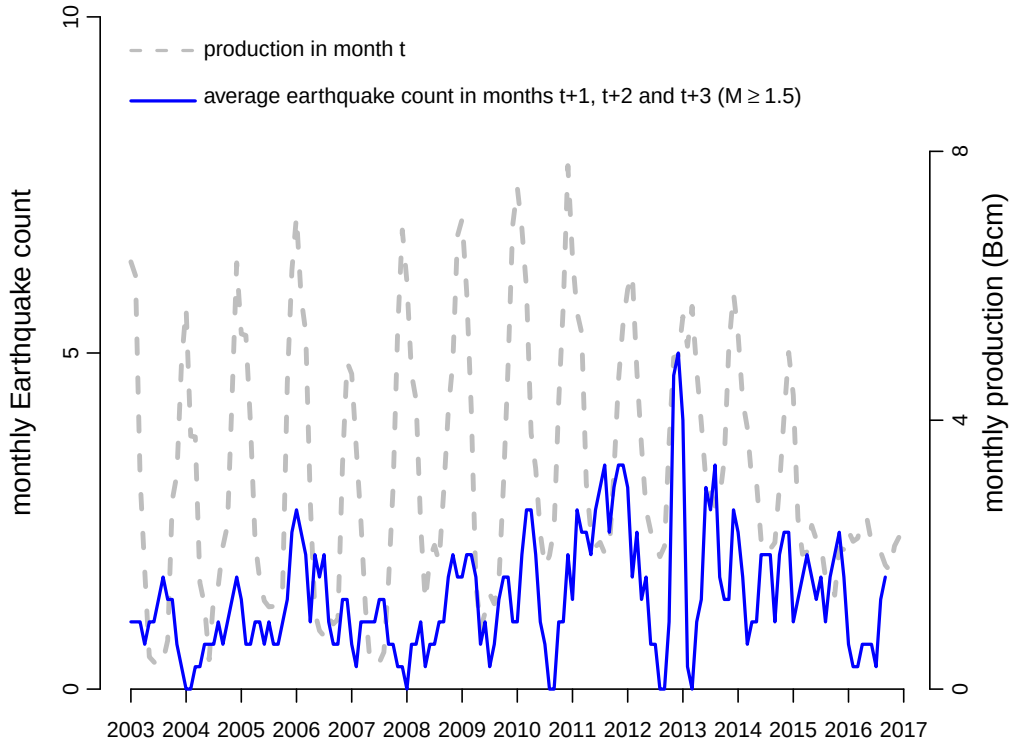

Figure 3.8.: Monthly field-wide gas production (grey dashed line) and smoothed catalogue rates for M ≥ 1.5 (blue line). The smoothed event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months and the averages of the counts of events in the three months are plotted on the graph and connected by lines.

4. Trends in the magnitude of completeness

Here, we apply empirical catalogue-based methods to estimate time-trends in M_c. A good overview of methods to estimate M_c, and their assumptions and weaknesses, is given in Mignan and Woessner [2012]. We note that there is not a single method which will yield a guaranteed unbiased estimate. The Groningen earthquake catalogue is sparse: the numbers of event occurrences per year are relatively low. Catalogue-based estimates of M_c are therefore surrounded with much uncertainty, in particular if these are broken down by epoch.

4.1. Trends in binomial proportions

Time-trends in binomial proportions are used as a descriptive method to visualise changes over time in M_c. Let M_s be a conservative (i.e. "high-end") estimate of M_c. Here, we take $M_s = 1.6$. Let M_p ($M_p < M_s$) be a *proposed* (hypothesised) less conservative value to be used as an estimate for M_c. The number of events with associated magnitudes equal to or larger than M_s is denoted by K_s. The number of events with associated magnitudes equal to or larger than M_p is denoted by K_p ($K_p \geq K_s$). Observed proportions $p_t = K_s(t)/K_p(t)$ for epoch t may be modelled as a function of time using the binomial distribution and a non-parametric smoothing spline for the underlying proportion (using a logit-link function). We used smooth functions represented using penalized regression splines. The amount of smoothness is estimated using generalised cross-validation. This model is implemented using the functionality for generalised additive modeling (gam; see e.g. Hastie and Tibshirani [1990]) in R (Wood [2006], Wood [2011]). In the above-mentioned implementation of gams in the R software (R Core Team [2014]), approximate multivariate posterior distributions of the fitted parameters are obtained by assuming a multivariate normal prior distribution and by using the derivatives of the likelihood with respect to the fitted parameters at convergence of the algorithm (see Wood [2000]).

An estimate of the expected proportion $\hat{p}$ if $M_p \geq M_c$, can be obtained based on the Gutenberg-Richter frequency-magnitude curve. For the Groningen catalogue, we expect values of $\hat{p}$ to be approximately between the proportions as expected given Gutenberg-Richter curves with b-values $b = 0.7$ and $b = 1.2$ (Harris [2015]).

In line with the advice from KNMI (Dost et al. [2012]), there are no obvious indications of time-trends in M_c for $M_p = 1.4$ or $M_p = 1.5$ (figure 4.1). For lower hypothesised values for the magnitude of completeness there are clear time-trends which reflect the gradual improvements in the capabilities of the geophone monitoring network. It may be reasonable to assume that the magnitude of completeness is lower than $M_p = 1$ for recent years, and potentially even considerably lower in the past year or two (figure 4.1).

We note that we have here used only two exceedance probabilities to characterise the frequency magnitude distribution. A more formal and possibly better analysis which uses the available data in a more efficient way would be to model the entire frequency-magnitude distribution, using either the Gutenberg-Richter curve or a peaks-over-threshold model such as the Generalised Pareto distribution for threshold exceedances (see e.g. Randell et al. [2015]).

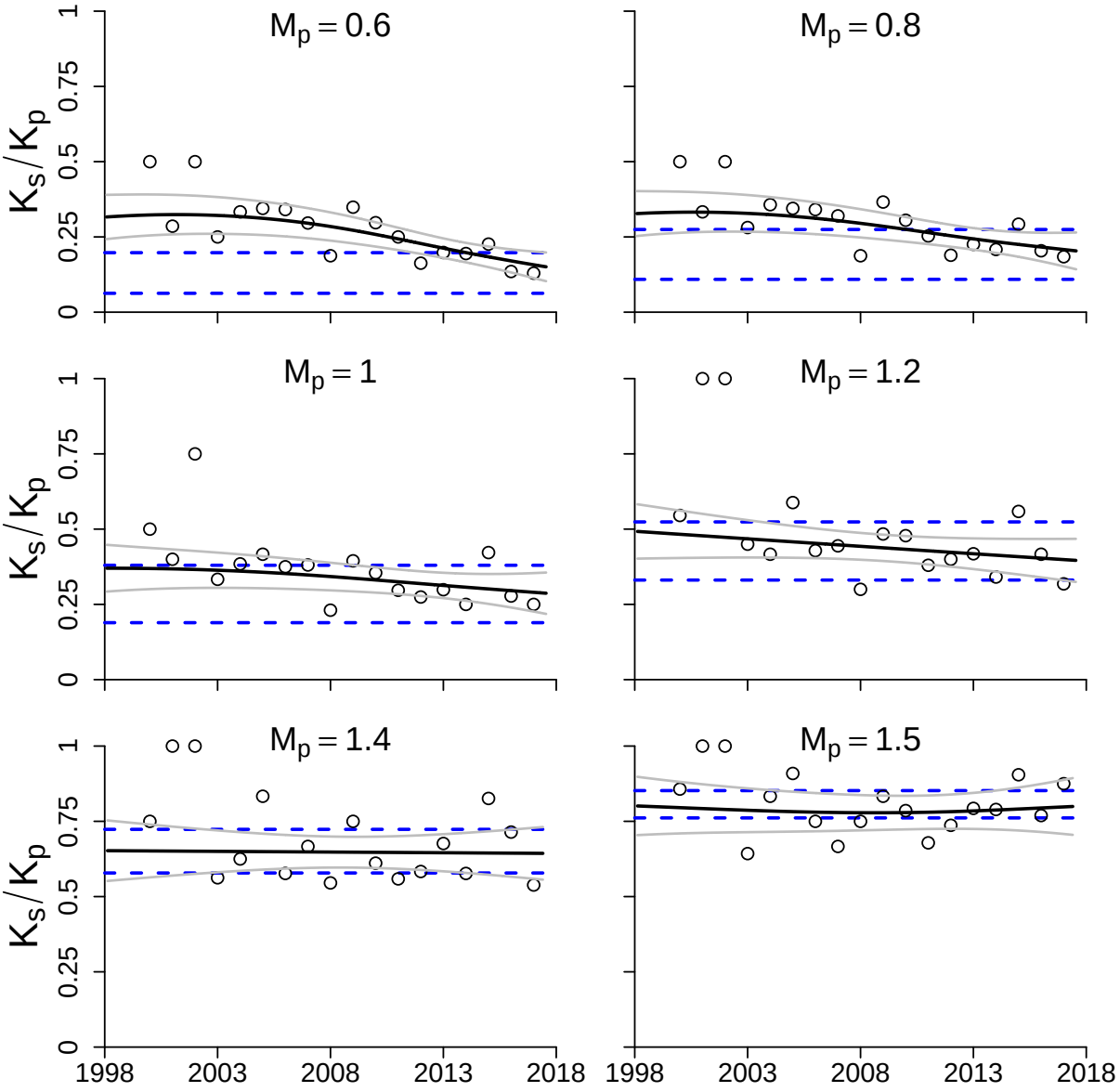

Figure 4.1.: Time-trends in numbers of events, K_p, with magnitudes above a hypothesised $M_c = M_p$, relative to numbers of events, $K_s < K_p$ with magnitudes above a conservative ("high-end") estimate of $M_c = M_s$. The open circles are annual observed proportions $p = K_s/K_p$ (see text). The black solid line gives the proportion (grey lines: 95% confidence interval of the mean) as predicted by the smoothing splines (assuming binomial proportions). The blue dashed lines indicate the expected proportions given Gutenberg-Richter curves with b-values $b = 0.7$ and $b = 1.2$.

4.2. The maximum curvature methods applied to epochs

In the Maximum Curvature (MAXC; Wiemer and Wyss [2000]) method, M_c is estimated as the maximum value of the first derivative of the magnitude-frequency (Gutenberg-Richter) curve. The MAXC method relies on the validity of the Gutenberg-Richter law (see page 14 in Mignan and Woessner [2012]). Here, the MAXC method is applied to epochs of 2 years each. The uncertainty surrounding the estimates of M_c is estimated by random resampling with replacement (bootstrap).

There are clear indications that M_c has decreased over time (figure 4.2), but the uncertainty surrounding the estimates is large due to the small number of events in each epoch. A value of $M_c = 1.5$ is certainly reasonable throughout the time-series, and a value of $M_c = 1$ appears reasonable since 2015.

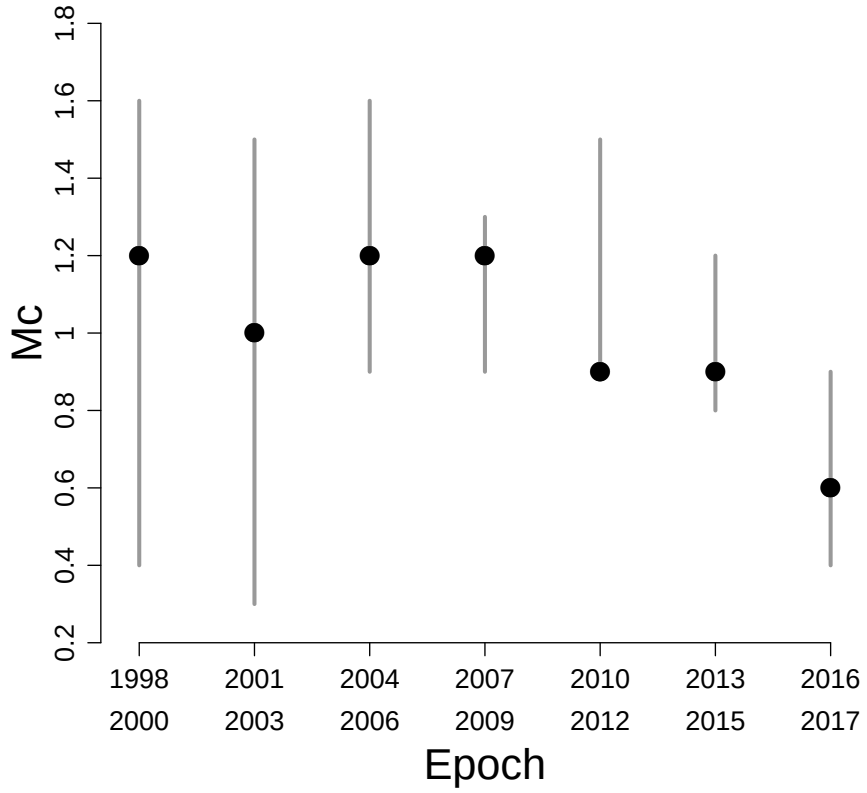

Figure 4.2.: Estimates (filled points) and 95% confidence intervals (grey vertical bars) of M_c obtained using the MAXC method, for different epochs.

4.3. Diurnality in event occurrence rates

As described in Bierman et al. [2015], a peculiar aspect of events with relatively low associated magnitudes is that the rate at which they occur in the catalogue appears to vary diurnally, with higher rates of events between approximately 20:00 in the evening and 04:00 in the morning (figure 4.3). Such a diurnal pattern is not immediately apparent for events with magnitudes M > 1. A hypothesis for the diurnal variation in counts of event occurrences is that this is caused by the influence of cultural noise on the detection of events with small associated magnitudes. We compare frequency histograms of the numbers of events per hour within the day (for different magnitudes) both before and after January 2014.

Post January 2014, diurnality is no longer apparent for events with magnitudes $0.5 \leq M \leq 1$ (figure 4.4). This indicates that the diurnality in event occurrence rates was caused by diurnal fluctuations in the noise floor and therefore detection capabilities of the geophone network. The absence of evidence of diurnality in event occurrence rates post January 2014 is a further indication of the improved ability of the geophone network to detect events with associated magnitudes well below $M = 1.5$.

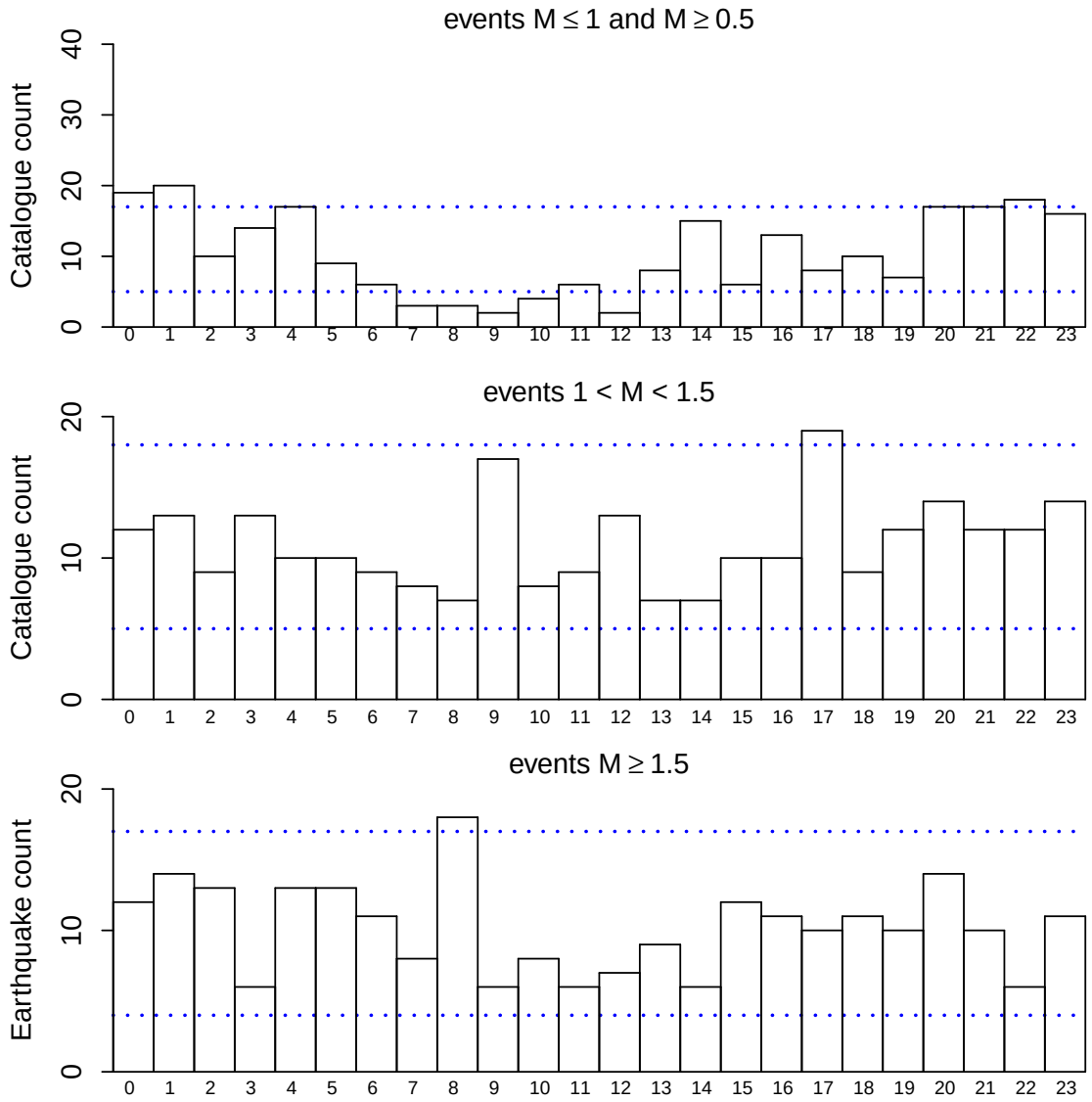

Figure 4.3.: Counts of numbers of events that occurred before 01-01-2015 for each of the 24 hours within the day (00:00 - 01:00, 01:00 - 02:00,...,23:00 - 24:00) for events in different categories of associated magnitudes. For ease of visual interpretation, 95% (blue dotted lines) confidence bounds are given. These confidence bounds are based on the assumption of a stationary Poisson process with mean event rate equal to the overall mean.

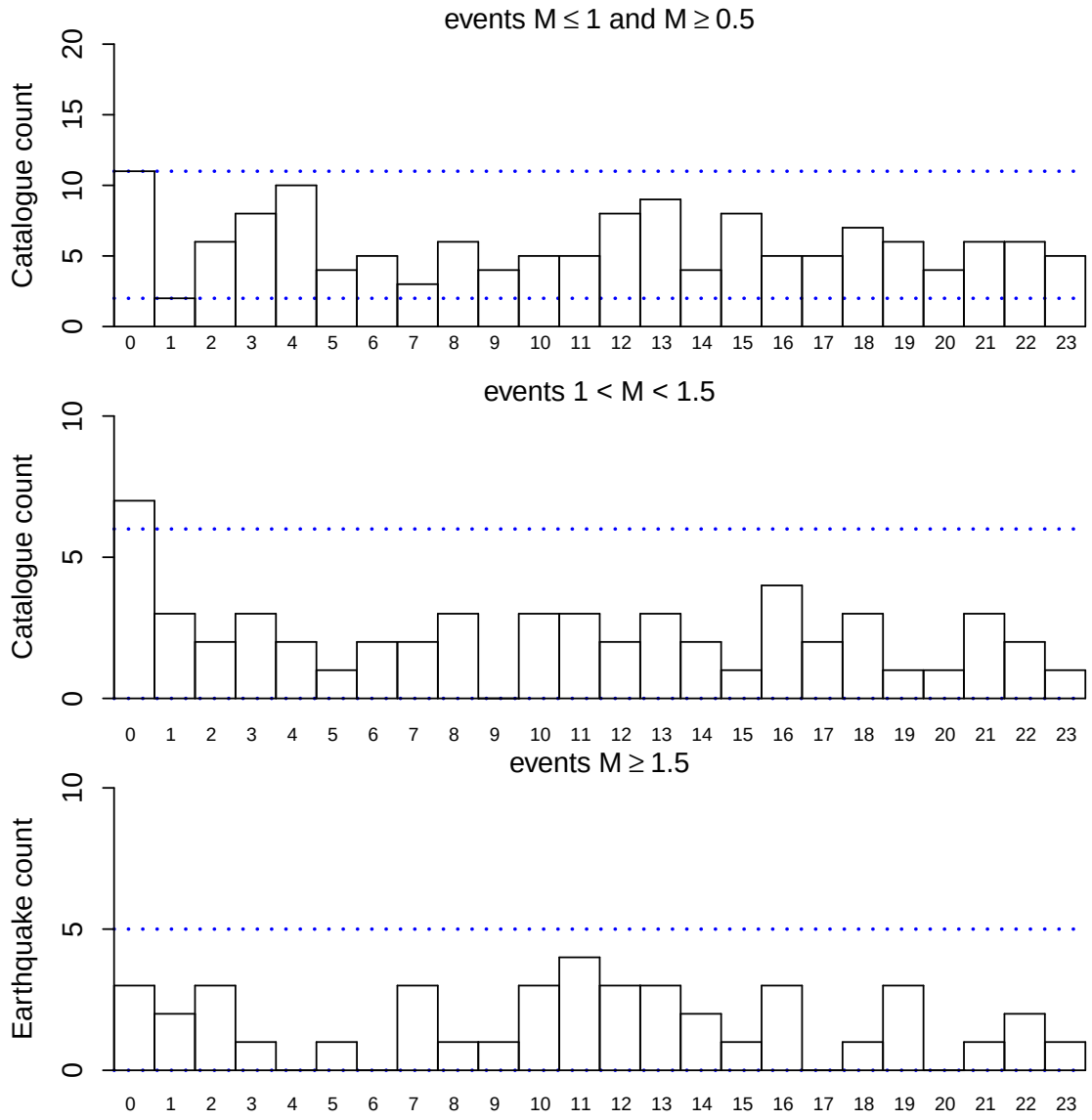

Figure 4.4.: Counts of numbers of events that occurred after 01-01-2015 for each of the 24 hours within the day (00:00 - 01:00, 01:00 - 02:00,...,23:00 - 24:00) for events in different categories of associated magnitudes. For ease of visual interpretation, 95% (blue dotted lines) confidence bounds are given. These confidence bounds are based on the assumption of a stationary Poisson process with mean event rate equal to the overall mean.

5. Correlations between time series of gas production rates and event counts

Let N_i be the number of events that occurred in "Julian month" i ($i = 1, 2, 3, ..., K$ with $K = 12 \times 12 = 168$ the total number of months from January 2003 up to an including December 2016). Let $\mathrm{m}(i)$ be the calendar month (within calendar year) of Julian month i ($\mathrm{m}(i) \in \{1, 2, ..., 12\}$). Daily field-wide gas production per month i is denoted by $P_{(\mathrm{m}(i))}$.

The Spearman rank correlation, $\rho_M = \mathrm{cor}\left\{P_{(\mathrm{m}-M)}, N_m\right\}$ can be used as a non-parametric measure of association between the vector of monthly counts of event occurrences N_m and time-lagged vector of monthly gas production $P_{(\mathrm{m}-M)}$ with lag M in months. We have used values of lags from zero (no lag) up to 11 calendar months ($M = 0, 1, 2, ..11$). There will be many identical values (ties) in the vector N_m and we note that we have used the ties-corrected Spearman rank correlation coefficient defined as the Pearson's sample correlation coefficient computed for the ranked sample data where ranks are averaged for tied values. Spearman's rank correlation provide a noteworthy alternative means of measuring the association between monthly production and monthly event counts because of it's simplicity and because:

- No assumption of linearity of the relationship between the two variables is made, because the measure is based on ranks.
- The measure is appropriate for measuring the association between a continuous variable (monthly gas production) and a discrete variable (monthly counts of event occurrences).

To test whether observed values of ρ_M were significantly different from zero (we note that $-1 \leq \rho \leq 1$), we calculated the distribution of ρ under the null hypothesis of no association between monthly production and monthly counts of event occurrences, by using a permutation test. Random synthetic data sets under the null hypothesis of no association were generated by randomly re-ordering, within each calendar year, the monthly counts of event occurrences over the calendar months within that year. For each random realisation, the Spearman rank correlation coefficient was calculated for each of the lags M. Confidence intervals were based on 5000 permutations.

For the period January 2003 - January 2014, correlation coefficients for events $M \leq 1$ were significantly different from those expected under the null hypothesis for a number of lags (Figure 5.1), with maximum positive correlation coefficients at lags of 3 and 4 months. Correlation coefficients for events with associated magnitudes $1 < M < 1.5$ were significantly different from those expected under the null hypothesis of no effect for lags of 2 and 8 months. There was insufficient evidence to reject the null hypothesis of no association between monthly production and monthly counts of event occurrences $M \geq 1.5$.

We note that correlation coefficients tended (on average) to be positive. This is a reflection of the overall increase in counts of event occurrences over the study period in combination with less pronounced monthly variation due to more gas being produced in the summer months in gas production near the end of the study period (from approximately 2011 onwards; see figure 3.1). For illustration, correlation coefficients are given for nine randomly generated synthetic data sets that were used in the permutation test for events with associated magnitudes $M \geq 1.5$ in figure 5.2. These cases illustrate that, due to the seasonally varying monthly gas production rates and the discrete nature of the monthly counts, apparent regular within-year fluctuations are expected to occur entirely by chance.

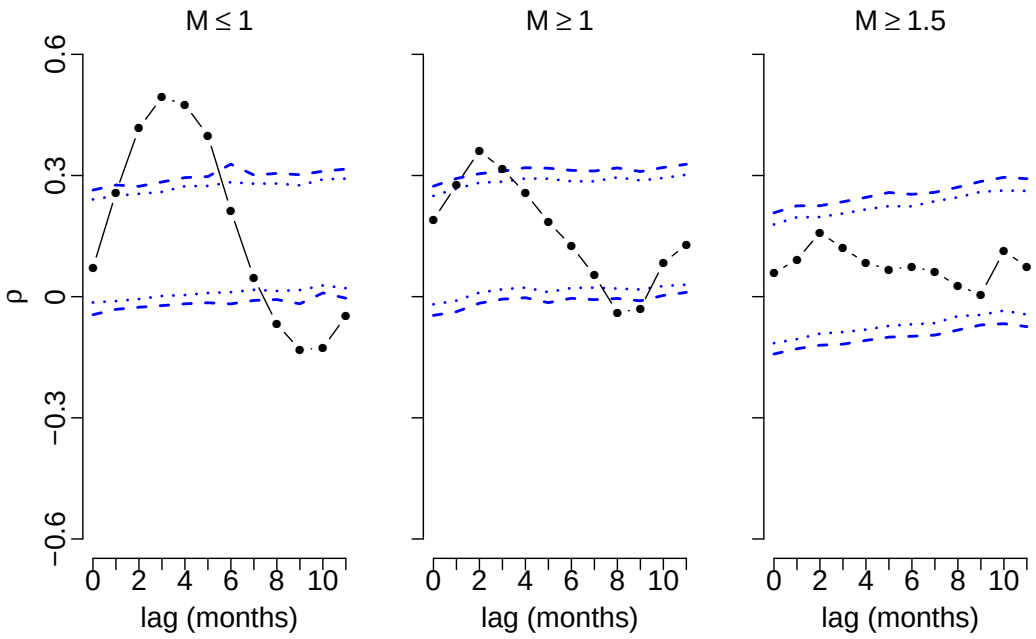

Figure 5.1.: Cross-correlations for the January 2003 - January 2014 epoch. Spearman rank correlation coefficients (ρ) as a measure of the association between monthly gas production and monthly counts of event occurrences. The correlation coefficients for each lag (in months) are given by the filled points. The dotted (inner) and dashed (outer) blue lines give the 95% and 99% quantiles respectively of the distribution of the correlation coefficients as expected under the null hypothesis of no association between monthly gas production and monthly event counts (see text).

For the period January 2014 - January 2017, there was insufficient evidence to reject the null hypothesis of no association between monthly production and monthly counts of event occurrences, regardless of the magnitude of events (figure 5.3). The power of the test to reject the null hypothesis is likely to be low due to the small number of event occurrences in this epoch.

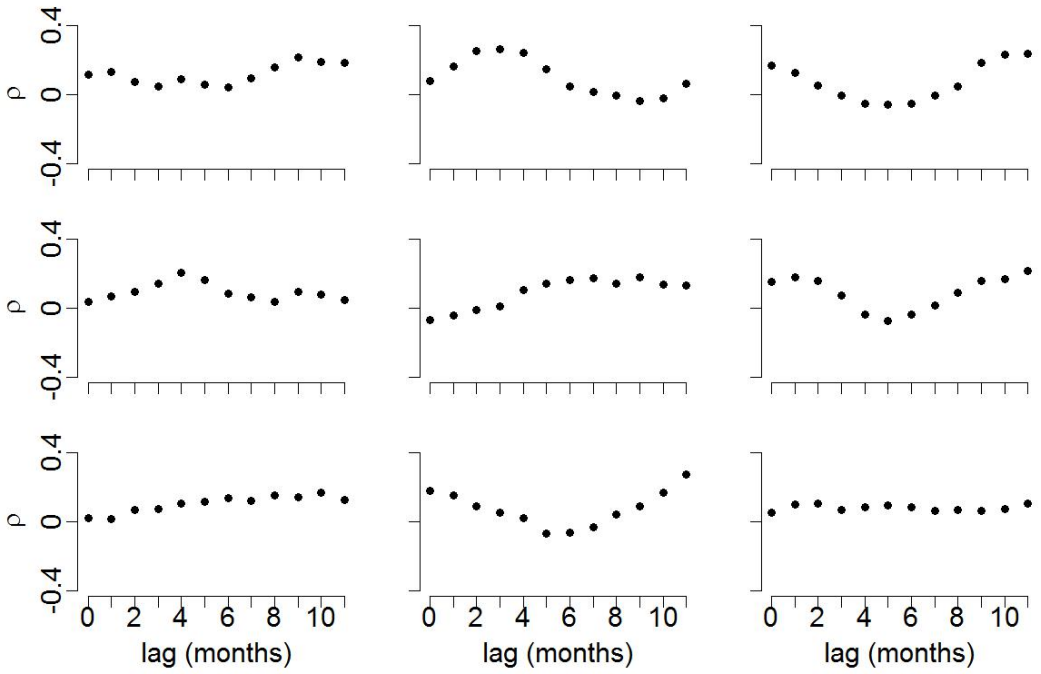

Figure 5.2.: Illustration of Spearman rank correlation coefficients (ρ) as a measure of the association between monthly gas production and monthly counts of event occurrences, for 9 (one for each panel) synthetic data sets of events with magnitudes $M \geq 1.5$. The synthetic data sets were generated by randomly re-ordering, within each calendar year, the monthly counts of event occurrences over the calendar months within that year (see text).

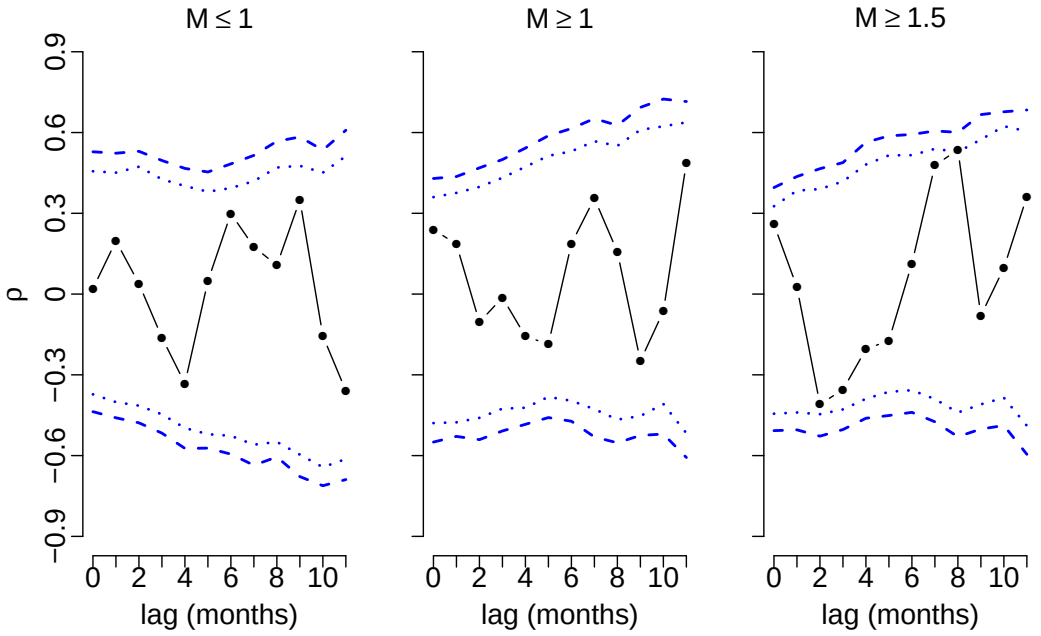

Figure 5.3.: Cross-correlations for the January 2014 - January 2017 epoch. Spearman rank correlation coefficients (ρ) as a measure of the association between monthly gas production and monthly counts of event occurrences. The correlation coefficients for each lag (in months) are given by the filled points. The dotted (inner) and dashed (outer) blue lines give the 95% and 99% quantiles respectively of the distribution of the correlation coefficients as expected under the null hypothesis of no association between monthly gas production and monthly event counts (see text).

6. Schuster spectrum test

The Schuster test (Schuster [1897]) can be used to test for evidence of the existence of a hypothesised periodicity in a catalogue of earthquake occurrence times. The Schuster test was further extended by Ader and Avouac [2013] to test for a range of periodicities by computing a spectrum of Schuster p-values. The statistical methodology as described by Ader and Avouac [2013] is suitable to test for evidence of periodicity in an earthquake catalog, both at hypothesised ("expected") and at unexpected periods. We have applied the Schuster spectrum method separately to subsets of the Groningen catalogue defined by combinations of epochs and magnitude classes:

- Epoch 1: January 2003 - January 2014, and Epoch 2: January 2014 - July 2017
- Magnitude classes: $M \geq 1.5$; $M \geq 1$; $M \leq 1$

For both epochs, for events with associated magnitudes $M \geq 1.5$, there was no evidence of any type periodicity (figure 6.1).

For the January 2003 - January 2014 epoch and for events with associated magnitudes $M \geq 1$ there was evidence of seasonality (annual periodicity; figure 6.2). This is in line with the findings as reported in Bierman et al. [2015] for this epoch. For the post January 2014 epoch there was no evidence for any type of periodicity.

For the January 2003 - January 2014 epoch and for events with associated magnitudes $M \leq 1$ there was strong evidence of diurnality and seasonality (annual periodicity; figure 6.3). This is line with the findings as reported in sections 4.3 and 3 and as reported in Bierman et al. [2015] for this epoch. For the post January 2014 epoch there was no evidence for any type of periodicity.

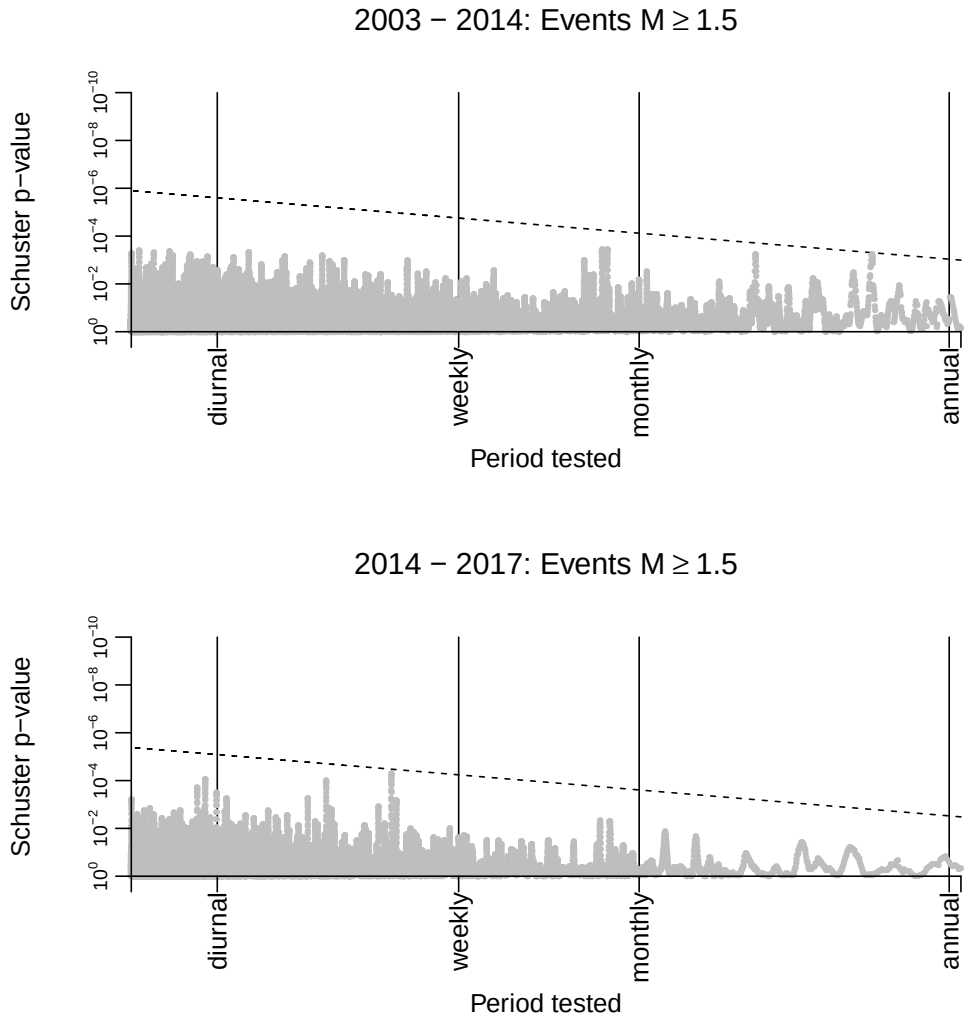

Figure 6.1.: Schuster spectrum test for events with associated magnitudes $M \geq 1.5$, for two epochs: Epoch 1: January 2003 - January 2014, and Epoch 2: January 2014 - July 2017. The diagonal dotted line delineates the critical region of the test at the 99% confidence level, All points below the line are not unexpected under the null hypothesis of no periodicity (see Ader and Avouac [2013]).

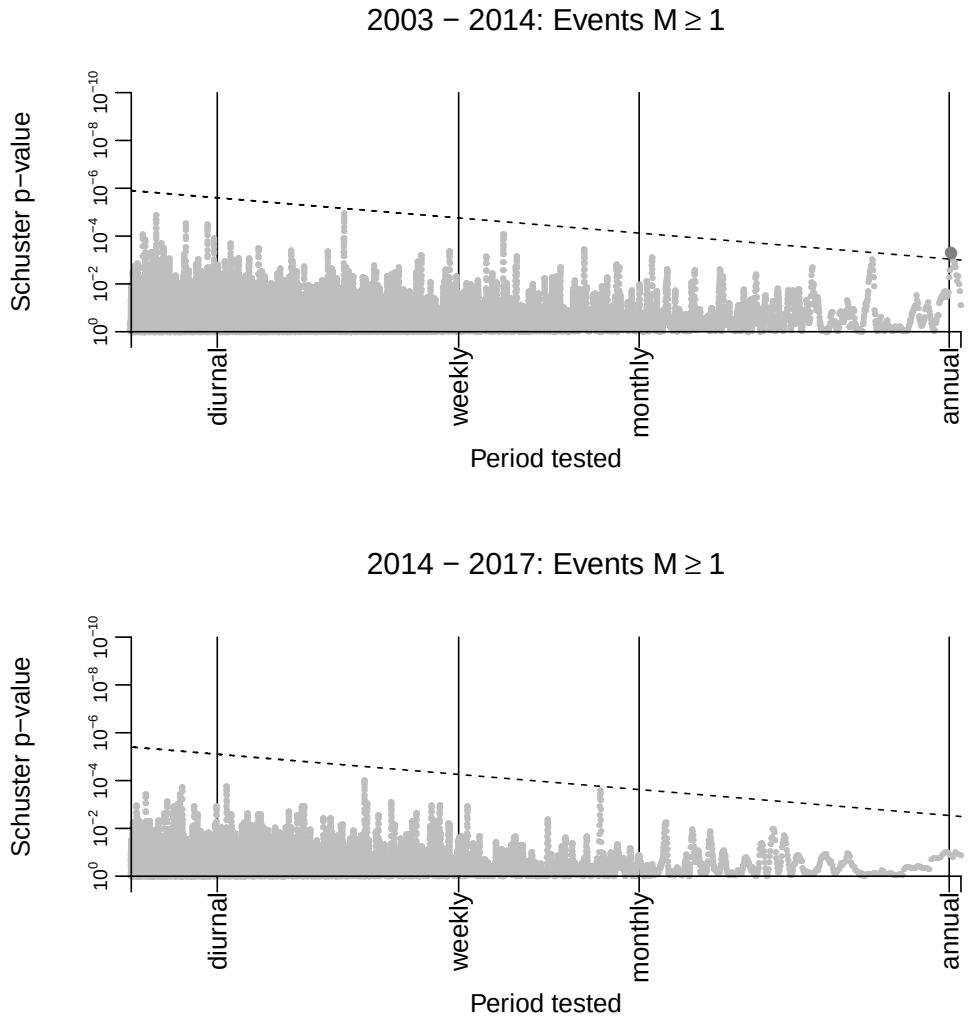

Figure 6.2.: Schuster spectrum test for events with associated magnitudes $M \geq 1$, for two epochs: Epoch 1: January 2003 - January 2014, and Epoch 2: January 2014 - July 2017. The diagonal dotted line delineates the critical region of the test at the 99% confidence level, All points below the line are not unexpected under the null hypothesis of no periodicity; points that exceed the threshold are enlarged (see Ader and Avouac [2013]).

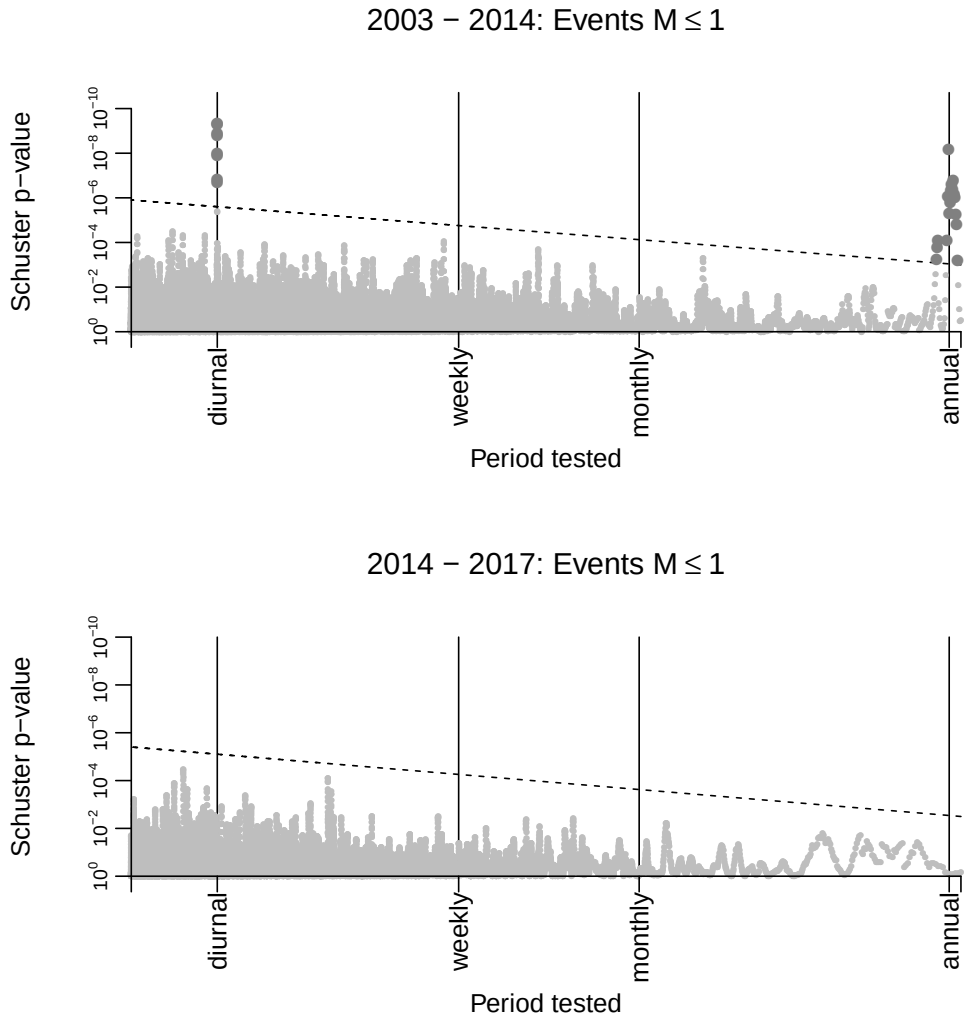

Figure 6.3.: Schuster spectrum test for events with associated magnitudes $M \leq 1$, for two epochs: Epoch 1: January 2003 - January 2014, and Epoch 2: January 2014 - July 2017. The diagonal dotted line delineates the critical region of the test at the 99% confidence level. All points below the line are not unexpected under the null hypothesis of no periodicity; points that exceed the threshold are enlarged (see Ader and Avouac [2013]).

7. Conclusions

In line with the findings as reported in Bierman et al. [2015], there is strong evidence of diurnal and annual (seasonal) periodicity in occurrence times of events with associated magnitudes $M \leq 1$, and some evidence of annual periodicity for events $M \geq 1$. There is no evidence of any type of periodicity for events $M \geq 1.5$. If data from the post-January 2014 epoch are analysed there is no longer any evidence of any periodicity for events regardless of their magnitude.

The absence of evidence of seasonality in occurrence rates in the post-January 2014 epoch for events with small associated magnitudes may be explained by:

- The low power of the statistical test to detect periodicities, due to the small size of the catalogue in the post-January 2014 epoch.
- A lower magnitude of completeness, due to improvements in the geophone network (under the hypothesis that the underlying cause was seasonal changes in event detection rates).
- The less pronounced monthly variations in gas production rates (under the hypothesis that the underlying cause was seasonal variation in gas production rates).
- A combination of the above.

The apparent disappearance of diurnality in occurrence rates of events with small magnitudes is most straightforwardly explained by improvements in the geophone network.

Causal inference is complicated because this is an entirely observational study: there is no scope to manipulate the activity rate using combinations of variables that may have some effect on this rate in the context of an experimental design. The ability to estimate the independent effects of variables depends therefore entirely on the availability of contrasts between the states of these variables between regions and epochs in the gas field.

It is often difficult to establish whether a potential correlation or pattern of interest in the data is "statistically significant". By this we mean that the observed pattern, summarised using a test statistic (computed from the data), is sufficiently extreme such that it is considered unlikely to have arisen under the null hypothesis of no effect between the production-related variable and activity rate. There are a number of reasons for this, in particular:

- If there is no "pre-commitment" in the form of a clearly specified and plausible hypothesis *before* the data are used in a statistical analysis, then it is often difficult to establish the statistical significance of associations (see e.g. Benjamini and Hochberg [1995] and Miller [1980]). Multiple inferences may be made, either consciously or unconsciously by looking at the data, and only the significant relationships may be selected or emphasised in discussions and in support of conclusions. Unguarded use of inference procedures which assume pre-commitment in the form of a single clearly specified hypothesis will lead to a (potentially severely) inflated false positive rate [1]. A skeptical viewpoint that can be taken in this context is that any potential interesting outcomes can only be used to specify new hypotheses that are subsequently tested using future data with pre-commitment.

[1] The rate at which it is erroneously concluded that there is evidence of a relationship (the statistical "type-1 error rate")

- Even if one simple pre-specified hypothesis is being tested, it is often still difficult to devise an appropriate and robust statistical inference procedure with good properties such as a well-controlled false positive rate. False positive rates may be inflated due to erroneous assumptions regarding the (in)dependence of earthquake occurrences due to the presence of aftershocks in the data (see e.g. Naylor et al. [2009] and Marsan and Wyss [2011]).

We note that correlation coefficients in chapter 5 are tested against the null hypothesis for multiple lags and multiple magnitude classes. This approach is particularly prone to false positives.

References

T. J. Ader and J.-P. Avouac. Detecting periodicities and declustering in earthquake catalogs using the schuster spectrum application to himalayan seismicity. *Earth and Planetary Science Letters*, 377: 97–105, 2013.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society series B*, 57(1):289–300, 1995.

S. Bierman, R. Paleja, and M. Jones. SR.15.13132: Statistical methodology for investigating seasonal variation in rates of earthquake occurrence in the groningen field. Technical report, Copyright Shell Global Solutions International, B.V., 2015.

S. Bierman, R. Paleja, and M. Jones. SR.15.13132: Measuring seasonal variation in rates of earthquake occurrence in the groningen field. improved methodology following independent external review. Technical report, Copyright Shell Global Solutions International, B.V., 2016.

B. Dost, F. Goutbeek, T. van Eck, and D. Kraaijpoel. Monitoring induced seismicity in the north of the netherlands: status report 2010. Technical report, KNMI Scientific report: WR 2012-03, 2012.

C. K. Harris. SR.15.binnedparetomle: Maximum likelihood estimates of b-value for induced seismicity in the groningen field. Technical report, Copyright Shell Global Solutions International, B.V., 2015.

Hastie and Tibshirani. *Generalized Additive Models*. Chapman and Hall, 1990.

D. Marsan and M. Wyss. Models and Techniques for Analyzing Seismicity: Seismicity Rate Changes. $http://www.corssa.org/articles/themev/marsan_wyss/marsan_wyss.pdf/$, 2011.

A. Mignan and J. Woessner. Estimating the magnitude of completeness for earthquake catalogs. Technical report, Community Online Resource for Statistical Seismicity Analysis., 2012.

R. Miller. Simultaneous Statistical Inference. *Springer Series in Statistics*, 1980.

M. Naylor, I. Main, and S.Touati. Quantifying uncertainty in mean earthquake interevent times for a finite sample. *Journal of geophysical research*, 114:B01316, 2009.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org/.

D. Randell, G. Feld, K. Ewans, and P. Jonathan. Distributions of return values for ocean wave characteristics in the south china sea using directional-seasonal extreme value analysis. *Environmetrics*, 26:442–450, 2015.

A. Schuster. On lunar and solar periodicities of earthquakes. *Proc. R. Soc. Lond.*, 61:455–465, 1897.

S. Touati, M. Naylor, and I. G. Main. The origin and non-universality of the earthquake inter-event time distribution. *Physical Review Letters*, 102(16):168501, 2009.

S. Wiemer and M. Wyss. Minimum magnitude of complete reporting in earthquake catalogs: examples from alaska the western united states and japan. *Bull. Seismol. Soc. Am.*, 90:859–869, 2000.

S. N. Wood. *Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.* London: Chapman and Hall/CRC, 2006.

S.N. Wood. Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society (B)*, 62:412–428, 2000.

S.N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73:3–36, 2011.

A. Further graphs

In this appendix, we provide further time-series graphs to illustrate variation in counts of events in the KNMI catalogue, for events in different magnitude classes (figure A.1). A Schuster spectrum see chapter 6) for the entire time-series 2003 - 2017 for events $M \geq 1.5$ with an extension to multi-annual periodicities is given in figure A.2.

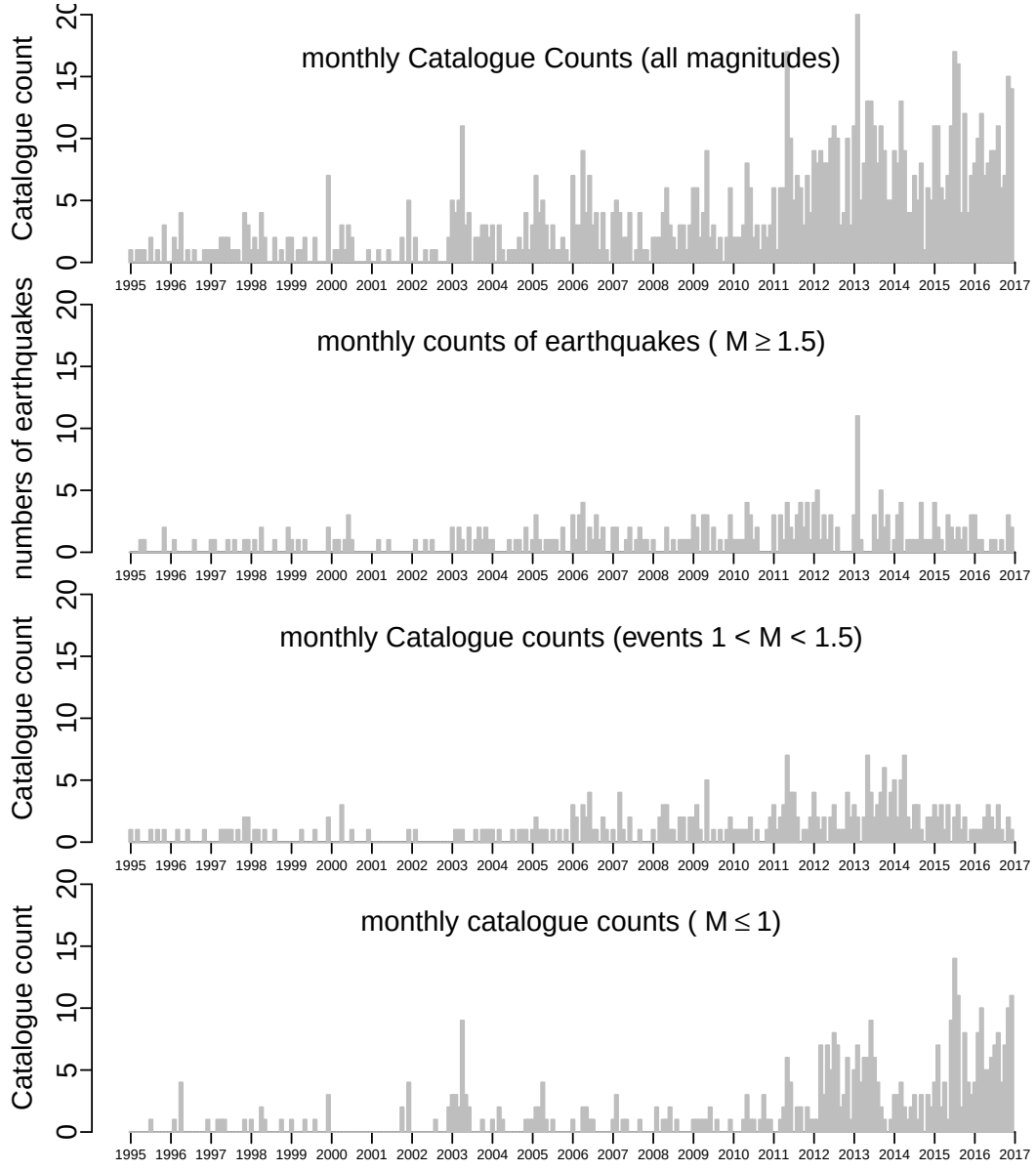

Figure A.1.: Time series of counts of events per calendar month for all events (top graph) or events in different categories of associated magnitudes.

For the January 2003 - January 2014 epoch and for events with associated magnitudes $M \leq 1$ there was strong evidence of diurnality and seasonality (annual periodicity; figure 6.3). This is line with the findings as reported in sections 4.3 and 3 and as reported in Bierman et al. [2015] for this epoch. For the post January 2014 epoch there was no evidence for any type of periodicity.

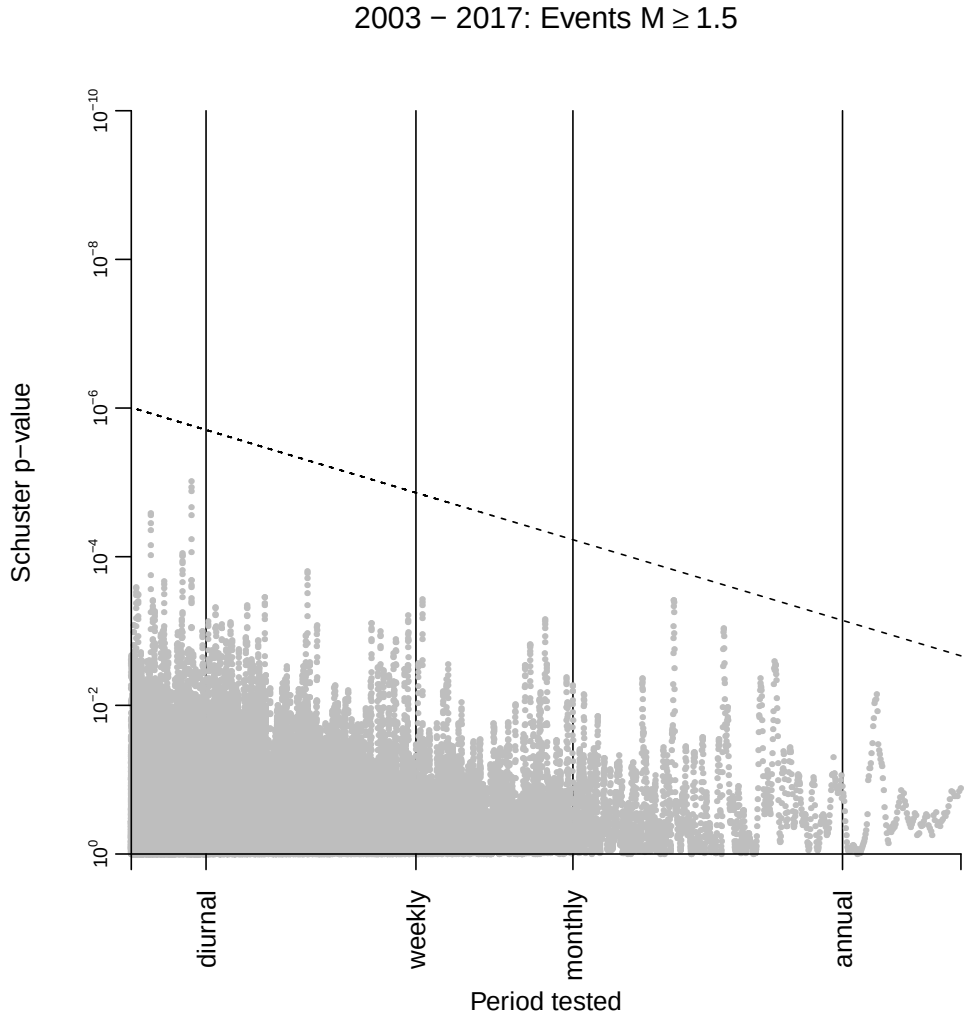

Figure A.2.: Schuster spectrum test for events with associated magnitudes $M \geq 1.5$, for the January 2003 - July 2017 epoch. The diagonal dotted line delineates the critical region of the test at the 99% confidence level. All points below the line are not unexpected under the null hypothesis of no periodicity (see chapter 6).

Bibliographic information

Classification	Unrestricted
Report Number	SR.17.00811
Title	Seasonal variation in rates of earthquake occurrences in the Groningen field
Authors	S.M. Bierman (GSNL-PTI/CA)
Keywords	Seismicity, Groningen, Seasonality, Statistical analysis
Issue Date	October 2018
Period of work	July 2017
US Export Control	Non US — Public Domain
WBSE Code	ZZPT/015656/010125
Reviewed by	D.F. Randell (GSNL-PTI/CA)
Approved by	P. Jonathan (GSUK-PTI/CA)
Sponsoring Company/ Customer	Nederlandse Aardolie Maatschappij
Spons./Cust. Address	Schepersmaat 2 9405 TA Assen, The Netherlands
Issuing Company	Shell Global Solutions International B.V., Amsterdam P.O. Box 38000 1030 BN Amsterdam The Netherlands

Report distribution

Electronic distribution (PDF)

Name, Company, Ref. Ind.

PT Information Services, PTT/TIKE, PT-Information-Services@Shell.com	PDF
Jonathan Philip, GSUK-PTD/TASE	PDF
Bierman Stijn, GSUK-PTD/TASE	PDF
van Elk Jan, NAM-UIO/T/DGR	PDF
Bourne Stephen, GSNL-PTI/RC	PDF
Harris Christopher, GSNL-PTI/RC	PDF
Wentinck Rick, GSNL-PTI/RC	PDF
Paleja Rakesh, GSUK-PTD/TASE	PDF
Randell David, GSUK-PTD/TASE	PDF
Kuperus Eddy, A NAM-UPO/T/GD	PDF
Wardenaar-Noordsij Mariette E, GSNL-PTI/CA	PDF
den Bezemer Taco, NAM-UPO/T/GD	PDF
Nevenzeel Keimpe Jan, SIEP-ITZ/AC	PDF
Oates Steve J, GSNL-PTU/E/S	PDF

Principal typesetting performed with LaTeX system (MiKTeX) using the PTreport2017.cls class employing the necessary modifications to match the Shell Projects and Technology house style.

The PTreport2017.cls class file is written and maintained by IdLTeX (www.idltex.com). All rights reserved.