DRAFT: Statistical methodology to test for evidence of seasonal variation in rates of earthquakes in the Groningen field

Restricted
PRELIMINARY DRAFT: SR.15.xxxxx
April 2015

Restricted **PRELIMINARY DRAFT: SR.15.xxxxx**

DRAFT: Statistical methodology to test for evidence of seasonal variation in rates of earthquakes 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 draft report, we apply statistical methodology to test for evidence of seasonality in rates of earthquakes 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. Events with magnitudes below M=1.5 may not always be detected with the current network of geophones and their epicenters cannot be reliably located. We therefore test for evidence of seasonality for all events and for events with magnitudes $M \geq 1.5$ and $M<1.25$ separately. Our preliminary findings are:

- There is strong evidence that rates of events with associated magnitudes $M<1.25$ vary seasonally within each year. Rates of events with $M < 1.25$ were estimated to be highest around (approximately) week 15 - 25 (April - June), and generally lower in the last half of the year (approximately July - December) compared to the first half (January - June) of the year.
- Some evidence was also found for seasonal variation in rates of earthquakes with magnitudes $M \geq 1.5$. Rates of events with $M \geq 1.5$ were estimated to be highest around (approximately) January and February, and lower in the last half of the year (approximately July - December) compared to the first half (January - June) of the year.
- Monthly variation in field-wide production could be used to explain a statistically significant proportion of the within-year variability in rates of events with magnitudes $M < 1.25$. High monthly rates of production were correlated with high monthly event counts with a delay of approximately 4 calendar months.
- No evidence, or at most statistically weak evidence, could be found of a relationship between monthly production rates and monthly rates of events with associated magnitudes $M \geq 1.5$.
- We note that this study does not provide any evidence of a causal relationship between variation in gas production and event rates.
- We note that in this report only temporal variation in event rates was investigated, and potential spatial variation in event rates and production rates was not taken into account.

Recommendations for further work are :

- Events with magnitudes below approximately M=1.5 may not always be detected with the current network of geophones, and may have large uncertainties in their estimated epicenters or hypocenters. Many analyses of earthquake events are for this reason done on events with magnitudes above M=1.5 only. We recommend that the possibility is investigated that the observed seasonal and diurnal variation in even rates is caused by variation in measurement accuracy of the network of geophones.
- A statistical analysis of seasonality in events outside of the Groningen field.
- An investigation into the extent to which rates of pressure decline vary seasonally within the reservoir.

Amsterdam, April 2015.

Table of Contents

Executive Summary **I**

1. Introduction **1**

2. Earthquake catalogue **3**

3. Seasonality in event rates: Data visualisation **7**

4. Seasonality in event rates: statistical analyses **15**

5. Results of statistical analyses **17**
5.1. Smooth seasonal trend in event rates 17
5.2. Relationship between monthly field-wide gas production and event rates 17

6. Conclusions **25**

References **26**

List of Figures

1.1. Monthly field-wide gas production (grey dashed line) and smoothed earthquake event rates (black solid line). The smoothed event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months. 1

2.1. Map of the outline of the Groningen reservoir (inner grey line) and a buffer of width 1000m (outer blue line). 3
2.2. Maps of epicenters of events in different of ranges of event magnitudes. 5
2.3. Counts of numbers of events in grid cells of 5000m by 5000m. The numbers in the grid cells in the map on the right are used as labels. Grid cells 27, 36 and 37 contain relatively high counts of events with magnitudes $M \leq 1.5$ in comparison to counts of events with magnitudes $M < 1.25$. 6
2.4. 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. 6

3.1. Time series of counts of events per calendar month for all events (top graph) or events in different categories of associated magnitudes. 8
3.2. Time series of counts of events per calendar month for all events, with a panel per year. 9
3.3. Time series of counts of events per calendar month events with magnitudes $M <$ 1.25, with a panel per year. 10
3.4. Time series of counts of events per calendar month events with magnitudes $M \geq$ 1.5, with a panel per year. 11
3.5. Counts of events per calendar month, summed over all years, for events with $M \geq$ 1.5. 12
3.6. Counts of events per calendar month, summed over all years, for events with $M<1.5$. 13
3.7. Field wide monthly gas production and monthly counts of events inside the field boundary with associated magnitudes $M < 1.25$ (all events with $M < 1.25$ or with exclusion of events that occurred within 3 days and 2500 m of a previous event (declustered)). 13
3.8. Field wide monthly gas production and monthly counts of events 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)). 14
3.9. Monthly field-wide gas production (grey dashed line) and smoothed earthquake event rates $M \geq 1.5$ (red solid line) or $M < 1.25$ (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. 14

5.1. Estimated monthly smooth deviations ("month effects") from the annual average event rate (equation 4.3). The solid and dotted lines depict the estimated rates and the 95% confidence interval of the estimates respectively. 18
5.2. Observed and predicted monthly counts of events with and without a smooth month effect for events with $M < 1.25$. 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. 19
5.3. 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. 20
5.4. Monthly counts of events versus monthly field-wide production with a delay of 4 calendar months (one panel per year). 24

List of Tables

2.1. Numbers of earthquakes in the KNMI catalogue with epicenters inside the Goningen 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. 4
2.2. Numbers of earthquakes in the KNMI catalogue with epicenters outside of the Goningen 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. 4

5.1. Overview of model parameters and indicators for evidence that rates of events (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 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. 21
5.2. Overview of model parameters and indicators for evidence that rates of events with $\text{M} < 1.25$ 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 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. 21
5.3. Overview of model parameters and indicators for evidence that rates of declustered events with $\text{M} < 1.25$ 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 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. 22
5.4. Overview of model parameters and indicators for evidence that rates of events with $\text{M} \geq 1.5$ 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 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. 22

5.5. Overview of model parameters and indicators for evidence that 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 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. . . . 23

1. Introduction

A visualisation of the catalogue of timings of earthquake events associated with the Groningen gas field (data obtained from the internet web-pages of the Dutch Meteorological Society: http://www.knmi.nl/seismologie/geinduceerde-bevingen-nl) suggests that event rates may vary seasonally and may, with some time-delay, be strongly correlated with the seasonal pattern in production rates (figure 1.1). The data visualisation is based on a moving average of counts of events, resulting in a temporally smooth trend in event 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 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).

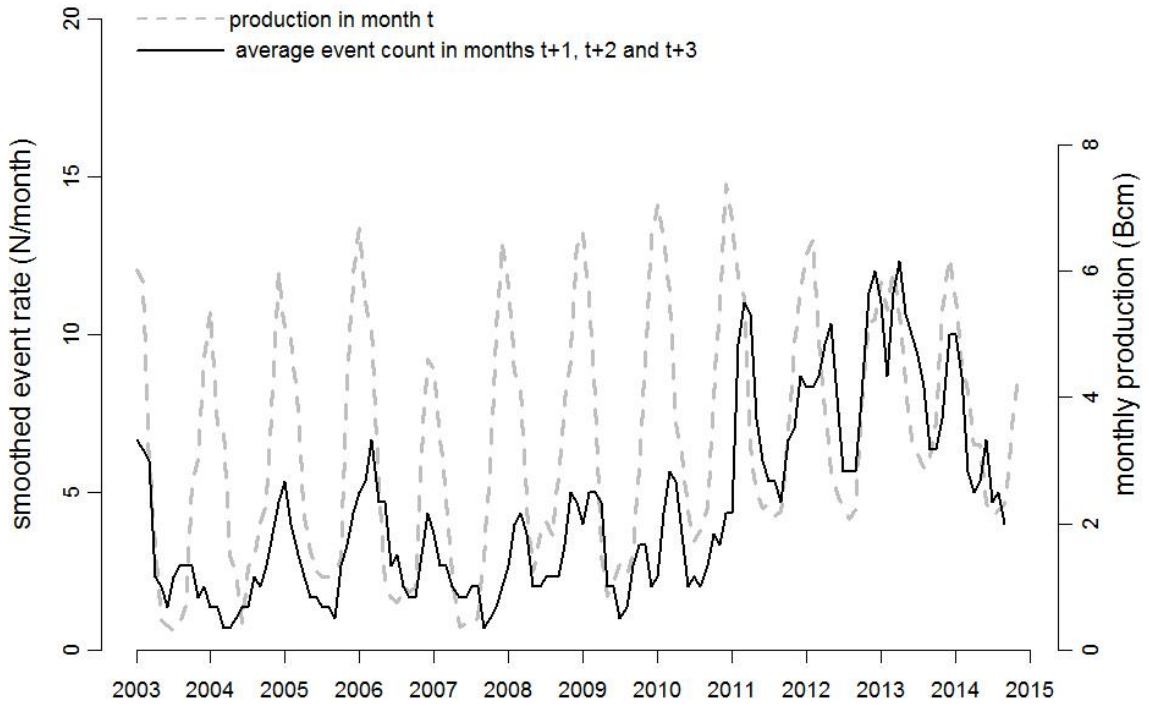

Figure 1.1.: Monthly field-wide gas production (grey dashed line) and smoothed earthquake event rates (black solid line). The smoothed event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months.

While figure 1.1 depicts a correlation between gas production and event 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). A formal analysis of the data is required to test for the presence or absence of seasonality.

In this report, we:

- Provide additional visualisations of the data that provide further insights 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.

We pay special attention to possible differences in apparent seasonality of rates of events within different ranges of event magnitudes. Events with magnitudes below M=1.5 may not always be detected with the current network of geophones, and may have large uncertainties in their estimated epicenters or hypocenters. Most analyses of earthquake events in Shell are for this reason done on events with magnitudes above M=1.5 only. We therefore test for evidence of seasonality for events with magnitudes above M=1.5 and M<1.25 separately, and investigate more generally what differences exist in spatial or temporal patterns of event rates between these two categories 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.

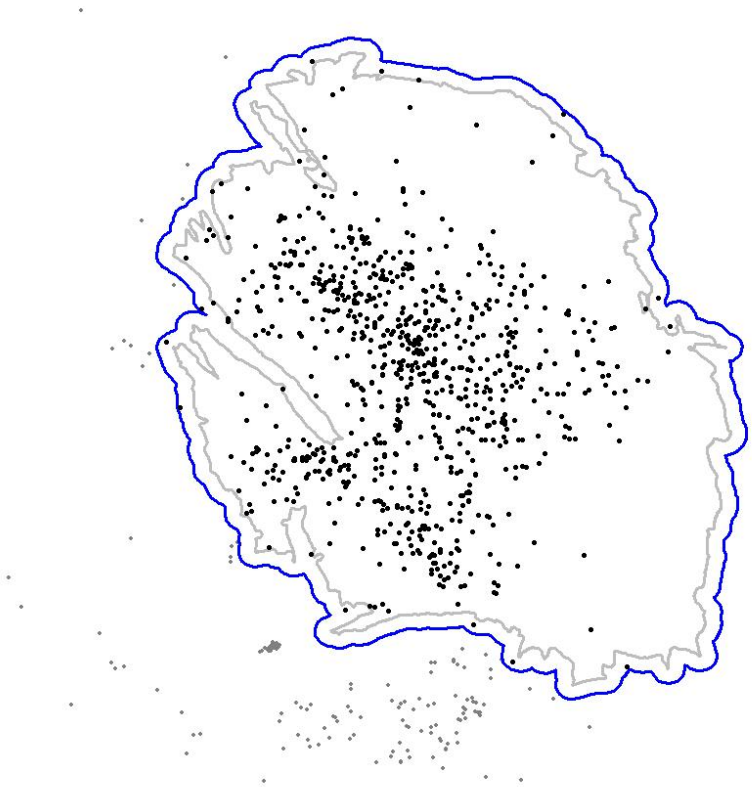

Figure 2.1.: Map of the outline of the Groningen reservoir (inner grey line) and a buffer of width 1000m (outer blue line).

The catalogue 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 Goningen 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.

Table 2.1.: Numbers of earthquakes 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
$\leq$ 1 hour	3	6	9	16	19
$\leq$ 4 hours	3	7	13	26	33
$\leq$ 12 hours	3	8	20	36	49
$\leq$ 1 day	3	11	26	52	74
$\leq$ 2 days	3	13	33	63	99
$\leq$ 3 days	3	13	34	67	119
$\leq$ 5 days	4	14	37	81	152

Table 2.2.: Numbers of earthquakes 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
$\leq$ 1 hour	10	23	26	28	30
$\leq$ 4 hours	12	35	39	42	44
$\leq$ 12 hours	13	42	45	48	51
$\leq$ 1 day	20	47	50	53	56
$\leq$ 2 days	23	49	55	59	64
$\leq$ 3 days	24	51	58	63	68
$\leq$ 5 days	28	54	59	65	71

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. However, a simple analysis based on counts of events in grid cells of 5000m by 5000m, indicates that there is at least one area with relatively high rates of events with magnitudes $M \geq 1.5$ (near the municipality of Loppersum) compared to the rates of events with magnitudes $M < 1.25$ (figure 2.3).

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 2.4). Such a diurnal pattern is not apparent for events with magnitudes $M \geq 1.5$.

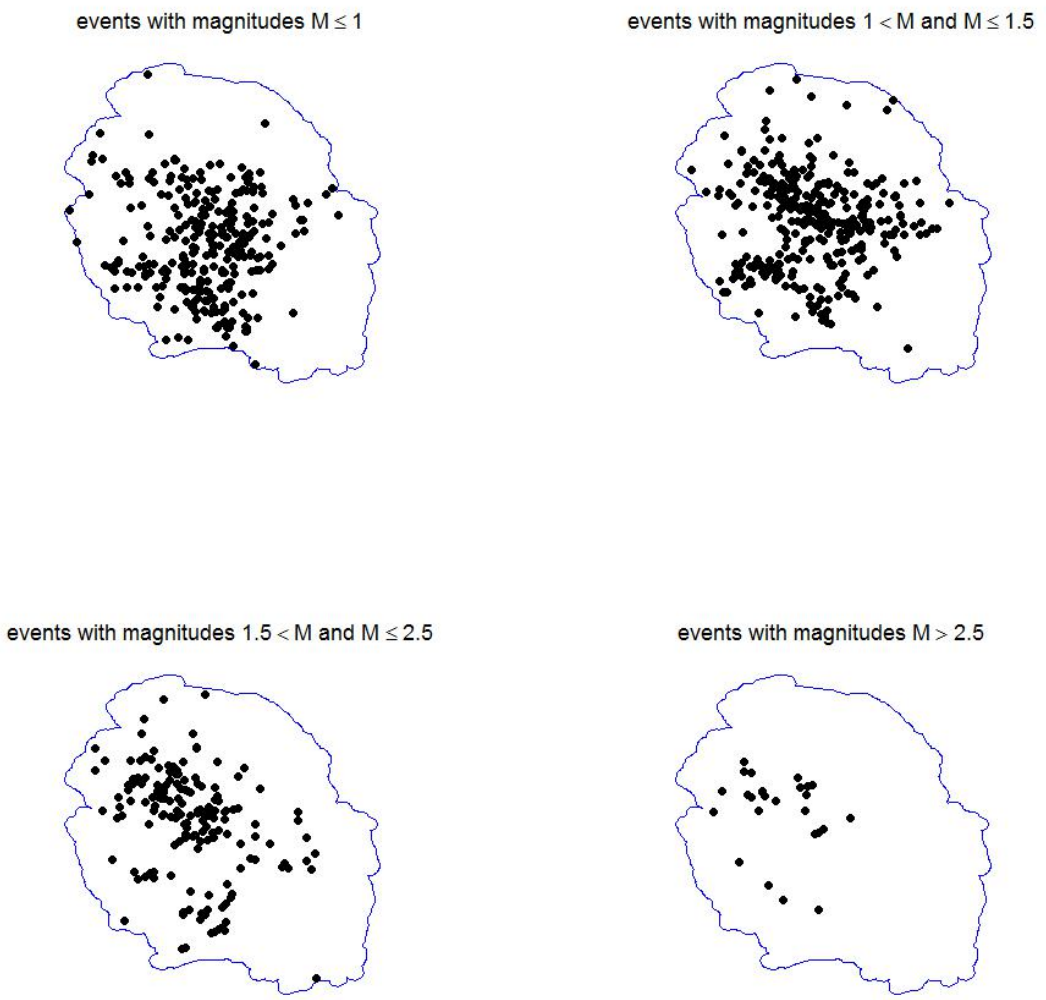

Figure 2.2.: Maps of epicenters of events in different of ranges of event magnitudes.

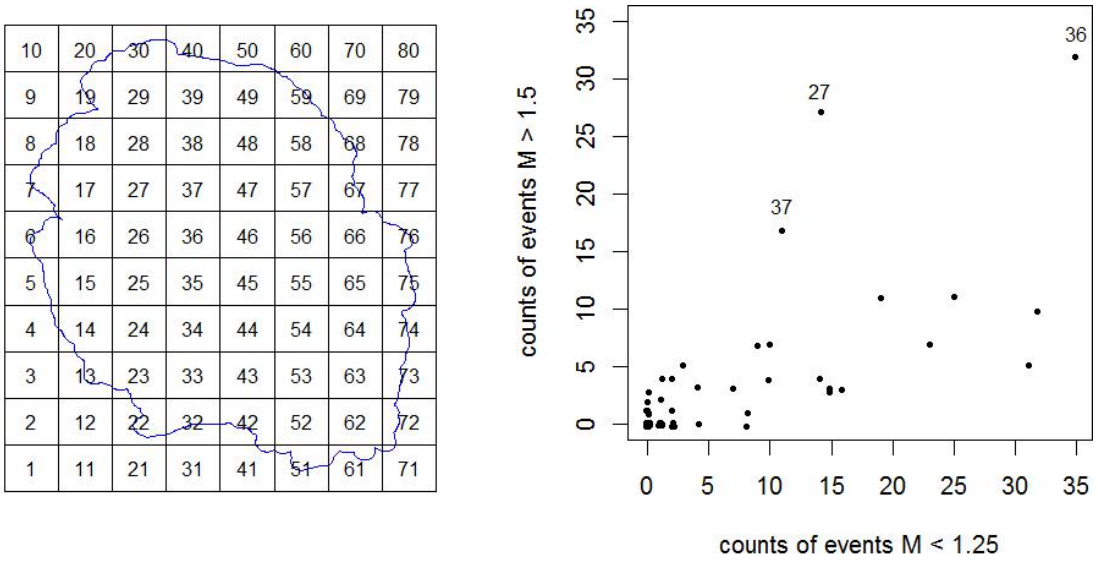

Figure 2.3.: Counts of numbers of events in grid cells of 5000m by 5000m. The numbers in the grid cells in the map on the right are used as labels. Grid cells 27, 36 and 37 contain relatively high counts of events with magnitudes $M \leq 1.5$ in comparison to counts of events with magnitudes $M < 1.25$.

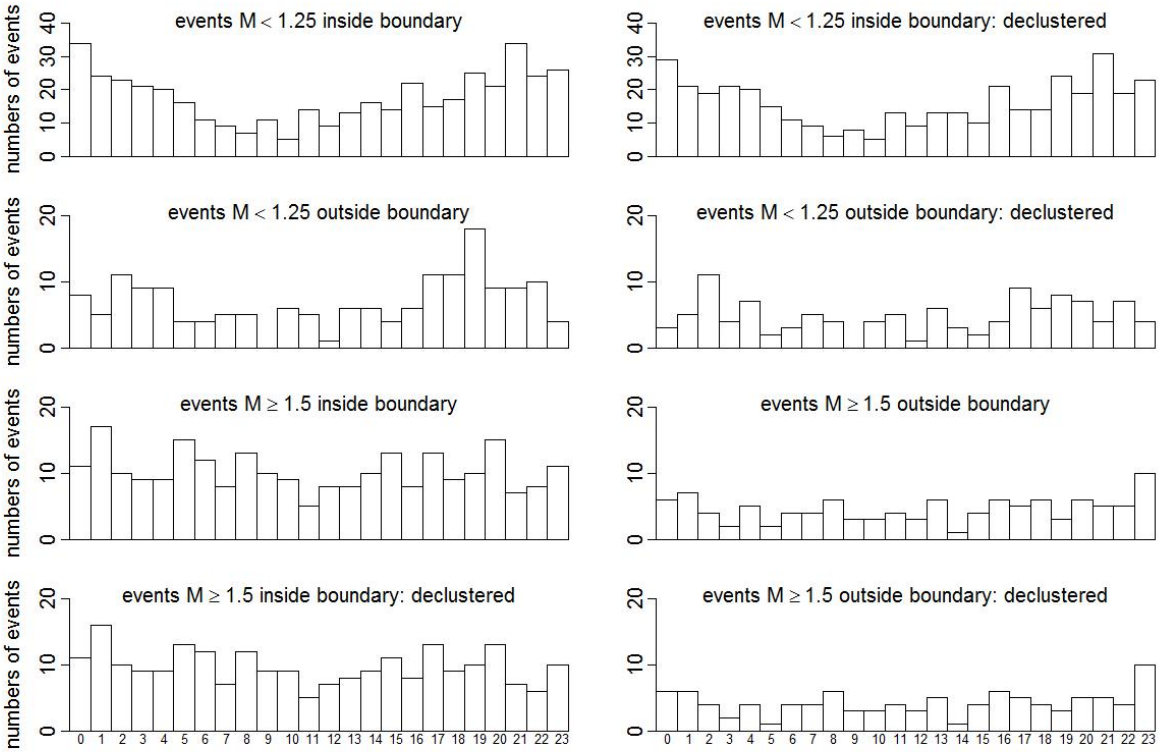

Figure 2.4.: Counts of numbers of events for each of the 24 hours within the day (00:00 - 01:00, 01:00 - 02:00,...,23:00 - 24:00) for events in different categories of associated magnitudes.

3. Seasonality in event rates: Data visualisation

Time series of counts of events per calendar month for all events, and events with magnitudes $M < 1.25$ and $M \geq 1.5$ are given in figure 3.1. To aid visual interpretation, the same information is given in three further graphs with a separate panel per calendar year (figures 3.2 3.3 3.4 for all events, and events with $M < 1.25$ and $M \geq 1.5$ respectively). In particular for events with magnitudes $M < 1.25$ it appears that, for most calendar years, the highest rates of events with associated magnitudes $M < 1.25$ occur in the first half of the year and are particularly high in April or May. This is reflected in higher total counts of events in these months across all years (figure 3.6). For events with associated magnitudes $M \geq 1.5$, events rates also appear higher in the first six months of the year, but most events occurred in January and February (figure 3.6).

Time series of numbers of events per calendar month and monthly field-wide gas production are given in figure 3.7 for events $M < 1.25$ and figure 3.8 for events $M \geq 1.5$. These figures provide a more direct representation of the available information than figure 1.1 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 both categories of magnitude may vary seasonally. If we apply the same smoothing as in figure 1.1, using a time-window of 3 calendar months, clear more-or-less regular seasonal fluctuations in rates occurs for events $M < 1.25$, whereas such fluctuations appear less regular for events $M \geq 1.5$ (figure 3.9).

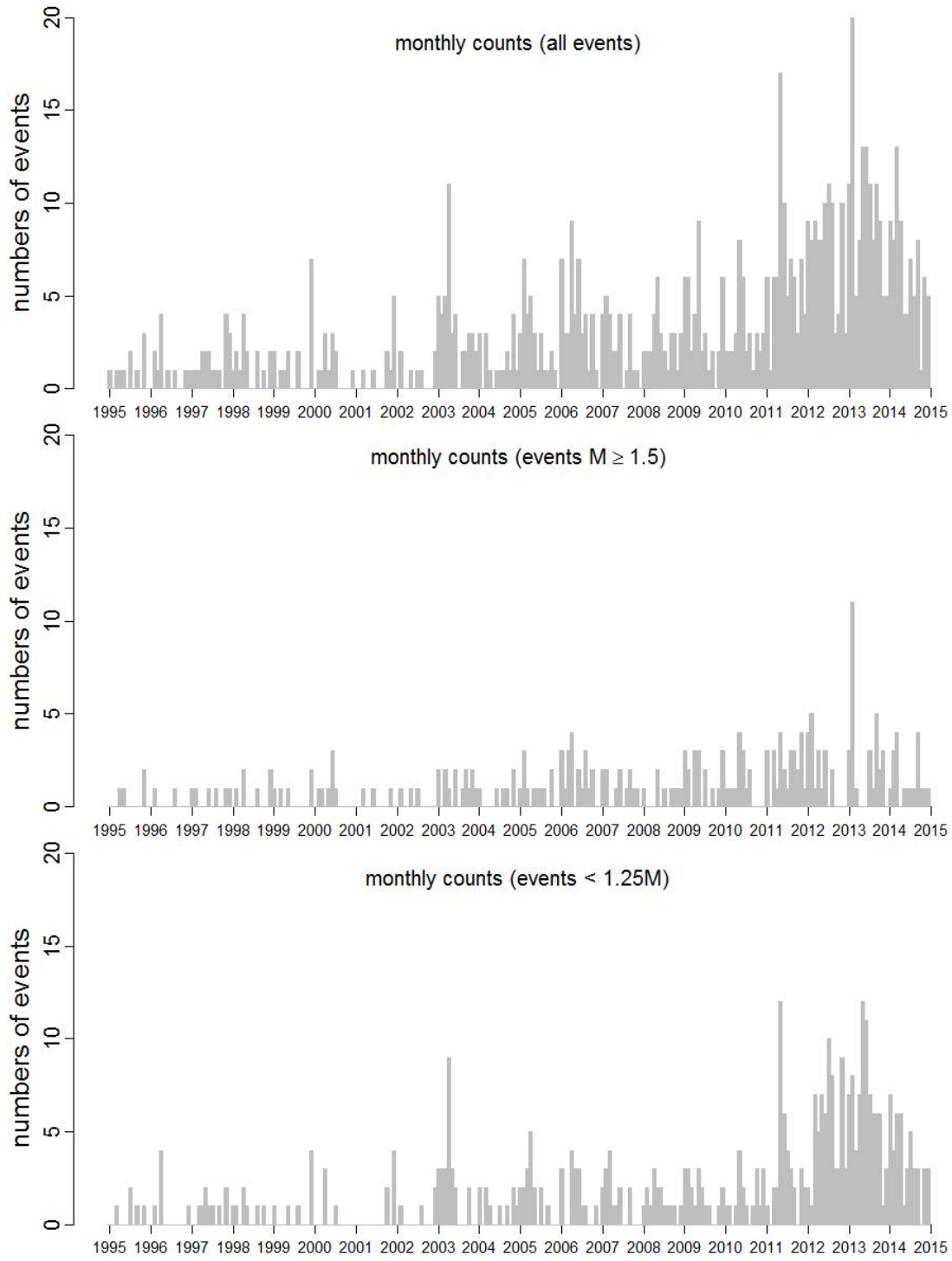

Figure 3.1.: Time series of counts of events per calendar month for all events (top graph) or events in different categories of associated magnitudes.

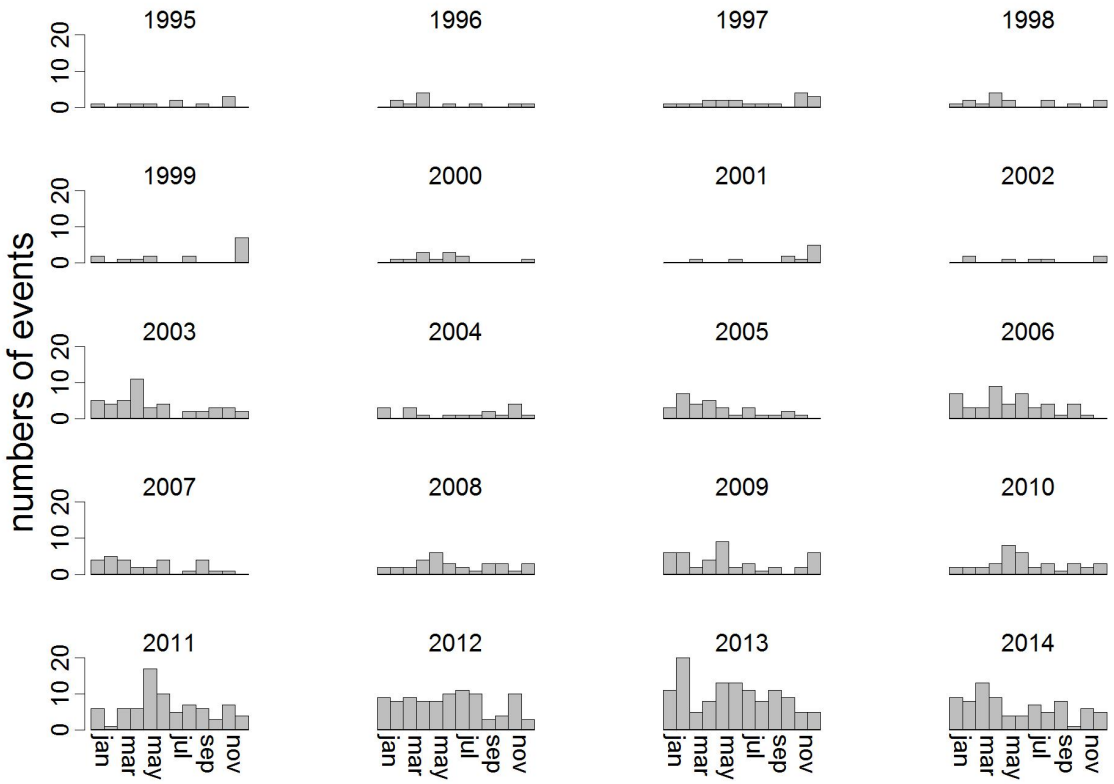

Figure 3.2.: Time series of counts of events per calendar month for all events, with a panel per year.

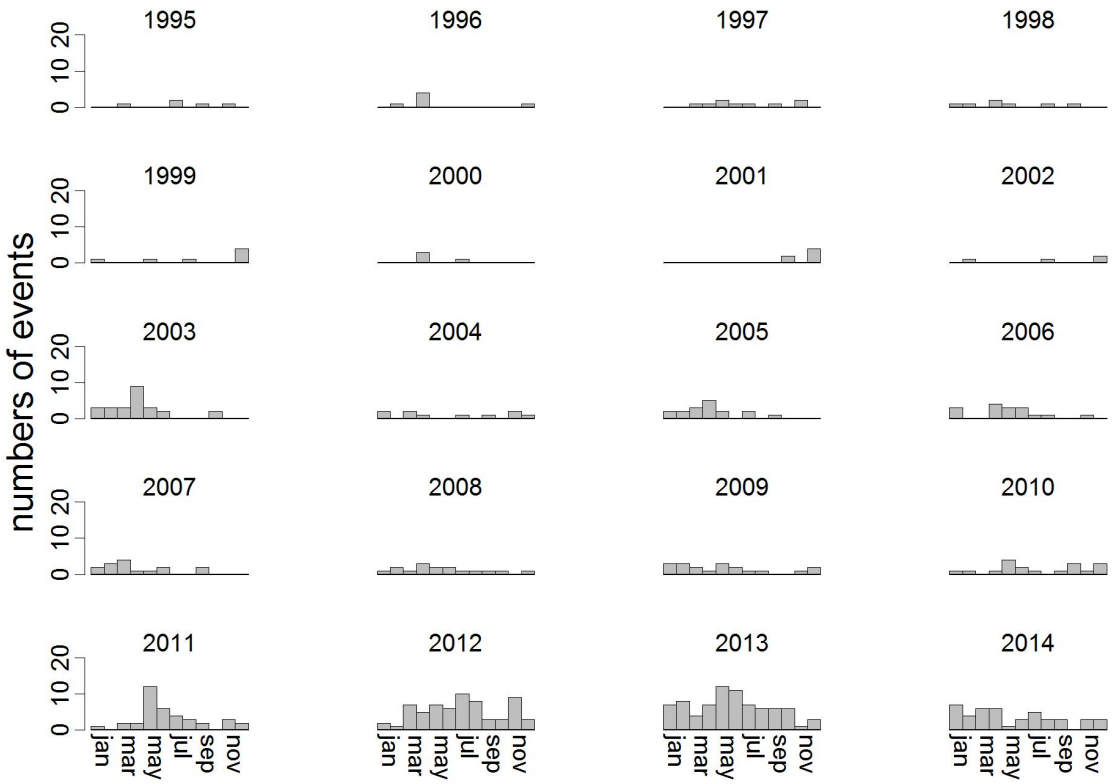

Figure 3.3.: Time series of counts of events per calendar month events with magnitudes M < 1.25, with a panel per year.

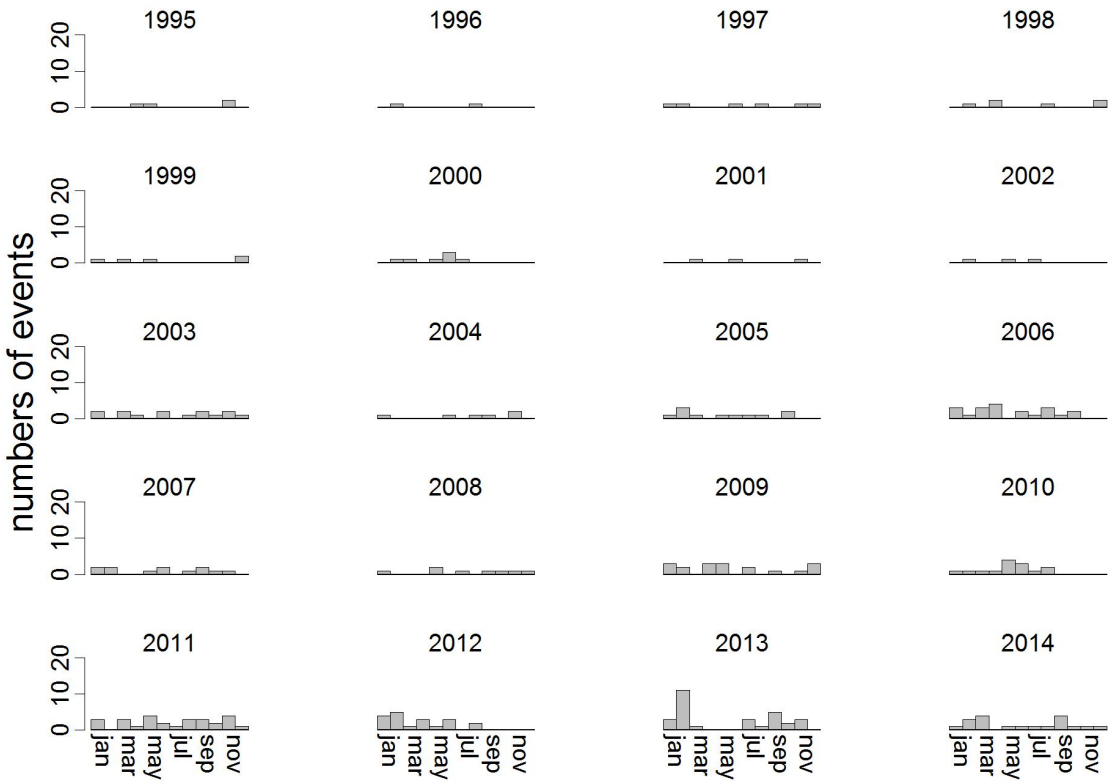

Figure 3.4.: Time series of counts of events per calendar month events with magnitudes M ≥ 1.5, with a panel per year.

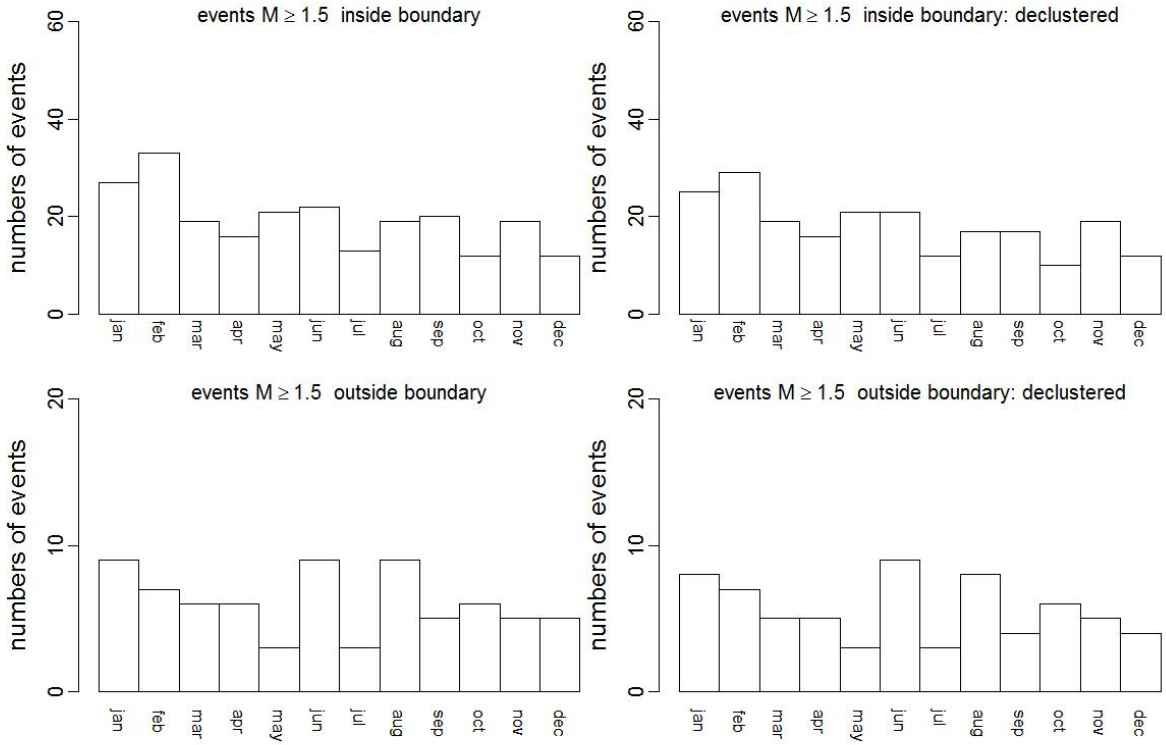

Figure 3.5.: Counts of events per calendar month, summed over all years, for events with M ≥ 1.5.

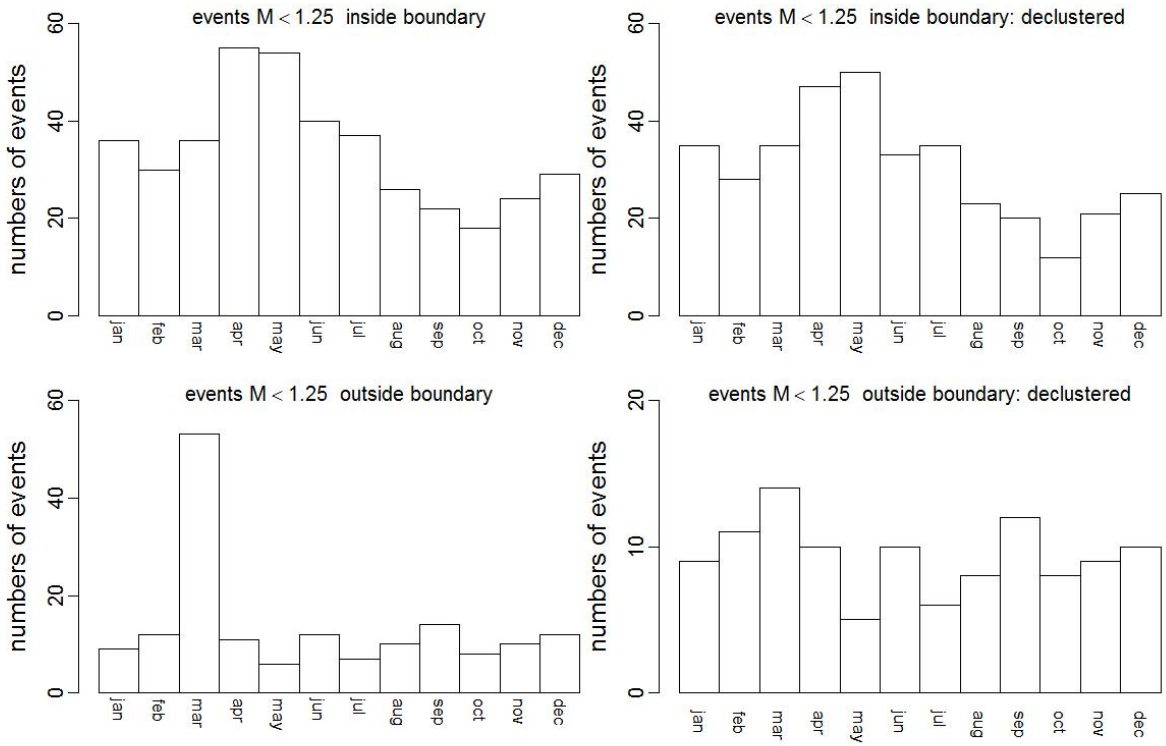

Figure 3.6.: Counts of events per calendar month, summed over all years, for events with M<1.5.

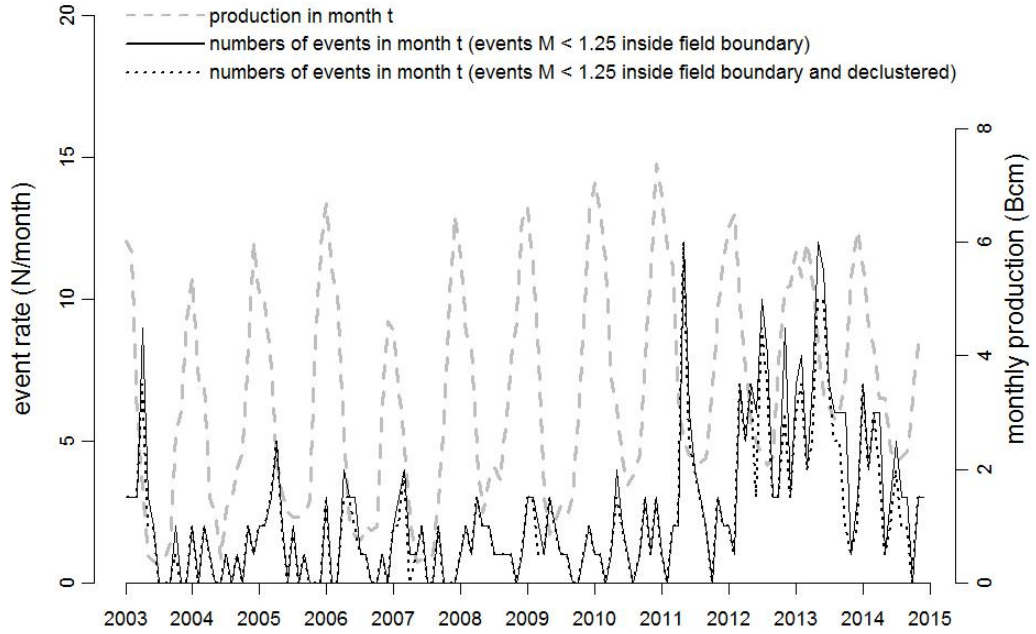

Figure 3.7.: Field wide monthly gas production and monthly counts of events inside the field boundary with associated magnitudes M < 1.25 (all events with M < 1.25 or with exclusion of events that occurred within 3 days and 2500 m of a previous event (declustered)).

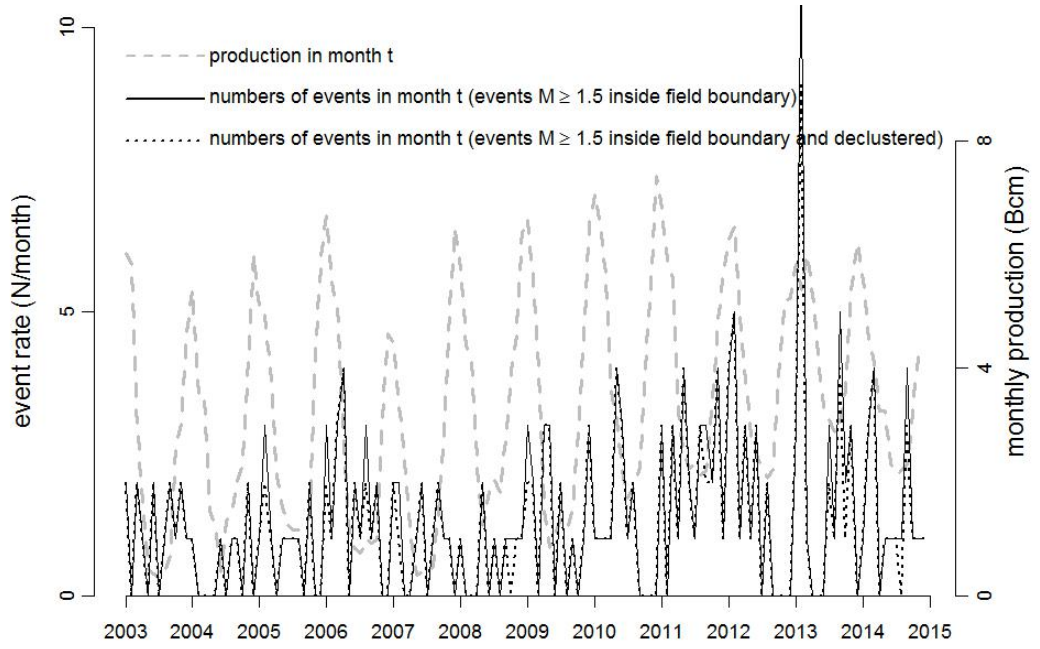

Figure 3.8.: Field wide monthly gas production and monthly counts of events inside the field boundary with associated magnitudes M ≥ 1.5 (all events with M ≥ 1.5 or with exclusion of events that occurred within 3 days and 2500 m of a previous event (declustered)).

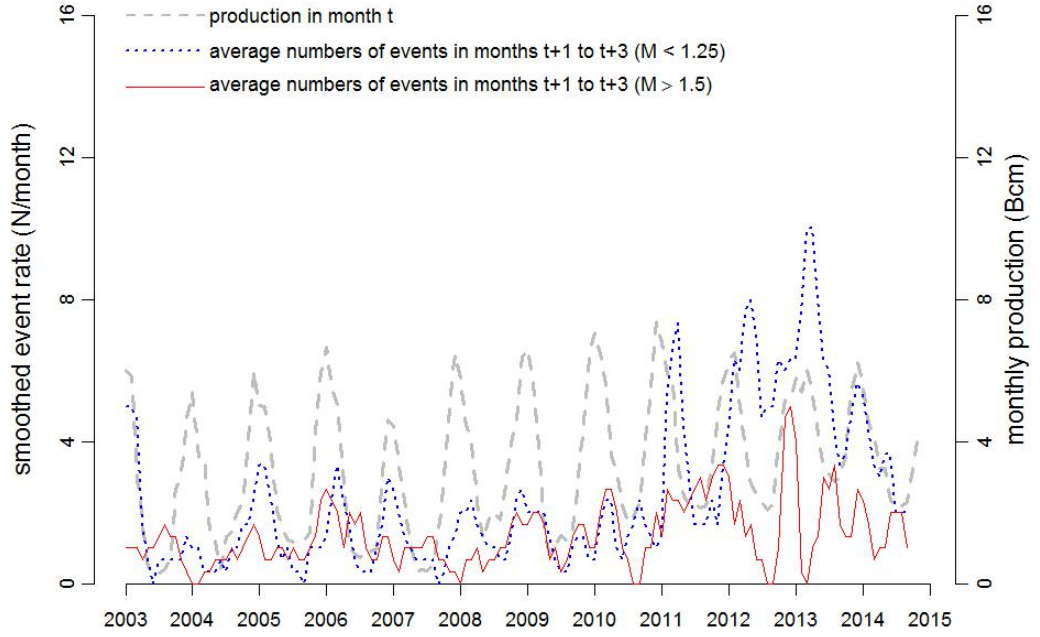

Figure 3.9.: Monthly field-wide gas production (grey dashed line) and smoothed earthquake event rates M ≥ 1.5 (red solid line) or M < 1.25 (blue dotted line). The smoothed event rates are calculated using a "sliding time-window" approach, where each time-window spans three calendar months and the averages 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 Poison distributed with average rate λ ($\lambda > 0$), it takes integer values $1, 2, ...$ 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.

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 event counts that occurred in 2003 through to 2014, because most events occurred in this period and 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 week i ($\mathrm{m}(i) \in \{1, 2, ..., 12\}$).

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

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

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 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:

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

where $\mathrm{O}_{\mathrm{m}(i)}$ is a small 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 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 null model (4.2) is further 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), $P_{(\mathrm{m}(i)-M)}$:

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

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

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 β in equation 4.4 and parameter(s) $\mathrm{s}(\mathrm{w}(i))$ in equation 4.3) 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. In practice, the smaller the value of AIC of a model the better. 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 of $\Delta\mathrm{AIC} \geq 2$ 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 ore 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.

5. Results of statistical analyses

5.1. Smooth seasonal trend in event rates

Events rates were estimated to vary as a function of calendar month over and above the average yearly rates for all events, and events M<1.25 or M $\geq$ 1.5 (figure 5.1). For events with M<1.25 there is strong evidence that rates of events vary seasonally with higher rates in the first half of the year (approximately January - June) and lower rates int he last half of the year (approximately July - December), and highest rates around April and May (figures 5.1 and 5.2). A visualisation of seasonal variation in predicted rates for events with M<1.25 is given in figure 5.2. For events with M $\geq$ 1.5 there is some evidence that rates of events vary seasonally with higher rates in the first half of the year (approximately January - June) and lower rates int he last half of the year (approximately July - December), and highest rates around January and February (figures 5.1 and 5.3). A visualisation of seasonal variation in predicted rates for events with M $\geq$ 1.5 is given in figure 5.3. 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).

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

Strong evidence was found that monthly variation in production rates can explain some of the variation in the within-year differences in rates of events of all magnitudes and of events with magnitudes M<1.25 (table 5.1, 5.2 and 5.3). No evidence, or at most weak evidence, was found of a relationship between monthly production and rates of events M $\geq$ 1.5 (table 5.4 and 5.5). The peak (if any) in monthly rates of events (if any) with associated magnitudes M $<$ 1.25 on average lags the annual peak in monthly production rates by approximately 4 calendar 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 (see e.g. table 5.2).

For most years, monthly production is positively correlated to monthly counts of events with M<12.5 with a lag of 4 months (figure 5.4). The lag which optimises this correlation seems to vary between years (see e.g. figure 3.9). For example, rates of events with M<1.25 appear uncorrelated with monthly production rates in 2012 if a lag of 4 months is applied (figure 5.4) whereas a larger lag would lead to a stronger positive relationship.

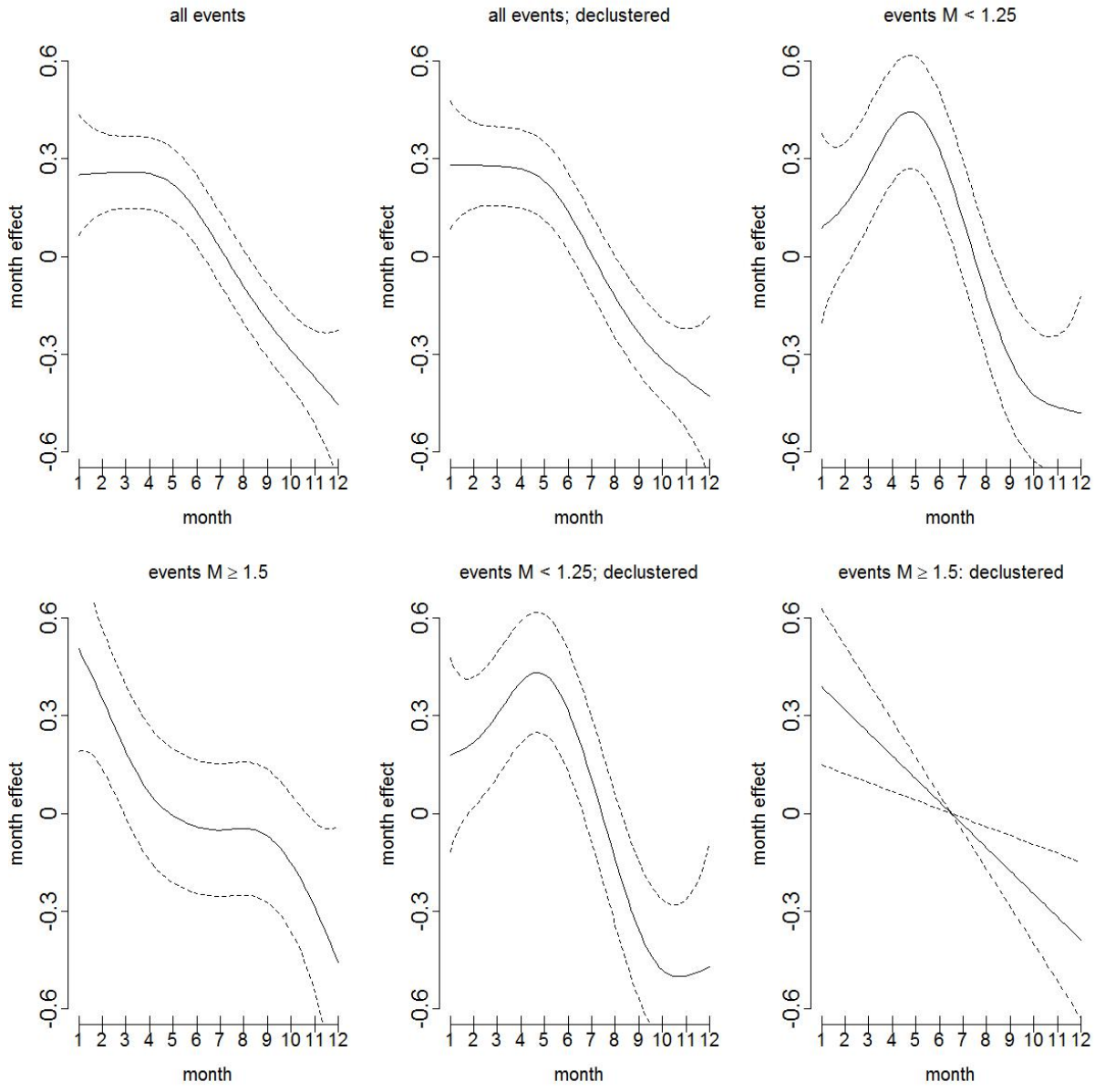

Figure 5.1.: Estimated monthly smooth deviations ("month effects") from the annual average event rate (equation 4.3). The solid and dotted lines depict the estimated rates and the 95% confidence interval of the estimates respectively.

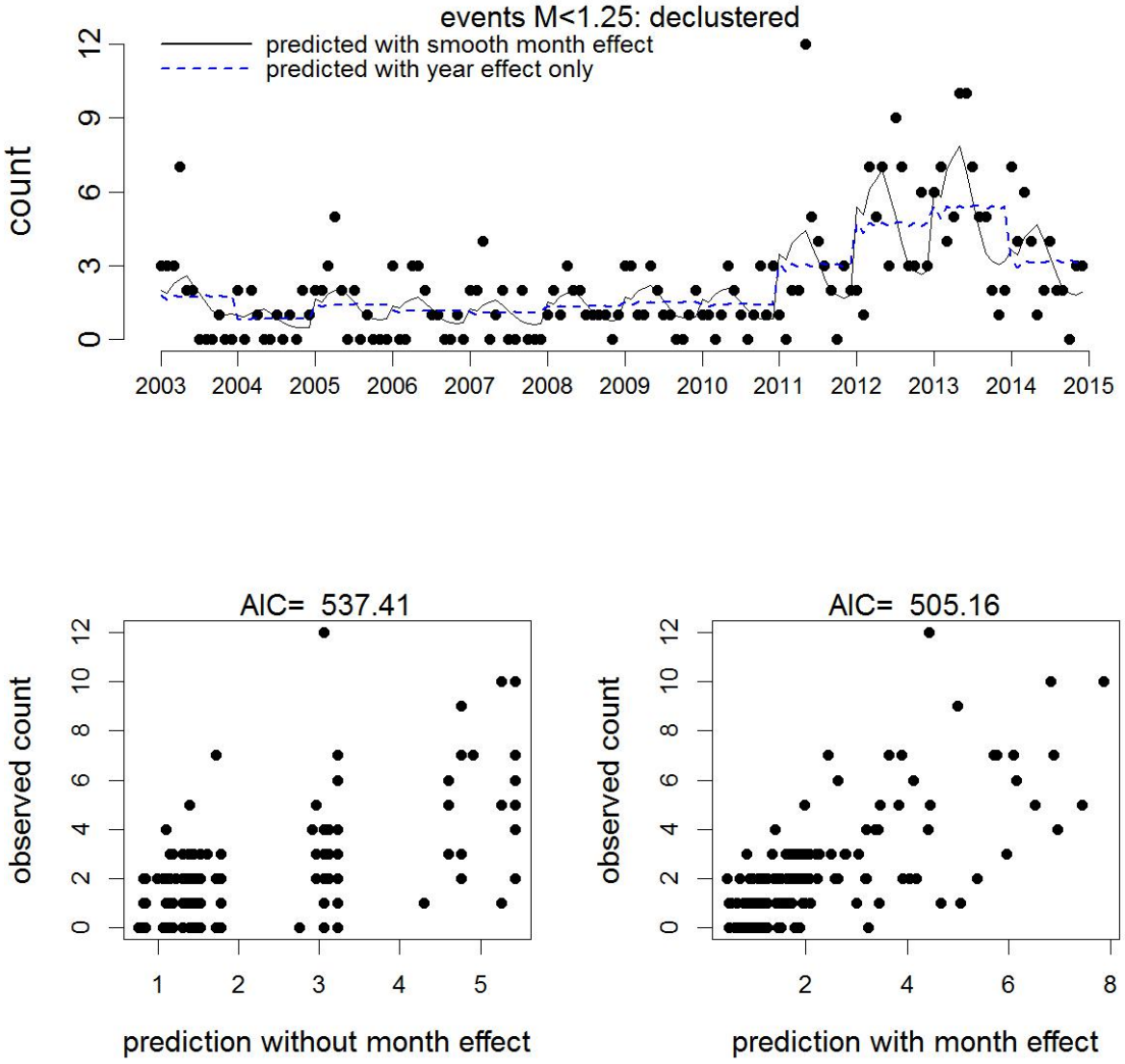

Figure 5.2.: Observed and predicted monthly counts of events with and without a smooth month effect for events with M < 1.25. Top panel: time series of observed and predicted monthly counts. Bottom left panel: predicted versus observed monthly counts for the "null" model with yearly average rates only (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.

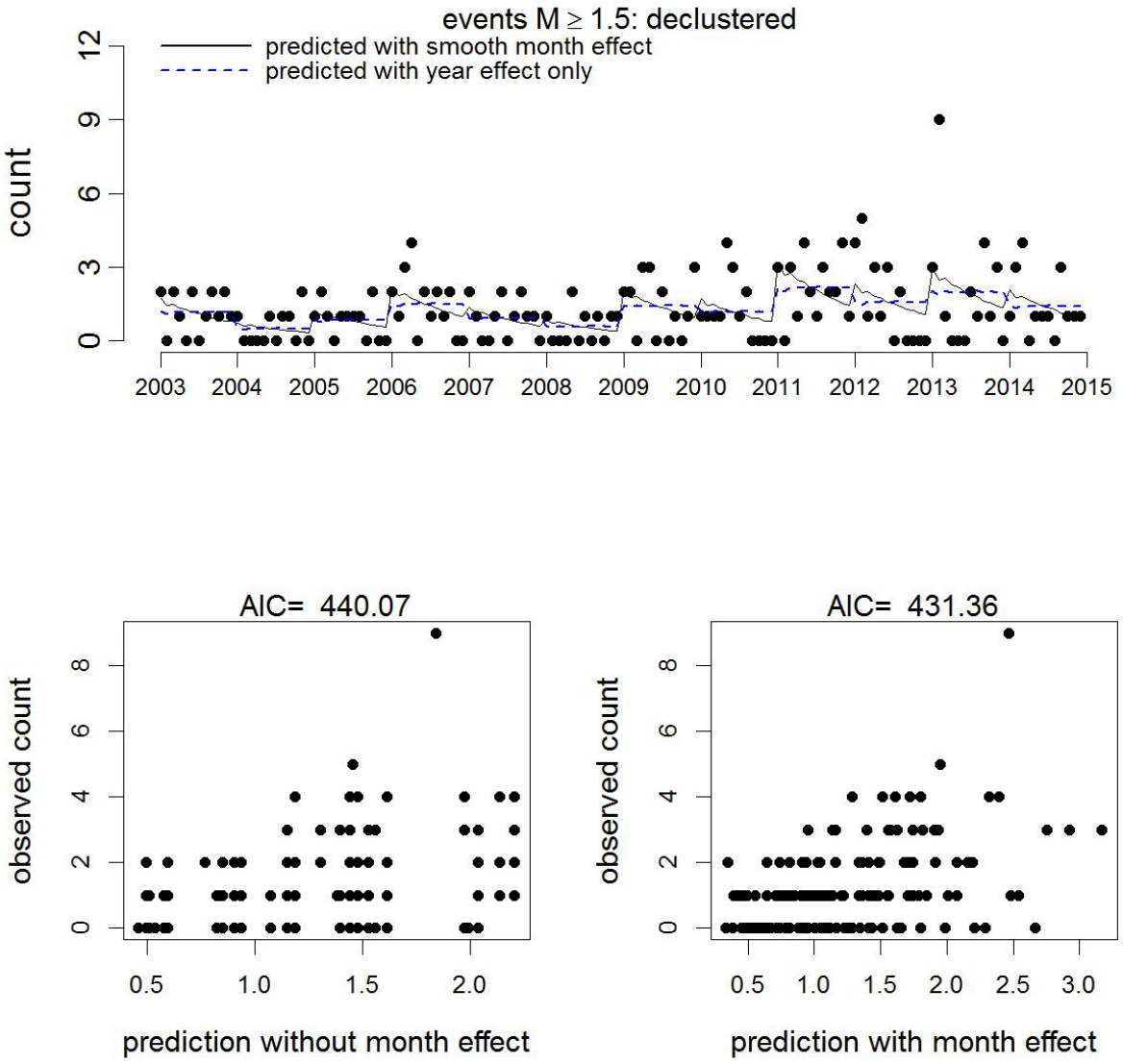

Figure 5.3.: Observed and predicted monthly counts of events with and without a smooth month effect (equation 4.3) for events with M $\geq$ 1.5. Top panel: time series of observed and predicted monthly counts. Bottom left panel: predicted versus observed monthly counts for the "null" model with yearly 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.

Table 5.1.: Overview of model parameters and indicators for evidence that rates of events (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 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

Table 5.2.: Overview of model parameters and indicators for evidence that rates of events with $M < 1.25$ 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 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	448.05	3.11	-2.47	1.106	0.026	-4.47	-4.10	-1.30	1.50	1.93
1	454.23	-1.85	0.42	1.065	0.696	-0.90	-0.41	1.82	3.94	4.30
2	444.25	8.76	3.48	1.057	0.001	2.33	2.69	5.15	7.79	8.59
3	430.65	24.57	5.47	1.063	<0.001	3.59	3.97	6.19	8.23	8.54
4	422.54	30.36	5.98	1.057	<0.001	2.24	2.88	5.46	7.56	7.82
5	426.73	25.15	5.40	1.041	<0.001	-0.66	0.29	3.64	6.26	6.79
6	446.59	4.58	2.62	1.018	0.010	-2.40	-1.95	0.41	2.88	3.41
7	452.71	-1.70	-0.56	1.017	0.582	-5.52	-5.05	-2.76	-0.59	-0.18
8	439.92	10.48	-3.63	1.048	0.001	-9.54	-9.00	-6.15	-3.48	-3.05
9	431.39	19.63	-4.90	1.089	<0.001	-10.50	-9.85	-6.77	-3.61	-3.07
10	426.02	23.87	-5.47	1.119	<0.001	-9.71	-9.01	-5.79	-2.60	-1.94
11	439.27	13.04	-4.06	1.076	<0.001	-6.65	-6.25	-3.36	-0.14	0.38

Table 5.3.: Overview of model parameters and indicators for evidence that rates of declustered events with M $<$ 1.25 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 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	433.09	1.75	-2.21	1.153	0.055	-4.44	-3.84	-1.05	1.85	2.33
1	438.22	-1.63	0.68	1.110	0.541	-0.99	-0.53	1.88	4.12	4.50
2	426.94	9.46	3.75	1.103	0.001	2.35	2.66	5.15	7.76	8.18
3	416.47	22.81	5.52	1.110	<0.001	3.53	3.74	6.03	8.21	8.65
4	409.77	26.39	5.85	1.103	<0.001	2.07	2.68	5.29	7.39	7.71
5	414.98	21.04	5.20	1.087	<0.001	-0.31	0.23	3.38	6.18	6.69
6	432.37	2.80	2.34	1.065	0.028	-2.30	-1.97	0.35	2.64	2.93
7	436.31	-1.41	-0.81	1.067	0.445	-5.15	-4.69	-2.71	-0.77	-0.47
8	423.77	10.36	-3.79	1.101	0.001	-8.93	-8.48	-5.75	-3.42	-3.03
9	414.83	19.92	-5.18	1.148	<0.001	-10.32	-9.65	-6.48	-3.57	-3.03
10	411.75	22.42	-5.57	1.174	<0.001	-9.20	-8.61	-5.56	-2.46	-1.90
11	424.31	12.14	-4.13	1.129	<0.001	-6.70	-6.08	-3.26	0.07	0.53

Table 5.4.: Overview of model parameters and indicators for evidence that 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 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

Table 5.5.: Overview of model parameters and indicators for evidence that 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 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

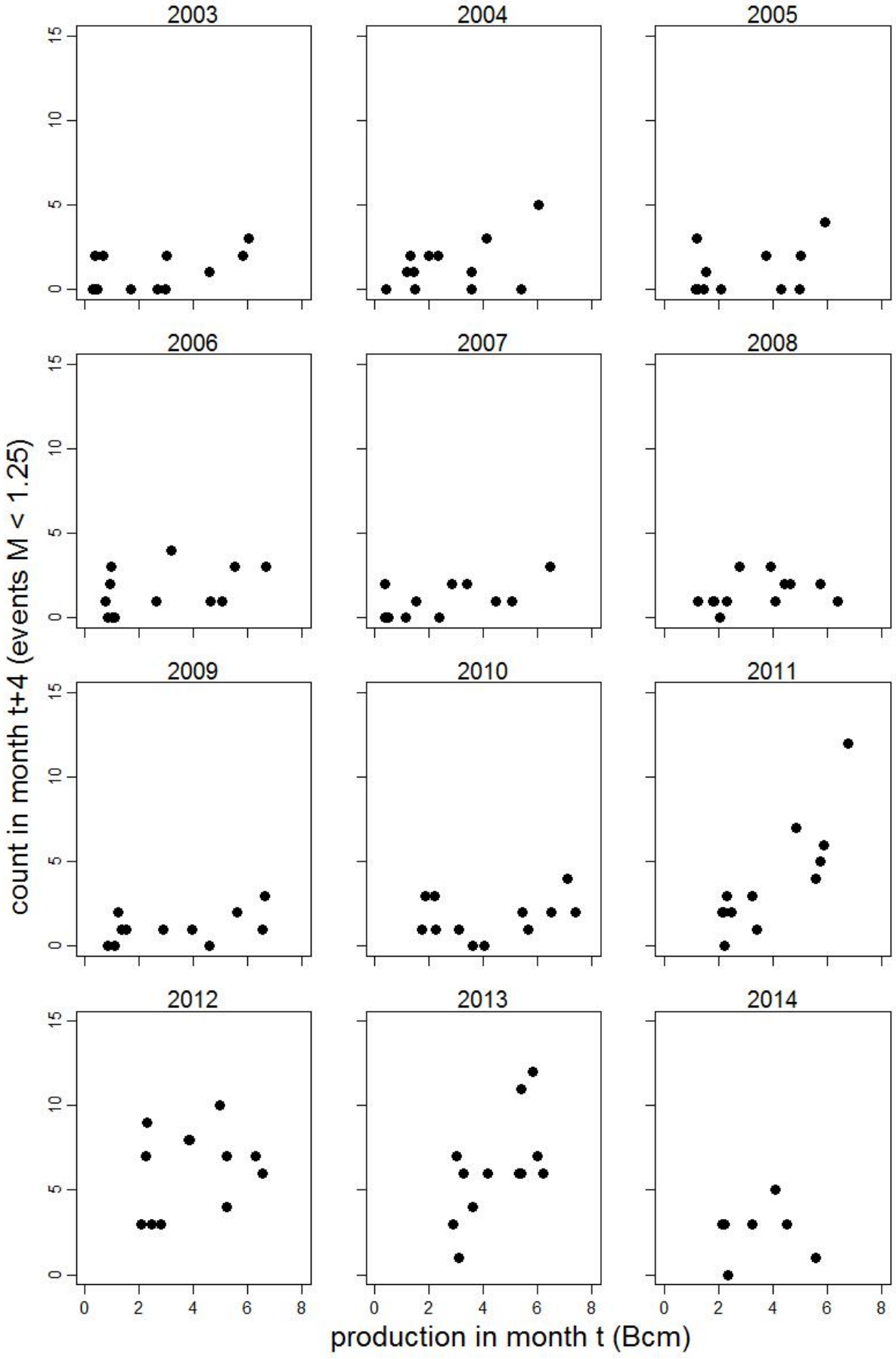

Figure 5.4.: Monthly counts of events versus monthly field-wide production with a delay of 4 calendar months (one panel per year).

6. Conclusions

In this draft report, we apply statistical methodology to test for evidence of seasonality in rates of earthquakes 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. Events with magnitudes below M=1.5 may not always be detected with the current network of geophones and their epicenters cannot be reliably located. We therefore test for evidence of seasonality for all events and for events with magnitudes $M \geq 1.5$ and $M<1.25$ separately. Our preliminary findings are:

- There is strong evidence that rates of events with associated magnitudes $M<1.25$ vary seasonally within each year. Rates of events with $M < 1.25$ were estimated to be highest around (approximately) week 15 - 25 (April - June), and generally lower in the last half of the year (approximately July - December) compared to the first half (January - June) of the year.
- Some evidence was also found for seasonal variation in rates of earthquakes with magnitudes $M \geq 1.5$. Rates of events with $M \geq 1.5$ were estimated to be highest around (approximately) January and February, and lower in the last half of the year (approximately July - December) compared to the first half (January - June) of the year.
- Monthly variation in field-wide production could be used to explain a statistically significant proportion of the within-year variability in rates of events with magnitudes $M < 1.25$. High monthly rates of production were correlated with high monthly event counts with a delay of approximately 4 calendar months.
- No evidence, or at most statistically weak evidence, could be found of a relationship between monthly production rates and monthly rates of events with associated magnitudes $M \geq 1.5$.
- We note that this study does not provide any evidence of a causal relationship between variation in gas production and event rates.
- We note that in this report only temporal variation in event rates was investigated, and potential spatial variation in event rates and production rates was not taken into account.

Recommendations for further work are :

- Events with magnitudes below approximately M=1.5 may not always be detected with the current network of geophones, and may have large uncertainties in their estimated epicenters or hypocenters. Many analyses of earthquake events are for this reason done on events with magnitudes above M=1.5 only. We recommend that the possibility is investigated that the observed seasonal and diurnal variation in even rates is caused by variation in measurement accuracy of the network of geophones.
- A statistical analysis of seasonality in events outside of the Groningen field.
- An investigation into the extent to which rates of pressure decline vary seasonally within the reservoir.

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.

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. 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	PRELIMINARY DRAFT: SR.15.xxxxx
Title	DRAFT: Statistical methodology to test for evidence of seasonal variation in rates of earthquakes in the Groningen field
Subtitle	
Authors	S. Bierman (GSNL PTD/TASE) R. Paleja GSUK PTD/TASE M. Jones GSUK PTD/TASE
Keywords	Keywords1, Keywords2, etc.
Issue Date	April 2015
Period of work	
US Export Control	Non US — Public Domain
WBSE Code	ZZPT/xxxxx/xx.xx
Reviewed by	
Approved by	
Sponsoring Company/ Customer	
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

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.