



Exploring Seismic Hazard in the Groningen Gas Field Using Adaptive Kernel Smoothing

M. N. M. van Lieshout^{1,2} · Z. Baki^{1,2}

Received: 17 January 2023 / Accepted: 16 June 2023
© The Author(s) 2023

Abstract

The discovery of gas in Groningen in 1959 has been a massive boon to the Dutch economy. From the 1990s onwards, though, gas production has led to induced seismicity. In this paper, we carry out a comprehensive exploratory analysis of the spatio-temporal earthquake catalogue. We develop a non-parametric adaptive kernel smoothing technique to estimate the spatio-temporal hazard map and to interpolate monthly well-based gas production statistics. Second- and higher-order inhomogeneous summary statistics are used to show that the state-of-the-art rate-and-state models for the prediction of seismic hazard fail to capture inter-event interaction in the earthquake catalogue. Based on these findings, we suggest a modified rate-and-state model that also takes into account changes in gas production volumes and uncertainty in the pore pressure field.

Keywords Adaptive bandwidth selection · Induced seismicity · Inhomogeneous summary statistics · Kernel smoothing · Pore pressure · Spatio-temporal point process

Mathematics Subject Classification 60G55

1 Introduction

In 1959, a large gas field was discovered in Groningen, a province in the north of the Netherlands. Its recoverable gas volume has been estimated at around 2,900 billion

This research was funded by the Dutch Research Council NWO through their DEEP.NL programme (grant number 2018.033).

✉ M. N. M. van Lieshout
m.n.m.van.lieshout@cwj.nl

¹ CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

² Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

normal cubic metres (Nbcm). The extraction rate has varied considerably over the years (NAM 2022). After a modest start, large amounts were being produced annually during the early 1970s, rising to approximately 85 Nbcm in 1976. During the next decade, the production volumes tended to decrease, followed by somewhat higher values during the first half of the 1990s. Production fell again during the second half of the decade, before rising in the new millennium to over 53 Nbcm in 2013.

From the 1990s, earthquakes were being registered in the previously tectonically inactive Groningen region (KNMI 2022). In particular, the earthquake near Huizinge in August 2012 with a magnitude of 3.6 led to public demand for a reduction of gas production volumes. The government reacted with legislation to phase out gas extraction, and by 2020, production had fallen to less than 8 Nbcm (NAM 2022).

Numerous studies have been conducted on the Groningen field. For example, Geerdink (2014) models the intervals between earthquakes in terms of the cumulative and annual production rates, pressure, subsidence and fault density. More recent studies include those by Post et al. (2021) and Trampert et al. (2022). Van Hove et al. (2015) propose a Poisson auto-regression model for the annual hazard maps in terms of subsidence, distance to major fault lines, and gas extraction in previous years. Both Hetteema et al. (2017) and Vlek (2019) explore the temporal development of seismicity in Groningen by proposing a linear model for the relation between the number of earthquakes over specific periods and gas production volumes. Sijacic et al. (2017) focus on the detection of changes in the rate of a temporal Poisson point process by Bayesian and frequentist methods. Bourne et al. (2018) modify Ogata's space–time model (Ogata 1988) to include changes in stress level and estimate the probability of fault failures. Other authors, notably Candela et al. (2019), Dempsey and Suckale (2017) and Richter et al. (2020), discuss the modelling of seismicity in relation to stress changes based on a differential equation and embed these in a space–time Poisson point process. For a more detailed review of the advantages and limitations of some important modelling approaches including deterministic physical models, statistical models, machine learning and hybrid approaches, we refer to Kühn et al. (2022).

In a previous paper (Baki and Van Lieshout 2022), we investigated the temporal development of seismicity in Groningen using cumulative and recent gas production as dependent variables in a regression model. We concluded that a decrease in production leads to decreased seismicity. Here, we extend the analysis to take into account spatial variations. First, we calculate non-parametric estimates for the spatio-temporal hazard map by means of an adaptive kernel smoother (Abramson 1982; Davies et al. 2018; Van Lieshout 2021) and generalise the bandwidth selection approach suggested by Van Lieshout (2021, 2022) to the space–time domain. Using this map, we investigate whether a Poisson point process model would suffice. Employing inhomogeneous summary statistics, we find that there is interaction which cannot be accounted for by Poisson models, including the state-of-the-art rate-and-state models (Candela et al. 2019; Dempsey and Suckale 2017; Richter et al. 2020). Since these models rely on differential equations for changes in Coulomb stress or pore pressure, we shift our attention to pressure and production data in the public domain. The production values are measured monthly at wells, whereas pressure is gauged at irregular times at some wells as well as at several observation and seismic monitoring stations. To obtain a gas production map, we use an adaptive kernel smoothing technique with local edge

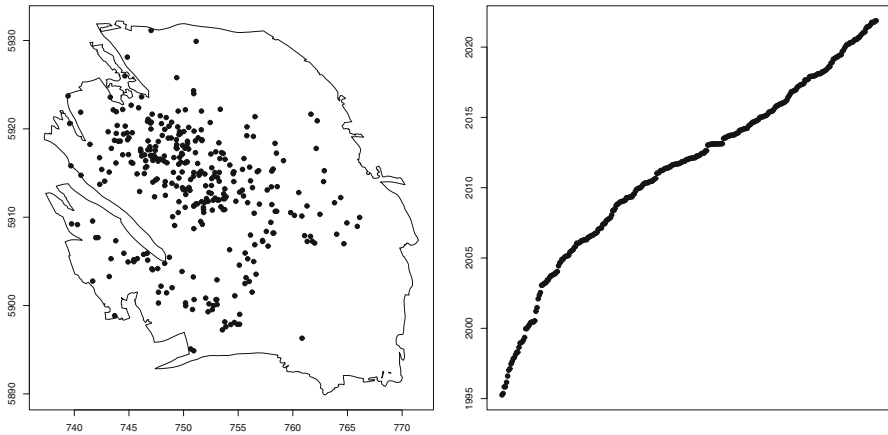


Fig. 1 Spatial (leftmost panel) and temporal (rightmost panel) projections of the 332 earthquakes of magnitude 1.5 or larger with epicentres in the Groningen gas field that occurred in the period from 1 January 1995 up to 31 December 2021. The coordinates are in the UTM system (zone 31), with kilometre as the unit of measure

correction (Van Lieshout 2012). For the pressure values, we fit a Gaussian random field, the mean function of which is modelled as a polynomial in space and time. Such an approach allows us to rapidly estimate pressure from sparse data and acknowledges the inherent uncertainty, but it ignores the field permeability. More elaborate approaches would require detailed expert knowledge and reservoir modelling (De Zeeuw and Geurtsen 2018). Finally, we propose a modification of the rate-and-state models of Candela et al. (2019), Dempsey and Suckale (2017) and Richter et al. (2020) that may exhibit clustering, accounts for the uncertainty in pore pressure, takes into account the varying gas production, and is amenable to monitoring by means of Markov chain Monte Carlo methods based on the history of recorded earthquakes.

This paper is organised as follows. In Sect. 2, we describe the data. Section 3 carries out a comprehensive second-order analysis, and Sect. 4 is devoted to extrapolation of gas production and pore pressure measurements from wells to the entire gas field. The paper closes with our proposed modification of the Coulomb rate-and-state seismicity model.

2 Data

Data on the Groningen gas field and the induced earthquakes are available from various sources.

2.1 Shapefiles for the Groningen Gas Field

Shapefiles for the Groningen gas field can be downloaded from the Geological Survey of the Netherlands TNO website (TNO 2022). The files are updated monthly and contain a polygonal approximation of the border of the field. In this paper, we use

the map that was published in April 2022. The coordinates of the field are given in the UTM system using zone 31 with metre as the spatial unit, which we rescale to kilometre. The boundary is outlined in the leftmost panel in Fig. 1.

2.2 Earthquake Catalogue

An earthquake catalogue for the Netherlands is maintained by the Royal Dutch Meteorological Office KNMI (2022). Data on the period before 1995 are not reliable due to the inaccuracy of the equipment used. Moreover, a threshold on the magnitude is necessary to guarantee data quality. According to Dost et al. (2012), for data from 1995, earthquakes of magnitude 1.5 or larger can be reliably recorded; a threshold of 1.3 can be used for the period from 2010 onwards due to an extension of the monitoring network (Hettema et al. 2017). We use data over the time window from 1995 to 2021 and therefore work with a magnitude 1.5 threshold. The coordinates of the epicentres are listed in terms of latitude and longitude. To avoid distortions and for compatibility with the gas field map, we project them to UTM coordinates (zone 31, in kilometres). This procedure results in 332 earthquakes, the spatial and temporal projections of which are shown in Fig. 1. The leftmost panel depicts the locations of all earthquakes regardless of time. To obtain the rightmost picture, the earthquakes are ordered chronologically. For each earthquake, its occurrence time is then plotted against its index number. Thus, when the time between successive earthquakes is long, the curve is steep; flatter pieces correspond to shorter inter-event times.

2.3 Wells

The exploration and production company NAM maintains a number of production, injection and observation wells. Their coordinates (in the Amersfoort projected coordinate system used in the Netherlands) are available from the production plans published on NAM (2022). For compatibility, we transform the Amersfoort coordinates to UTM (zone 31, in kilometres). Of the 52 locations in the Groningen gas field shown in Fig. 2, 29 are production wells (indicated by a cross), one is an injection well (indicated by a circle) and 22 are observation wells (indicated by a triangle).

A few remarks are in order. Firstly, some production wells (at Midwolda, Noordbroek, Nieuw Scheemda and Uiterburen) were taken out of production around the year 2010 and are no longer in use. Secondly, two south-westerly observation wells (at Kolham and Harkstede) were drilled in a peripheral field rather than in the main reservoir. Finally, up to the middle of the 1970s, small amounts of gas were extracted from wells not earmarked for production.

2.4 Gas Extraction

Monthly production values from the start of preliminary exploration in February 1956 up to and including December 2021 were kindly provided by Mr Rob van Eijs from

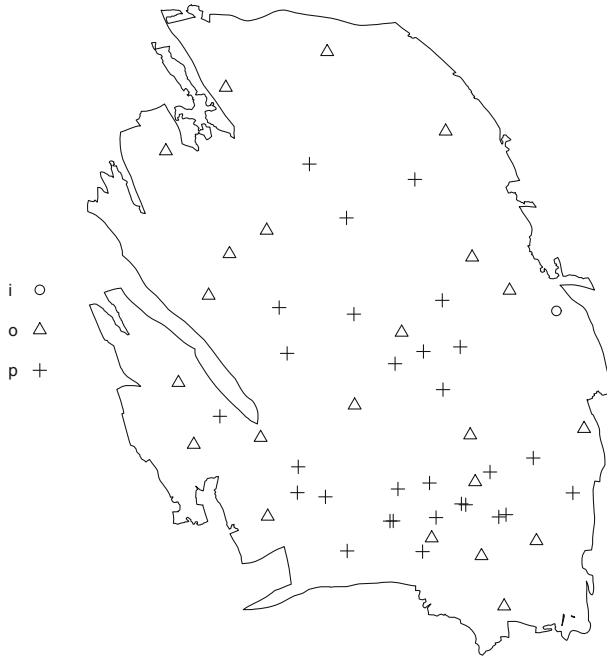


Fig. 2 Production (crosses), observation (triangles) and injection (circles) well locations in the Groningen gas field. The units are as in Fig. 1

Shell for all 29 production wells. The figures were given in cubic metres, which we rescale to Nbcm.

In Fig. 4, we show the production time series for six wells chosen to show a range of different patterns: Bierum in the north-east, De Paauwen in the centre and Eemskanaal in the west of the gas field, Amsweer in the central east, Tusschenklappen in the south and Zuiderpolder in the south-east. Their locations are indicated on the map in Fig. 3.

One may observe that not all wells were drilled at the same time and that some were not in use during the entire period. Also, there are differences in the amount of gas extracted: The production figures for Eemskanaal-13 are lower than average. The sharp decline in production from 2014 following legislation is readily apparent.

2.5 Pore Pressure Observations

On the website of NAM (2022b), pore pressure observations are available over the period from April 1960 until November 2018. In total, there are 2,056 observations. However, these data need some cleaning, as discussed in the Appendix.

After cleaning, we are left with 2,009 pore pressure measurements (over the period from April 1960 until November 2018). Figure 5 shows time series at four locations: a production well in the south-west (Slochteren), two observation wells (Harkstede and Kolham) in peripheral fields and the injection well at Borgweer in the east. Their locations are indicated in the map in Fig. 3. Generally, the graph is initially flat, followed

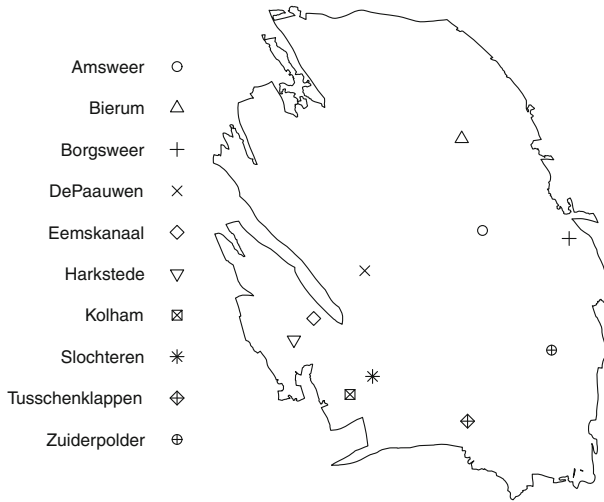


Fig. 3 A selection of wells indicated by name. The units are as in Fig. 1

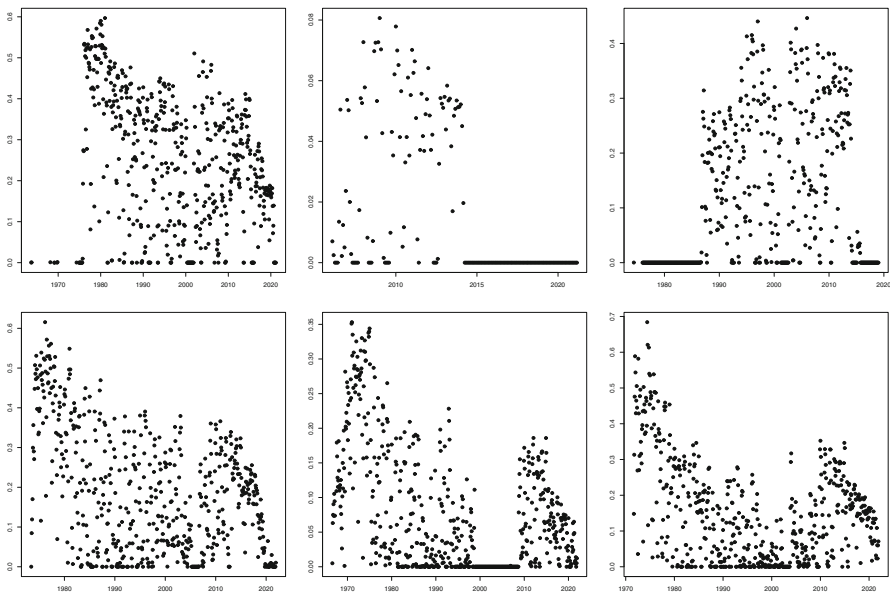


Fig. 4 Monthly production in Nbcm against time for the wells Bierum, Eemskanaal-13 and De Pauuwen (left to right, top row) and Amsweer, Tusschenklappen and Zuiderpolder (left to right, bottom row)

by a decrease. Note that the measurements in the periphery are higher than those in the main reservoir.

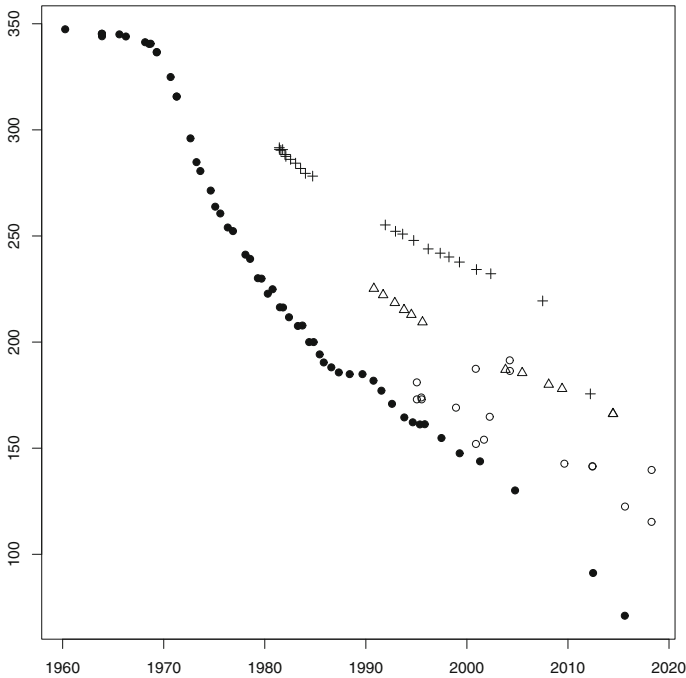


Fig. 5 Pore pressure measurements in bara against time for Slochteren (discs), Harkstede (crosses), Kolham (triangles) and Borgsweer (circles)

3 Exploratory Data Analysis

In this paper, we treat the earthquake catalogue as a spatio-temporal point pattern, a realisation of a point process in space and time. Formally, let Ψ be a simple spatio-temporal point process in $W_S \times W_T$ for bounded open sets $W_S \subset \mathbb{R}^d$ and $W_T \subset \mathbb{R}$, and suppose that its first-order moment measure exists and is finite and absolutely continuous with respect to the product of Lebesgue measures ℓ in space and time (see e.g. Chiu et al. 2013). Write λ for its Radon–Nikodym derivative, known as the intensity function, and $N(\Psi \cap (W_S \times W_T))$ for the number of points placed by Ψ in $W_S \times W_T$. Intuitively speaking, $\lambda(s, t)dsdt$ is the probability that Ψ places a point in the infinitesimal region $dsdt$ around $(s, t) \in W_S \times W_T$.

As usual in spatial statistics, we start with an empirical exploration of the trend and interaction.

3.1 Adaptive Kernel Estimation of the Intensity Function

Our first step is to estimate the intensity function of the point process of earthquakes, which describes how many tremors are expected as a function of place and time.

The standard technique to do so is kernel estimation as proposed by Diggle (1985). This technique can be seen as a generalisation of the histogram. Briefly, a Gaussian

kernel (say) with a given standard deviation h_S in space and h_T in time is centred at each of the points in a realisation of Ψ and their sum is reported. The choice of the bandwidths h_S and h_T is crucial. If one chooses bandwidths that are too small, the kernels concentrate around the observed earthquakes, which leads to spurious hot spots, especially for clustered patterns. If one takes bandwidths that are too large, the kernels spread out too far, which in turn results in the loss of structural details. Some asymptotic results for optimal choices of bandwidth are available (Chacón and Duong 2018; Van Lieshout 2020; Lo 2017), but practical rules of thumb seem to be lacking.

The risk of using the same bandwidths h_S and h_T at all points of Ψ is that, as with all one-size-fits-all solutions, this approach tends to oversmooth in regions that are rich in points, while at the same time it does not smooth enough in sparser regions. To overcome this drawback, we propose using an adaptive smoother as introduced for classic random variables by Abramson (1982). In a spatial context, such estimators were studied by Davies et al. (2018) for Poisson point processes and by Van Lieshout (2021, 2022) for general point processes with interaction between the points. The underlying idea is to weight the bandwidth at $(s, t) \in \Psi$ by a scalar $c(s, t)$ that is inversely proportional to $\sqrt{\lambda(s, t)}$. In this way, the kernel estimator adapts itself to the observed pattern: the bandwidth decreases in point-rich regions and is increased in regions with fewer points. The power 1/2 is motivated by asymptotics (Abramson 1982; Van Lieshout 2021). In practice, other powers could also be used.

Formally, set

$$\hat{\lambda}_A(x_0; h_S, h_T) = \sum_{(s,t) \in \Psi \cap (W_S \times W_T)} \frac{\kappa(H(c(s, t)h_S, c(s, t)h_T)^{-1}(x_0 - (s, t)))}{w((s, t), h_S, h_T) c(s, t)^3 h_S^2 h_T} \quad (1)$$

for $x_0 \in W_S \times W_T$. Here, $H(h_1, h_2)$, $h_1, h_2 > 0$, is a 3×3 diagonal matrix whose first two entries are h_1 and whose third entry is h_2 ,

$$c(s, t) = \left(\frac{\lambda(s, t)}{\left(\prod_{z \in \Psi \cap (W_S \times W_T)} \lambda(z) \right)^{1/N(\Psi \cap (W_S \times W_T))}} \right)^{-1/2}, \quad (2)$$

$w((s, t), h_S, h_T)$ is an edge correction term (see below) and κ is a symmetric probability density function, the kernel.

We make a few remarks. Firstly, because of the geometric averaging in (2), c is dimension-less. Secondly, since λ is unknown, $c(s, t)$ must be estimated. We will therefore replace $\lambda(s, t)$ in (2) by kernel estimates $\hat{\lambda}(s, t)$ using a global bandwidth (Steps 1 and 2 in Algorithm 1).

The adaptive kernel estimator (1) relies on spatial and temporal bandwidths, h_S and h_T . To select the bandwidths, several approaches can be taken. For example, asymptotic expansions of the mean integrated squared error provide explicit formulas for the optimal choice of h_S and h_T (Van Lieshout 2022). However, these formulas depend on the unknown intensity and its derivatives, making them unsuitable for direct practical application. For Poisson point processes, for which the likelihood is available in explicit form, a leave-one-out cross-validation approach might be taken.

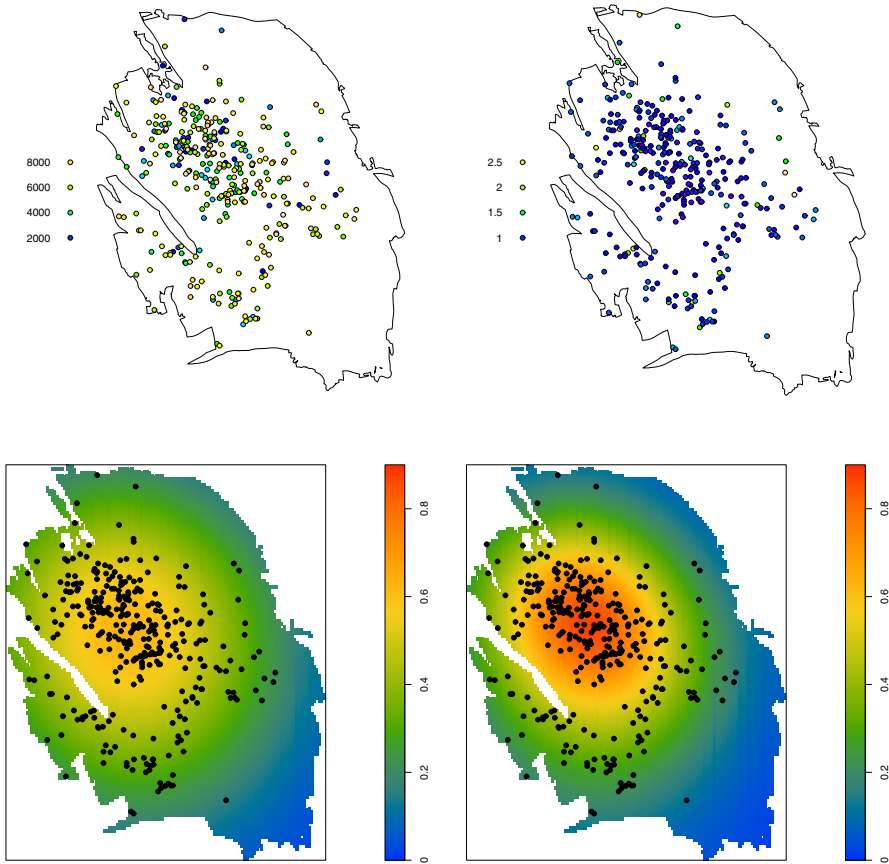


Fig. 6 Top row: Spatial projection of the 332 earthquakes of magnitude 1.5 or larger in the Groningen gas field that occurred in the period from 1 January 1995 up to 31 December 2021, marked by time (in days from 1 January 1995, leftmost panel) and by Abramson weights $c(s, t)$ (rightmost panel). Bottom row: Projections on space of global (leftmost panel) and adaptive (rightmost panel) kernel estimates of intensity (per square km) for the data in Fig. 1. The optimal pilot bandwidths are $h_{g,S} = 9.4$ km and $h_{g,T} = 182.5$ days, whilst $h_{a,S} = 6.9$ km and $h_{a,T} = 212.9$ days

From a numerical point of view, as the likelihood involves an integral over the spatio-temporal observation window, the approach is computationally costly. Indeed, the window must be discretised into a lattice, and for each lattice point and every potential bandwidth, the kernel estimator must be evaluated. Moreover, the Poisson assumption may not hold. Instead, we propose the following computationally easy algorithm that does not rely on any model assumptions. It relies on the Campbell–Mecke theorem (Chiu et al. 2013), which states that

$$\mathbb{E} \left[\sum_{x \in \Psi \cap (W_S \times W_T)} \frac{1}{\lambda(x)} \right] = \ell(W_S)\ell(W_T).$$

The right-hand side, the volume of the space–time window $W_S \times W_T$, is readily available. The left-hand side may be estimated from the data pattern ψ and an estimate $\hat{\lambda}(\cdot; h_S, h_T)$ of the intensity using a kernel estimator with bandwidths h_S and h_T . For good choices of bandwidth, the result should be close to the theoretical value. The idea is then to minimise the difference between the estimated left-hand side and $\ell(W_S)\ell(W_T)$ over h_S and h_T .

Algorithm 1 *Let ψ be a non-empty spatio-temporal point pattern that is observed in the window $W_S \times W_T$. Then*

1. Choose global bandwidths $h_{g,S}, h_{g,T}$ by minimising

$$\left| \sum_{x \in \psi \cap (W_S \times W_T)} \frac{1}{\hat{\lambda}(x; h_S, h_T)} - \ell(W_S)\ell(W_T) \right| \tag{3}$$

over $h_S, h_T > 0$ (with minimal $h_S^2 h_T$ in the case of multiple solutions), where

$$\hat{\lambda}(x; h_S, h_T) = \frac{1}{h_S^2 h_T} \sum_{y \in \psi \cap (W_S \times W_T)} \kappa \left(H(h_S, h_T)^{-1}(x - y) \right).$$

2. Calculate, for each $x \in \psi \cap (W_S \times W_T)$, the edge-corrected pilot estimator

$$\hat{\lambda}_c(x; h_{g,S}, h_{g,T}) = \frac{1}{h_{g,S}^2 h_{g,T}} \sum_{y \in \psi \cap (W_S \times W_T)} \frac{\kappa \left(H(h_{g,S}, h_{g,T})^{-1}(x - y) \right)}{w(y, h_{g,S}, h_{g,T})}$$

with local edge correction weights

$$w(y, h_{g,S}, h_{g,T}) = \frac{1}{h_{g,S}^2 h_{g,T}} \int_{W_S \times W_T} \kappa \left(H(h_{g,S}, h_{g,T})^{-1}(z - y) \right) dz.$$

3. Choose adaptive bandwidths $h_{a,S}, h_{a,T}$ by minimising

$$\left| \sum_{x \in \psi \cap (W_S \times W_T)} \frac{1}{\hat{\lambda}_A(x; h_S, h_T)} - \ell(W_S)\ell(W_T) \right|$$

over $h_S, h_T > 0$ (with minimal $h_S^2 h_T$ in the case of multiple solutions), where $\hat{\lambda}_A$ is given by (1) and (2) with $w \equiv 1$ upon plugging in the edge-corrected pilot estimator $\hat{\lambda}_c$ for λ .

Local edge correction weights, as suggested by Van Lieshout (2012), for the adaptive kernel estimator take the form

$$w(y, h_{a,S}, h_{a,T}) = \int_{W_S \times W_T} \frac{\kappa \left(H(\hat{c}(y)(h_{a,S}, h_{a,T}))^{-1}(z - y) \right)}{h_{a,S}^2 h_{a,T} (\hat{c}(y))^3} dz$$

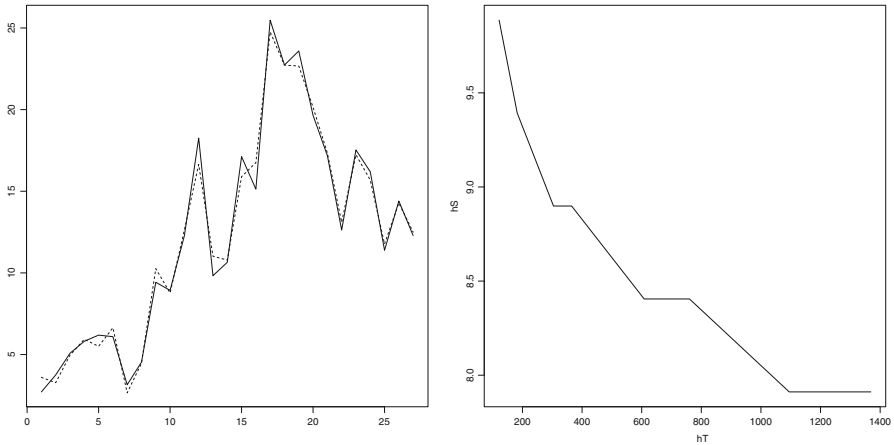


Fig. 7 Left: Expected number of earthquakes against time using global (broken line) and adaptive (solid line) kernel estimates for the data in Fig. 1 over 27 equal time intervals of approximately a year. The optimal pilot bandwidths are $h_{g,S} = 9.4$ km and $h_{g,T} = 182.5$ days, whilst $h_{a,S} = 6.9$ km and $h_{a,T} = 212.9$ days. Right: Curve of (h_T, h_S) for which the objective function (3) in Step 1 of Algorithm 1 is equal to zero. The spatial unit is a kilometre. The temporal unit is a day

and ensure that the integral of (1) over $W_S \times W_T$ is equal to the number of points in $\Psi \cap (W_S \times W_T)$. In selecting the bandwidth in Steps 1 and 3 of Algorithm 1, no edge correction is applied in order to obtain a clear optimum (Cronie and Van Lieshout 2018; Van Lieshout 2022).

The first and third steps are modifications to space–time of, respectively, the Cronie–Lieshout bandwidth selector (Cronie and Van Lieshout 2018) and the adaptive selector of Van Lieshout (2022) for purely spatial point processes. An important difference is that in space, one optimises over a single parameter, and usually, but not always, there is only one minimiser. In the space–time domain, the optimisation is done with respect to two parameters, h_S and h_T . Therefore, as a rule, a curve of minimisers is found in Steps 1 and 3 of Algorithm 1 (cf. Fig. 7). We pick the optimiser having the smallest scale $h_S^2 h_T$.

For our earthquake catalogue, the results are shown in Figs. 6 and 7. The minimal-scale minimisers of Eq. (3) are $h_{g,S} = 9.4$ km and $h_{g,T} = 182.5$ days. The corresponding edge-corrected projections on space and time are shown in, respectively, the bottom left panel in Fig. 6 and the broken line in the left panel in Fig. 7. They should be compared with the projections of the edge-corrected adaptive kernel estimate based on the minimisers $h_{a,S} = 6.9$ km and $h_{a,T} = 212.9$ days obtained in Step 3 of Algorithm 1 shown in the bottom right panel in Fig. 6 and the solid line in the left panel in Fig. 7. Note that $\hat{\lambda}_A$ attains higher values in the central reservoir than $\hat{\lambda}$, and lower values near the eastern border of the gas field. From a temporal perspective, the extremes in years with a large number of earthquakes are somewhat more pronounced for the adaptive kernel estimate. Figure 6 also depicts the Abramson weights $c(s, t)$. The weights are larger in the periphery and early in time, to compensate for the smaller number of earthquakes.

3.2 Inhomogeneous Spatio-Temporal K- and J-Functions

Having estimated the trend, we now turn our attention to the inter-point interactions. To quantify such interaction, information about the joint distributions of tuples of points is required, which is formalised by the higher-order analogues of the intensity function, the product densities $\lambda^{(n)}$. Heuristically, $\lambda^{(n)}((s_1, t_1), \dots, (s_n, t_n)) ds_1 dt_1 \cdots ds_n dt_n$ is the probability that Ψ places points at each of the infinitesimal regions $ds_i dt_i$ around $(s_i, t_i), i = 1, \dots, n$.

A spatio-temporal point process Ψ on $\mathbb{R}^2 \times \mathbb{R}$ is said to be intensity-reweighted moment stationary (IRMS) (Cronie and Van Lieshout 2015) if its product densities $\lambda^{(n)}$ of all orders exist, $\bar{\lambda} = \inf_{(s,t)} \lambda(s, t) > 0$ and, for all $n \geq 1, \xi_n$ is translation-invariant in the sense that

$$\xi_n((s_1, t_1) + (a, b), \dots, (s_n, t_n) + (a, b)) = \xi_n((s_1, t_1), \dots, (s_n, t_n))$$

for almost all $(s_1, t_1), \dots, (s_n, t_n) \in \mathbb{R}^2 \times \mathbb{R}$ and all $(a, b) \in \mathbb{R}^2 \times \mathbb{R}$. Here, ξ_n are the n -point correlation functions defined in terms of the $\lambda^{(n)}$ by setting $\xi_1 \equiv 1$ and for other n recursively by

$$\frac{\lambda^{(n)}((s_1, t_1), \dots, (s_n, t_n))}{\prod_{k=1}^n \lambda(s_k, t_k)} = \sum_{k=1}^n \sum_{D_1, \dots, D_k} \prod_{j=1}^k \xi_{|D_j|}((s_i, t_i) : i \in D_j), \quad (4)$$

where \sum_{D_1, \dots, D_k} is a sum over all possible k -sized partitions $\{D_1, \dots, D_k\}, D_j \neq \emptyset$, of the set $\{1, \dots, n\}$, and $|D_j|$ denotes the cardinality of D_j . For a Poisson point process, in which there are no correlations between the points, $\xi_n \equiv 0$ for $n \geq 2$. The point process is said to be second-order intensity-reweighted stationary (SOIRS) if the translation invariance holds up to $n = 2$ (Gabriel and Diggle 2009).

Various summary statistics exist to explore inter-point interactions. Write

$$S_{h_S}^{h_T} = \{(s, t) \in \mathbb{R}^2 \times \mathbb{R} : \|s\| \leq h_S, |t| \leq h_T\}$$

for the spatio-temporal cylinder with spatial radius h_S and temporal range h_T . It has volume $2\pi h_S^2 h_T$. Let Ψ be an IRMS spatio-temporal point process. Then, for $n = 1, 2, \dots$, the statistics

$$J_n(h_S, h_T) = \int_{S_{h_S}^{h_T}} \cdots \int_{S_{h_S}^{h_T}} \xi_{n+1}((0, 0), (s_1, t_1), \dots, (s_n, t_n)) \prod_{i=1}^n ds_i dt_i$$

quantify cumulative correlations up to ranges $h_S \geq 0$ in space and $h_T \geq 0$ in time. Multiple orders can be combined. For example, Van Lieshout (2011) proposed

$$J_{\text{inhom}}(h_S, h_T) = 1 + \sum_{n=1}^{\infty} \frac{(-\bar{\lambda})^n}{n!} J_n(h_S, h_T) \quad (5)$$

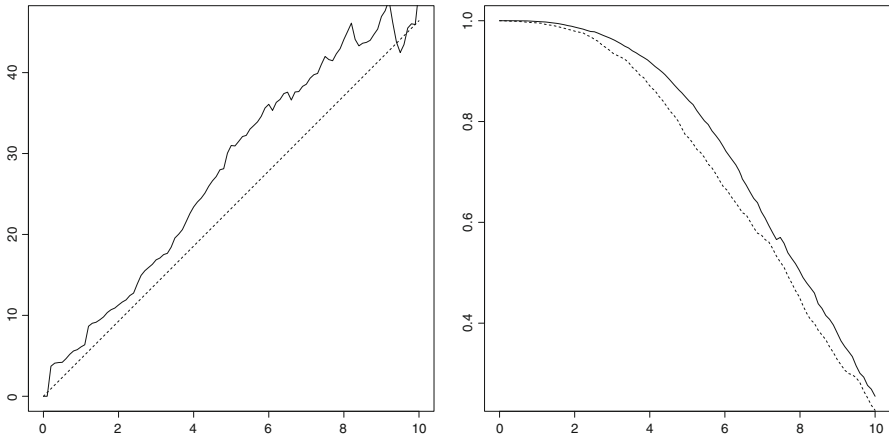


Fig. 8 Graphs of $(\hat{K}_{\text{inhom}}(r, 100r)/(2\pi))^{1/3}$ against r (solid line, leftmost panel), and graphs of $\hat{G}(1 - u_{r,100r}^y)$ (solid line, rightmost panel) and $\hat{G}^{!y}(1 - u_{r,100r}^y)$ (broken line, rightmost panel) against r for the data in Fig. 1. The spatial unit is a kilometre. The temporal unit is a day

for all spatial radii $h_S \geq 0$ and temporal ranges $h_T \geq 0$ for which the series is absolutely convergent. Values greater than 1 are indicative of repulsion, whereas values smaller than 1 suggest clustering. For further details, the reader is referred to Cronie and Van Lieshout (2015). Truncating at $n = 1$,

$$J_{\text{inhom}}(h_S, h_T) - 1 \approx -\bar{\lambda} \left(K_{\text{inhom}}(h_S, h_T) - \ell \left(S_{h_S}^{h_T} \right) \right)$$

in terms of the inhomogeneous K -function

$$K_{\text{inhom}}(h_S, h_T) = \int_{S_{h_S}^{h_T}} [\xi_2((0, 0), (s_1, t_1)) + 1] ds_1 dt_1$$

of Gabriel and Diggle (2009), which is well defined under the SOIRS assumption. Values greater or smaller than the volume of $S_{h_S}^{h_T}$ are indicative of, respectively, clustering and inhibition between points.

In practice, one must estimate these statistics based on a pattern observed in $W_S \times W_T$. The definition of the inhomogeneous J -function does not immediately suggest a suitable estimator. However, Cronie and Van Lieshout (2015) showed that, for all $y \in \mathbb{R}^2 \times \mathbb{R}$,

$$J_{\text{inhom}}(h_S, h_T) = \frac{G^{!y} \left(1 - u_{h_S, h_T}^y \right)}{G \left(1 - u_{h_S, h_T}^y \right)} \tag{6}$$

whenever well defined and the denominator is greater than zero. Here, the function u_{h_S, h_T}^y is defined by

$$u_{h_S, h_T}^y(s, t) = \frac{\bar{\lambda} 1\{\|a - s\| \leq h_S, |b - t| \leq h_T\}}{\lambda(s, t)}, \quad y = (a, b) \in \mathbb{R}^2 \times \mathbb{R},$$

and G is the generating functional of Ψ , that is,

$$G\left(1 - u_{h_S, h_T}^y\right) = \mathbb{E} \left[\prod_{(s, t) \in \Psi} \left(1 - \frac{\bar{\lambda} 1\{\|a - s\| \leq h_S, |b - t| \leq h_T\}}{\lambda(s, t)}\right) \right]$$

for $h_S, h_T \geq 0$, under the convention that empty products take the value 1. G^{1y} is defined similarly in terms of the conditional distribution of $\Psi \setminus \{y\}$ given that there is a point of Ψ at y . Intuitively, the numerator in (6) may be thought of as the probability that the nearest neighbour of a typical point y of the point process does not lie in a cylinder $S_{h_S}^{h_T}$ centred at y . Similarly, the denominator represents the probability that this cylinder centred at some arbitrary origin does not contain any points of the point process.

Being expectations, the numerator and denominator in Eq. (6) can be estimated in a straightforward manner. Write $W_S^{\ominus h_S}$ for the set of points in W_S that are at least h_S away from the border of W_S , and let $W_T^{\ominus h_T}$ be the similarly eroded temporal domain. Then, given a finite point grid $L \subseteq W_S \times W_T$,

$$\frac{1}{N\left(L \cap \left(W_S^{\ominus h_S} \times W_T^{\ominus h_T}\right)\right)} \sum_{l \in L \cap \left(W_S^{\ominus h_S} \times W_T^{\ominus h_T}\right)} \left[\prod_{x \in \Psi \cap \left(l + S_{h_S}^{h_T}\right)} \left(1 - \frac{\bar{\lambda}}{\lambda(x)}\right) \right]$$

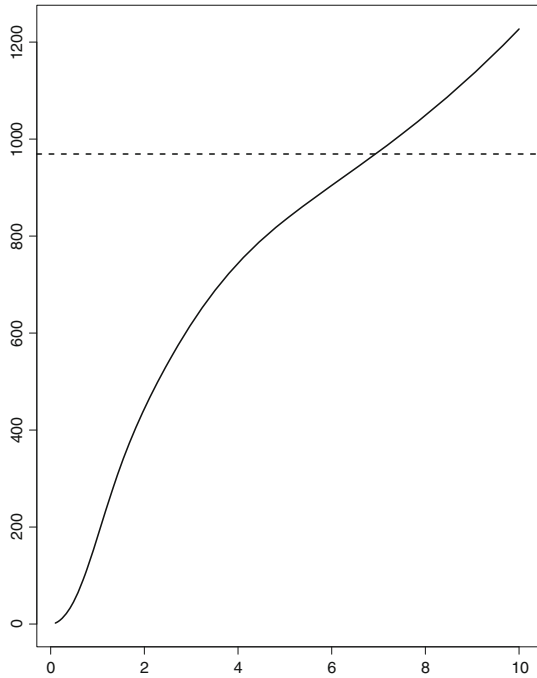
is an unbiased estimator of the denominator in (6). An unbiased estimator for the numerator is obtained analogously. Finally,

$$\hat{K}_{\text{inhom}}(h_S, h_T) = \frac{1}{\ell\left(W_S^{\ominus h_S}\right) \ell\left(W_T^{\ominus h_T}\right)} \sum_{x \in \Psi \cap \left(W_S^{\ominus h_S} \times W_T^{\ominus h_T}\right)} \sum_{x \neq y \in \Psi \cap \left(x + S_{h_S}^{h_T}\right)} \frac{1}{\lambda(x)\lambda(y)}$$

is an unbiased estimator of $K_{\text{inhom}}(h_S, h_T)$. Note that the intensity functions are unknown. A practical solution is to plug in their estimated counterparts (cf. Sect. 3.1).

For the earthquake catalogue depicted in Fig. 1, consider Fig. 8. The solid line in the leftmost panel in the figure is the graph of $(\hat{K}_{\text{inhom}}(r, 100r)/(2\pi))^{1/3}$ for the earthquake data. It lies above the graph of the same function for a Poisson point process (shown as a broken line in the leftmost panel in the figure), suggesting attraction between the points. In the rightmost panel in Fig. 8, the graph of $\hat{G}(1 - u_{r, 100r}^y)$ estimated from the data lies above that of $\hat{G}^{1y}(1 - u_{r, 100r}^y)$, which confirms the suggested clustering.

Fig. 9 The graph of $\sum_{(s,t) \in \phi} \hat{\lambda}_A((s,t), \phi; h)^{-1}$ as a function of bandwidth h (solid line, in kilometres). The dotted horizontal line is drawn at 969.2445, the area of the gas field in square kilometres. The pilot bandwidth is $h_{g,S} = 7.6$ km



4 Explanatory Variables

Gas production and pore pressure (cf. Sects. 2.4 and 2.5) may well have an effect on the earthquake rates. However, they are measured only for a limited number of locations and times. Therefore, in this section, we discuss how to calculate appropriate values for the entire space–time domain.

4.1 Non-parametric Smoothing of Gas Production Values

Monthly production values are available for the 29 production wells shown by a cross in Fig. 2 over the time period from 1956 to 2021. The gas extracted from the Eemskanaal-13 borehole is listed separately from the other ones at the Eemskanaal site. This is because the Eemskanaal-13 pipe is not drilled vertically but bends underground in such a way that it is depleting the Harkstede block (cf. Fig. 3). Therefore, we assign its production to the location of the observation well at Harkstede (Alternatively, one may use the deviation data from TNO (2022)). Thus, we end up with a spatial pattern of 30 points: the locations of the 29 production wells as well as the Harkstede proxy for the Eemskanaal-13 pipe.

In order to smooth out the production, an adaptive kernel method similar to that proposed in Sect. 3 for the intensity function can be used. More specifically, write $V(s, t) \geq 0$ for the volume of gas produced at a well at s in month t , and write ϕ for the well pattern. We then use the spatial counterpart of Algorithm 1 proposed by

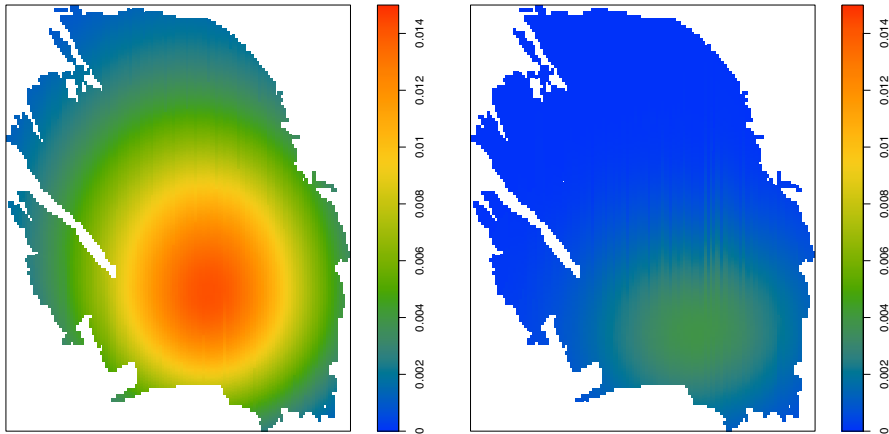


Fig. 10 Smoothed monthly gas production over January 2012 (leftmost panel) and January 2021 (rightmost panel) in Nbcm per square kilometre. The pilot bandwidth is $h_{g,S} = 7.6$ km, whilst $h_{a,S} = 6.9$ km

Van Lieshout (2021, 2022) to select bandwidths $h_{g,S}$ and $h_{a,S}$ and set, for $s_0 \in W_S$, $t_0 \in W_T$,

$$\hat{\lambda}_V((s_0, t_0), \phi) = \sum_{(s,t) \in \phi} \frac{V(s, t) 1\{t_0 \in i(t)\}}{w(s, h_{a,S}) h_{a,S}^2 \hat{c}(s; h_{g,S})^2 \ell(i(t))} \kappa\left(\frac{s_0 - s}{h_{a,S} \hat{c}(s; h_{g,S})}\right)$$

based on spatial kernel κ and using local edge correction weights w (cf. Sect. 3). In time, $V(s, t)$ is spread evenly over the $\ell(i(t))$ days in the month $i(t) \subset W_T$ indexed by t . Since

$$\begin{aligned} \int_{W_S \times W_T} \hat{\lambda}_V((s, t), \phi) ds dt &= \sum_{(s,t) \in \phi} \frac{V(s, t)}{w(s, h_{a,S})} \int_{W_S} \frac{1}{h_{a,S}^2 \hat{c}(s; h_{g,S})^2} \kappa\left(\frac{s_0 - s}{h_{a,S} \hat{c}(s; h_{g,S})}\right) ds_0 \\ &= \sum_{(s,t) \in \phi} V(s, t), \end{aligned}$$

the total production is preserved.

For the data discussed in Sect. 2.4, Algorithm 1 yields a pilot bandwidth $h_{g,S} = 7.6$ km and $h_{a,S} = 6.9$ km. We illustrate Step 3 of Algorithm 1 in Fig. 9. The solid line is the graph of the sum of inverse point estimates of intensity, and the horizontal dotted line indicates the volume of W_S . The point of intersection defines $h_{a,S}$, here 6.9 km. The resulting gas production maps for January 2012, before legislation to phase out gas extraction came into effect, and for January 2021 are shown in Fig. 10. The decrease in extracted volume is evident. Moreover, the remaining production is located mostly in the southern part of the gas field.

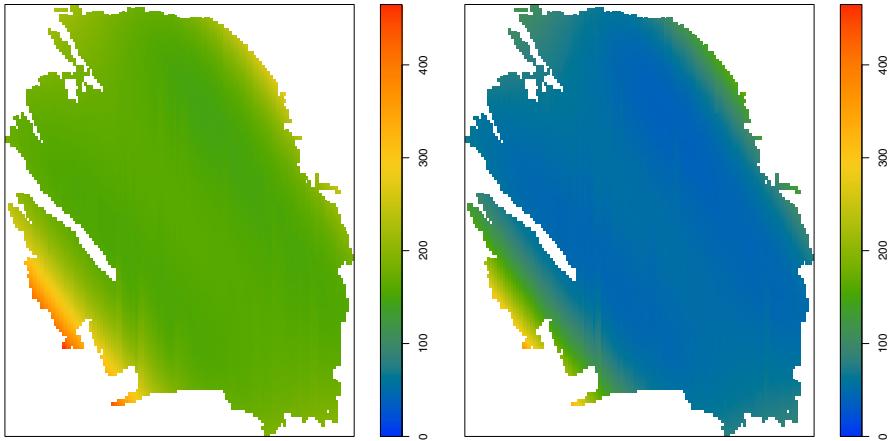


Fig. 11 Estimated pore pressure maps at midnight on 1 January in the years 1995 (leftmost panel) and 2022 (rightmost panel) in bara

4.2 Regression Analysis for Pore Pressure

From Fig. 5, it is clear that the pore pressure in the reservoir is decreasing over time. The trend is similar for most wells, with some exceptions, for example, because of large faults or at the periphery of the field. Recall from Sect. 2 that our earthquake catalogue contains tremors from 1 January 1995 onwards. If we wish to use the pore pressure as an explanatory variable in a monitoring model, we need their values over the same time period. Thus, our goal in this section is to perform a regression analysis using only the 352 pore pressure measurements from 1 January 1995 and later.

Suppose that the observed pore pressures can be seen as realisations of random variables $X(s, t)$ that can be decomposed in a trend term $m((s, t); \beta)$ and noise $E(s, t)$ as

$$X(s, t) = m((s, t); \beta) + E(s, t).$$

We assume that the measurement errors $E(s, t)$ are independent and normally distributed with a mean of zero and variance of σ^2 . For the trend, we fit a polynomial in space and time,

$$\begin{aligned} m((s, t); \beta) = & \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 (s - s_0)_1 + \beta_5 (s - s_0)_2 + \beta_6 (s - s_0)_1^2 \\ & + \beta_7 (s - s_0)_1 (s - s_0)_2 + \beta_8 (s - s_0)_2^2 + \beta_9 (s - s_0)_1^3 \\ & + \beta_{10} (s - s_0)_1^2 (s - s_0)_2 + \beta_{11} (s - s_0)_1 (s - s_0)_2^2 + \beta_{12} (s - s_0)_2^3 \\ & + \beta_{13} (s - s_0)_1^4 + \beta_{14} (s - s_0)_1^3 (s - s_0)_2 + \beta_{15} (s - s_0)_1^2 (s - s_0)_2^2 \\ & + \beta_{16} (s - s_0)_1 (s - s_0)_2^3 + \beta_{17} (s - s_0)_2^4 + \beta_{18} t (s - s_0)_1 + \beta_{19} t (s - s_0)_2 \\ & + \beta_{20} t (s - s_0)_1^2 + \beta_{21} t (s - s_0)_1 (s - s_0)_2 + \beta_{22} t (s - s_0)_2^2 + \beta_{23} t (s - s_0)_1^3 \\ & + \beta_{24} t (s - s_0)_1^2 (s - s_0)_2 + \beta_{25} t (s - s_0)_1 (s - s_0)_2^2 + \beta_{26} t (s - s_0)_2^3 \end{aligned}$$

Table 1 Analysis of variance table for pore pressure

	Sum Sq	Pr(> F)
$\beta_1 + \beta_2 t$	234,611	2.2e-16
$\beta_3 t^2$	1,976	6.085e-09
$\beta_4(s - s_0)_1 + \beta_5(s - s_0)_2$	15,261	2.2e-16
$\beta_6(s - s_0)_1^2 + \dots + \beta_8(s - s_0)_2^2$	58,675	2.2e-16
$\beta_9(s - s_0)_1^3 + \dots + \beta_{12}(s - s_0)_2^3$	12,153	2.2e-16
$\beta_{13}(s - s_0)_1^4 + \dots + \beta_{17}(s - s_0)_2^4$	25,106	2.2e-16
$\beta_{18}t(s - s_0)_1 + \beta_{19}t(s - s_0)_2$	1,502	2.189e-06
$\beta_{20}t^2(s - s_0)_1^2 + \dots + \beta_{22}t^2(s - s_0)_2^2$	1,773	1.041e-06
$\beta_{23}t^3(s - s_0)_1^3 + \dots + \beta_{26}t^3(s - s_0)_2^3$	1,148	0.0004642
Residual sum of squares	18,052	

A first-order linear model in time is updated with a second-order term, first-order up to fourth-order polynomials in space, and first-order up to third-order space-time interaction terms. The second column lists the sum of squares for the coefficients in the first column, and the third column quantifies the significance as the *p* value of the *F* test

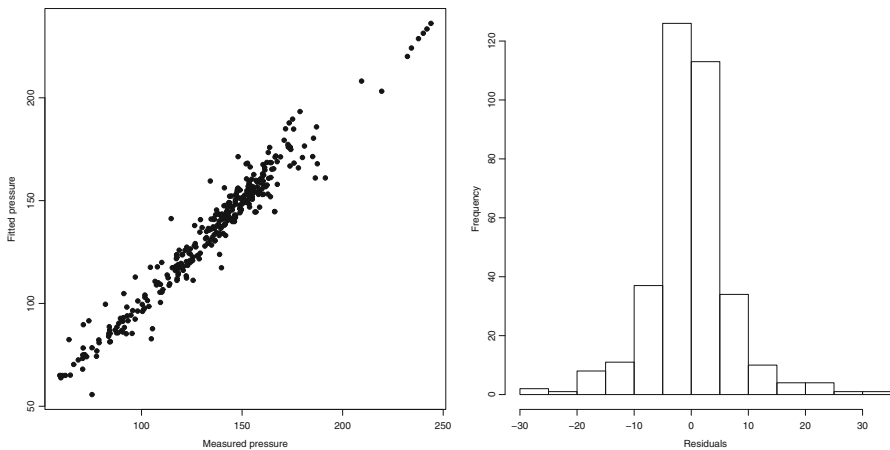


Fig. 12 Observed pore pressure values in bara against fitted values (leftmost panel) and the histogram of residuals (rightmost panel)

where we centre the spatial locations at $s_0 = ((s_0)_1, (s_0)_2) = (750, 5,900)$ in the UTM system (zone 31, in kilometres) and take a day as the temporal unit counting from 1 January 1995. The parameter $\beta = (\beta_1, \dots, \beta_{26})$ can be estimated by minimising the sum of squared residuals (the difference between the observed and the fitted pore pressure values). The analysis of variance table is given in Table 1.

The fitted pore pressure maps at midnight on 1 January in the years 1995 and 2022 are given in Fig. 11. One may note the elevated values in the Harkstede block in the south-west as well as those in the northern border regions.

To validate the model, the actual pore pressure measurements are plotted against their fitted values in the leftmost panel in Fig. 12. The graph seems reasonably close

to a straight line. The histogram of the residuals is shown in the rightmost panel in Fig. 12. Most residuals are quite small, and the histogram is centred around zero, with an estimated standard deviation $\hat{\sigma} = 7.17$ bara.

5 Discussion and Further Work

In this paper, we carried out an exhaustive second-order exploratory analysis of the spatio-temporal point pattern of earthquakes recorded in the Groningen gas field since January 1995. To do so, we needed to develop a new methodology. We proposed an adaptive kernel smoothing technique for estimating the intensity function and suggested a practical algorithm for selecting the spatial and temporal bandwidths. The estimated intensity function was then plugged into state-of-the-art inhomogeneous summary statistics to quantify the degree of clustering in the earthquake catalogue. We also applied our new adaptive kernel smoothing technique to monthly gas production figures. Finally, we performed a regression analysis on pore pressure data for the gas field.

In the rate-and-state models (Candela et al. 2019; Dempsey and Suckale 2017; Richter et al. 2020) that can be seen as the state of the art in modelling the seismic hazard (Kühn et al. 2022) and that are being used for planning, the earthquake intensity λ (the rate) is assumed to be inversely proportional to a state variable Γ , that is,

$$\lambda(s, t) \propto \Gamma(s, t)^{-1}, \quad (s, t) \in W_S \times W_T.$$

The state variable $\Gamma(s, t)$ is defined by the ordinary differential equation

$$d\Gamma(s, t) = \alpha [dt + \Gamma(s, t)dX(s, t)],$$

where X is the pore pressure at spatial location s and time t . Multiplying both sides by $\exp(-\alpha X(s, t))$, it follows that

$$\frac{d}{dt} \left[\Gamma(s, t)e^{-\alpha X(s,t)} \right] = \alpha e^{-\alpha X(s,t)}. \tag{7}$$

The Euler discretisation reads

$$\Gamma(s, t + \Delta) = (\Gamma(s, t) + \Delta\alpha) \exp[\alpha(X(s, t + \Delta) - X(s, t))], \quad s \in W_S,$$

upon discretising W_T in time steps of length Δ .

In the rate-and-state model, the earthquakes constitute a Poisson point process with intensity function λ , possibly modified by a fault map. The parameter α and the initial state $\Gamma(s, 0)$ (as well as the proportionality constant and any parameters associated with the fault map) are treated as unknowns and can be estimated, for example, by the maximum likelihood method.

Based on the exploratory analysis in Sects. 3 and 4, the rate-and-state model can be criticised on several points. As we saw in Sect. 3, the earthquake pattern exhibits

clustering. Since by definition the points in any Poisson point process do not interact with one another, the apparent clustering of earthquakes cannot be described by the rate-and-state model. Secondly, the pressure values are assumed to be known everywhere, in practice by interpolation of the measurements. Proceeding in this way, the uncertainty in the interpolations is ignored. It would be better to treat the $X(s, t)$ as a random field. Lastly, the varying gas extraction is not taken into account.

Motivated by the above considerations, we propose the following model. Set $\Gamma(s, 0) \equiv \gamma_0$ and iterate

$$\Gamma(s, t + \Delta) = (\Gamma(s, t) + \alpha \Delta) \exp[\alpha(m(s, t + \Delta) - m(s, t))] \exp[\alpha(E(s, t + \Delta) - E(s, t))]$$

where m and E are as in Sect. 4.2. Gas production and random effects can be included in the following way. Let Ψ be a Cox process (Chiu et al. 2013) on $W_S \times W_T$ with driving random measure

$$\Lambda(s, t) = \exp\left[\theta_1 + \theta_2 \tilde{V}(s, t) + U(s, t)\right] \frac{\gamma_0}{\Gamma(s, t)} \quad (8)$$

where θ_1, θ_2 are real-valued parameters, $\tilde{V}(s, t)$ is the gas extracted at s during the year preceding time t , that is, the integral of $\hat{\lambda}_V((s, \cdot), \phi)$ (cf. Sect. 4.1) over this period, and $U(s, t)$ is a correlated Gaussian field that accounts for random effects. Instead of or in addition to \tilde{V} , other covariates are easily incorporated. For example, one might add the cumulative gas extraction, subsidence or compaction information, the distance to the nearest fault or fault density terms to the exponent in Eq. (8). Monitoring can then be based on the posterior distribution of Λ or, equivalently, Γ and U , given the recorded earthquakes. The implementation requires careful use of Markov chain Monte Carlo techniques and is the topic of our ongoing research.

Acknowledgements We are grateful to Professor Van Dinther for expert advice and to Mr Van Eijs for providing us with data on gas production. We thank two anonymous referees for their constructive comments.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Data Cleaning

Errors in Recorded Dates

There are some anomalies in the recorded dates. For example, in 2015, the entry 10/7/2015 should be interpreted as 7/10/2015, July the tenth. There are eight other such errors: May 10, 2016, May 6–7, 2017, June 5 and 7, 2017, April 9 and 11, 2018, and November 6, 2018.

Missing Coordinates

Since the coordinates of one of the stations are not listed in the production plans (cf. Sect. 2.3) and are therefore unknown, we omit the corresponding 28 pore pressure measurements from consideration. We also disregard the two measurements from an observation well located outside the Groningen gas field.

The Eemskanaal-13 well is the only one depleting a peripheral field, the so-called Harkstede block. Moreover, as can be seen from Fig. 4, it is extracting less gas than other wells. According to an expert, setting its location to either that of the Eemskanaal plant or to that of the installation at Harkstede would lead to biases, and it is therefore preferable to ignore the seven observations for Eemskanaal-13 altogether.

Invalid Measurements

The NAM file mentions 10 cases in which the observations are invalid for various reasons. We disregard these measurements.

References

- Abramson IS (1982) On bandwidth variation in kernel estimates—a square root law. *Ann Stat* 10:1217–1223
- Baki Z, Van Lieshout MNM (2022) The influence of gas production on seismicity in the Groningen field. In: Proceedings of the 10th international workshop on spatio-temporal modelling METMA X, pp 163–167
- Bourne SJ, Oates SJ, Van Elk J (2018) The exponential rise of induced seismicity with increasing stress levels in the Groningen gas field and its implications for controlling seismic risk. *Geophys J Int* 213:1693–1700
- Candela T, Osinga S, Ampuero JP, Wassing B, Pluymaekers M, Fokker PA, Van Wees JD, De Waal HA, Muntendam-Bos AG (2019) Depletion-induced seismicity at the Groningen gas field: Coulomb rate-and-state models including differential compaction effect. *J Geophys Res Solid Earth* 124:7081–7104
- Chacón JE, Duong T (2018) *Multivariate kernel smoothing and its applications*. CRC Press, Boca Raton
- Chiu SN, Stoyan D, Kendall WS, Mecke J (2013) *Stochastic geometry and its applications*, 3rd edn. Wiley, Chichester
- Cronie O, Van Lieshout MNM (2015) A J -function for inhomogeneous spatio-temporal point processes. *Scand J Stat* 42:562–579
- Cronie O, Van Lieshout MNM (2018) A non-model based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika* 105:455–462
- Davies TM, Flynn CR, Hazelton ML (2018) On the utility of asymptotic bandwidth selectors for spatially adaptive kernel density estimation. *Stat Probab Lett* 138:75–81
- De Zeeuw Q, Geurtsen L (2018) Groningen dynamic model update 2018. Report NAM

- Dempsey D, Suckale J (2017) Physics-based forecasting of induced seismicity at Groningen gas field, The Netherlands. *Geophys Res Lett* 22:7773–7782
- Diggle P (1985) A kernel method for smoothing point process data. *Appl Stat* 34:138–147
- Dost B, Goutbeek F, Van Eck T, Kraaijpoel D (2012) Monitoring induced seismicity in the North of The Netherlands: status report 2010. Scientific report KNMI, WR 2012–03
- Gabriel E, Diggle PJ (2009) Second-order analysis of inhomogeneous spatio-temporal point process data. *Stat Neerl* 63:43–51
- Geerdink E (2014) Modeling the induced earthquakes in Groningen as a Poisson process using GLM and GAM. BSc thesis, University of Groningen
- Hettema MHH, Jaarsma B, Schroot BM, Van Yperen GCN (2017) An empirical relationship for the seismic activity rate of the Groningen gas field. *Neth J Geosci* 96:149–161
- KNMI (2022) www.knmi.nl/kennis-en-datacentrum/dataset/aardbevingscatalogus. Downloaded April 2022
- Kühn D, Hainzl S, Dahm T, Richter G, Vera Rodriguez I (2022) A review of source models to further the understanding of the seismicity of the Groningen field. *Neth J Geosci* 101:e11
- Lo PH (2017) An iterative plug-in algorithm for optimal bandwidth selection in kernel intensity estimation for spatial data. PhD Thesis, Technical University of Kaiserslautern
- NAM (2022a) www.nam.nl/gas-en-olie/groningen-gasveld/winningsplan-groningen-gasveld.html. Downloaded April 2022
- NAM (2022b) www.nam-feitenencijfers.data-app.html/gasdruk.html. Downloaded April 2022
- Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83–401:9–27
- Post RAJ, Michels MAJ, Ampuero JP, Candela T, Fokker PA, Van Wees JD, Van der Hofstad RW, Van den Heuvel ER (2021) Interevent-time distribution and aftershock frequency in non-stationary induced seismicity. *Sci Rep* 11:3540
- Richter G, Hainzl S, Dahm T, Zöller G (2020) Stress-based statistical modeling of the induced seismicity at the Groningen gas field, The Netherlands. *Environ Earth Sci* 79:252
- Sijacic D, Pijpers F, Nepveu M, Van Thienen-Visser K (2017) Statistical evidence on the effect of production changes on induced seismicity. *Neth J Geosci* 96:27–38
- TNO (2022) www.nlog.nl/bestanden-interactieve-kaart. Downloaded April 2022
- Trampert J, Benzi R, Toschi F (2022) Implications of the statistics of seismicity recorded within the Groningen gas field. *Neth J Geosci* 101:e9
- Van Hove E, Van Lingen R, Riemens S (2015) Geïnduceerde aardbevingen in gasveld Groningen. Een statistische analyse. BSc thesis, University of Twente
- Van Lieshout MNM (2011) A J -function for inhomogeneous point processes. *Stat Neerl* 65:183–201
- Van Lieshout MNM (2012) On estimation of the intensity function of a point process. *Methodol Comput Appl Probab* 14:567–578
- Van Lieshout MNM (2020) Infill asymptotics and bandwidth selection for kernel estimators of spatial intensity functions. *Methodol Comput Appl Probab* 22:995–1008
- Van Lieshout MNM (2021) Infill asymptotics for adaptive kernel estimators of spatial intensity. *Aust N Z J Stat* 63:159–181
- Van Lieshout MNM (2022) Non-parametric adaptive bandwidth selection for kernel estimators of spatial intensity functions. [arXiv:2210.11902](https://arxiv.org/abs/2210.11902)
- Vlek C (2019) Rise and reduction of induced earthquakes in the Groningen gas field, 1991–2018: statistical trends, social impacts, and policy change. *Environ Earth Sci* 78:1–14