Ludwig-Maximilians-Universität München
Fakultät für Mathematik, Informatik und Statistik
Institut für Statistik

Spatio-Temporal Infectious Disease Epidemiology based on Point Processes

Sebastian Meyer

Master's Thesis

Supervision:
Prof. Dr. Ludwig Fahrmeir
Michael Höhle, Ph.D.

18 December 2009
(Revised edition as of 29 January 2010)

Abstract: In this Master's Thesis, a novel combination of point process models continuous in space-time is proposed for infectious disease data. Modelling is driven by the conditional intensity function, which enables a step towards a regression framework for self-exciting spatio-temporal point processes. The model is an extension of the discrete space additive-multiplicative conditional intensity model proposed by Höhle (2009a), and also borrows from earthquake research, especially from the formulation of the space-time ETAS model in Ogata, Katsura & Tanemura (2003). Estimation is performed by means of full maximum likelihood, which for general point processes in space requires the evaluation of two-dimensional integrals. Therefore, various methods of numerical integration are investigated as a prerequisite for maximum likelihood inference.
The particular application of interest is the stochastic modelling of the transmission dynamics of the two most common meningococcal strains observed in Germany in 2002–2008. The application showed that the proposed model and its estimation by the provided thoroughly `R` implementation are applicable and valuable for the analysis of spatio-temporal infectious disease data.

Keywords: stochastic epidemic modelling, spatio-temporal point processes, conditional intensity function, two-dimensional numerical integration, invasive meningococcal disease

---

Zusammenfassung: In dieser Masterarbeit wird eine neuartige Modellkombination von stetigen, räumlich-zeitlichen Punktprozessen für Daten über das Auftreten von Infektionskrankheiten vorgeschlagen. Die Modellierung beruht auf der bedingten Intensitätsfunktion, die einen Schritt in Richtung eines Regressionskontextes für „selbsterregende" räumlich-zeitliche Punktprozesse ermöglicht. Das Modell ist eine Erweiterung des von Höhle (2009a) vorgeschlagenen, räumlich diskreten, additiv-multiplikativen Modells für die bedingte Intensitätsfunktion, und ist ebenfalls inspiriert durch stochastische Erdbeben-Modelle, insbesondere durch die Formulierung des räumlich-zeitlichen ETAS-Modells in Ogata et al. (2003). Die Modellschätzung basiert auf voller Likelihood-Inferenz, die für allgemeine Punktprozesse im Raum die Berechnung von zweidimensionalen Integralen benötigt. Als Voraussetzung für Maximum Likelihood-Inferenz werden deshalb auch verschiedene Methoden der numerischen Integration untersucht.
Motivierendes Anwendungsbeispiel ist die stochastische Modellierung der beobachteten Ausbreitungsdynamik der zwei häufigsten Meningokokkenstämme in Deutschland in den Jahren 2002–2008. Die Anwendung zeigte, dass das vorgeschlagene Modell und seine Schätzung mittels der sorgfältigen Implementierung in `R` geeignet und wertvoll für die Analyse von räumlich-zeitlichen Daten zu Infektionskrankheiten sind.

Schlagwörter: Stochastische Modellierung von Epidemien, Räumlich-zeitliche Punktprozesse, Bedingte Intensitätsfunktion, Zweidimensionale numerische Integration, Invasive Meningokokken-Erkrankung

---

Résumé : Dans ce mémoire, une nouvelle combinaison de modèles de processus ponctuels continus dans l'espace-temps est proposée concernant des données sur les maladies infectieuses. La modélisation provient de l'intensité conditionnelle qui permet de se diriger vers un cadre de régression pour des processus ponctuels spatio-temporels et autoexcités. Le modèle est une extension du modèle additif-multiplicatif de l'intensité conditionnelle en espace discret proposé par Höhle (2009a), et il s'inspire aussi de la recherche sur la modélisation des séismes, en particulier de la formulation spatio-temporelle du modèle ETAS dans Ogata et al. (2003). L'estimation est réalisée à partir de l'estimation du maximum de vraisemblance complète. De manière générale, pour les processus spatiaux, cela exige l'évaluation d'intégrales bidimensionnelles. Par conséquent, diverses méthodes d'intégration numérique sont étudiées en tant que condition indispensable pour l'inférence du maximum de vraisemblance.
L'application d'intérêt particulière concerne la modélisation stochastique de la dynamique de propagation des deux souches de méningocoques les plus communes observée en Allemagne au cours des années 2002–2008. Les recherches ont montré que le modèle proposé avec l'estimation obtenue par la soigneuse mise en œuvre sous `R` sont applicables et utiles pour l'analyse spatio-temporelle des données sur les maladies infectieuses.

Mots-clés: modélisation stochastique des épidémies, processus ponctuels spatio-temporels, intensité conditionelle, intégration numérique de deux dimensions, méningococcie invasive

Preface

This Master's Thesis has been registered on 3 June 2009 and submitted on 18 December 2009 to the Department of Statistics of the Ludwig-Maximilians-Universität München.

I wish to thank Michael Höhle, Ph.D., very much for his outstanding mentoring, and I am grateful to Prof. Dr. Ludwig Fahrmeir for enabling this thesis and for several valuable discussions. Dr. Johannes Elias from the German National Reference Centre for Meningococci in Würzburg provided me with the data on the cases of meningococcal disease analysed in this thesis. He also supported my understanding of the data and the microbiological aspects of the disease, which is greatly appreciated.

In the context of this work I implemented several functions for the application of the presented methodology in `R` – a free software environment for statistical computing and graphics (R Development Core Team, 2009a). In doing so I obeyed most of the "R Coding Conventions" of Bengtsson (2005) and the handbook "Writing R Extensions" of the R Development Core Team (2009b). During the implementation of the thesis' methodology I ran across a bug in `R` and one in the `spatstat` package, which hindered my functions of working correctly. I would like to thank Prof Brian Ripley and Prof Adrian Baddeley, respectively, for the rapid bug fixes. As a consequence, the functions on the attached CD require the latest `R` version `2.10.1` and the `spatstat` version `1.17-0` at least.

The thesis has been compiled with LaTeX in conjunction with the `R` function `Sweave` (Leisch, 2002) and the driver `weaver` (Falcon, 2009) for handy integration of code and generated graphics. The computations were all run on the same 64-bit Linux machine with a 1.8 GHz CPU and 1 024 MB of RAM, which should be considered for the runtimes mentioned in the text. The `pdf` version of this Master's Thesis, as well as all implementations are available from the CD attached as Appendix A.

The comments of my faithful lectors Valentin Wimmer, Julia Meyer, Steffen Kobilke, Sebastian Pöhlmann, Hans-Werner Kammer, Martin Dittmer, and Aurore Puy are appreciated as well. Last but not least I thank my wife Julia Meyer and my parents Heike & Carl Meyer very much for their continuous mental support and understanding during the whole writing period.

Contents

1 Introduction

Infectious diseases are a matter of tremendous public concern especially gaining attention in case of outbreaks. For instance, each winter waves of influenza remind people of the ease of infection by person-to-person contact, Norovirus outbreaks often lead to temporarily closed schools, and the detection of new highly infectious or pathogenic viruses like the "bird flu" (influenza subtype A/H5N1) or the current "swine flu" (influenza subtype A/H1N1) concern the World Health Organisation and its members trying to prevent a pandemic like the Spanish flu in 1918. There are also infectious diseases in animals not threatening humans directly like the classical swine fever or the foot and mouth disease, which often lead to the preventive culling of animals in susceptible farms in the neighborhood of infected ones, and cause an enormous economic damage. Collaboration of health professionals, veterinaries, geneticists, epidemiologists, statisticians and many others is indispensable for controlling and understanding the evolution of such epidemics.

This Master's Thesis concentrates on stochastic modelling and associated inference for spatio-temporal epidemic data. The aim is to establish a regression framework, where the transmission dynamics of an infectious disease and its dependency on covariates can be quantified. Specifically, the statistical methodology is motivated by the invasive meningococcal disease (IMD), outbreaks of which are supposed to be linked to waves of influenza. Meningococci also seem to exhibit different degrees of infectivity according to the specific type of pathogen eliciting the disease. The available data stem from an epidemiological database with high geographical and temporal resolution on occurence of the disease, which are becoming more widely available. Herein, the cases of disease are neither aggregated temporally nor within administrative borders, which otherwise is often a consequence of the data sampling. This motivates the use of modelling techniques which honour the high resolution. Point processes constitute an appropriate framework for the underlying stochasticity as they represent the individual occurence of events in space and/or time. The observations may be correlated spatially as well as temporally, which differs from classical i.i.d. statistics and is essential for the representation of epidemics.

This Master's Thesis is organised as follows: the remainder of this introduction will in Section 1.1 give an overview of invasive meningococcal disease (IMD) and present the available data in Section 1.2. Chapter 2 moves towards point process modelling. It first sketches the theory of point processes, as I do not assume the reader to have in-depth knowledge in this

field of statistics, and then describes a preliminary spatio-temporal point process model for a finite and fixed set of possible event locations. As a prerequisite for likelihood inference of point process models continuous in space-time, methods of numerical integration will be presented and compared in Chapter 3. Chapter 4 then investigates a novel combination of point process models with full likelihood inference aiming at a spatio-temporal regression framework for infectious disease data. The final general model class is then applied to the IMD data in Chapter 5. The thesis concludes with a discussion of the results of data analysis and the presented methodology.

1.1 Invasive Meningococcal Disease (IMD)

Throughout the thesis, I will examine a specific data set (see Section 1.2 below), which serves as a motivating example of application for the presented methodology. It contains cases of invasive meningococcal disease (IMD) recorded by the German National Reference Centre for Meningococci (Nationales Referenzzentrum für Meningokokken, NRZM) during the years 2002–2008.

IMD is a life-threatening human disease involving meningitis (50% of cases), septicemia (5 to 20 percent) and/or pneumonia (5 to 15 percent) caused by the infection with the bacterium *Neisseria meningitidis* also termed meningococcus (Rosenstein, Perkins, Stephens, Popovic & Hughes, 2001, p. 1382). The clinical picture ranges from temporary asymptomatic bacteraemia to a fulminant septic course of disease, which can lead to death within a few hours (Hellenbrand, Vogel & Elias, 2008, p. 266). Meningococci can be transmitted airborne (e.g. by coughing or sneezing) or by other mucous secretions from infected humans, where they colonise the nasopharynx, their only natural reservoir (Rosenstein et al., 2001, p. 1379). The incubation time is around four days. A special characteristic of the bacterium is that 5 to 10 percent of adults are *asymptomatic* nasopharyngeal carriers of meningococcal strains. Actually, most meningococci are commensal in humans, but only a few isolates are virulent and cause invasive disease. “Before the 1920s, meningococcal disease was fatal in up to 70 percent of cases”, whereas “the overall case fatility rates have remained relatively stable over the past 20 years, at 9 to 12 percent” (Rosenstein et al., 2001, p. 1382). In Germany within the years 2001–2007, the overall lethality was 8.4% (Hellenbrand et al., 2008, p. 266). This is also a success of the immediate administering of effective antibiotics such as penicillin. However, up to 40 percent of patients with meningococcal sepsis do not survive the disease (Rosenstein et al., 2001, p. 1382), especially those who develop the Waterhouse-Friderichsen syndrome (acute adrenal insufficiency).

As determined by the German Infection Protection Act (Infektionsschutzgesetz, IfSG, §6), IMD is a notifiable disease. Health authorities mandatorily transmit every case to the Robert

Koch institute (RKI), which surveys the spread of the disease (IfSG, §4). Based on the case numbers from RKI the rates of IMD during the years 2001–2008 in Germany are at approximately 0.53–0.95 cases per 100 000 population per year, where the rate has roughly been decreasing over the years (cf. Hellenbrand et al., 2008).

Of course, the risk of contracting IMD is much higher inside the household of an infected person. However, secondary cases have become rare due to effective antimicrobial chemoprophylaxis of household members and anyone exposed to an infected patient's oral secretions. The risk of secondary infections "is highest during the first few days" (Rosenstein et al., 2001, p. 1384). Other important risk factors include the active or passive exposure to tobacco smoke and a concurrent viral infection of the upper respiratory tract (Rosenstein et al., 2001, p. 1381). Both of these risk factors damage the mucous membranes and thus facilitate the penetration of the mucosa by virulent meningococci, which then gain access to the bloodstream.

The highly variable bacterium *Neisseria meningitidis* is classified into at least 13 serogroups (antigen groups of bacterial strains), "with serogroups B and C responsible for the majority of cases in Europe and the Americas and serogroups A and C predominating throughout Asia and Africa" (Rosenstein et al., 2001, p. 1379). There is however some fluctuation in the distribution of serogroups, e.g. the serogroups Y and W-135, both commonly associated with pneumonia, accumulated at times in the past decades in the USA. Figure 1.1 shows the locations of all IMD cases registered by the NRZM during 2002–2008 in Germany marked by the corresponding serogroup. Serogroups B and C dominate the map, where the former appears in more obvious and intense clusters in and around agglomerations than the latter. Note that regarding the *incidence* of the disease, i.e. the number of cases in relation to the population size, northern and eastern (new) federal states have higher rates than southern and western (old) federal states (Hellenbrand et al., 2008, p. 269).

As an option of disease prevention, vaccination is possible against serogroups A, C, Y, and W-135. Since 2006, vaccination of all children over 12 months using a conjugate vaccine against serogroup C is recommended by the German Standing Vaccination Committee (STIKO) at the RKI (Hellenbrand et al., 2008, p. 266). Unfortunately, there is no licensed vaccine against serogroup B, which is why "most cases of meningococcal disease in Germany are not preventable by vaccination" (Elias et al., 2006, p. 1693). In the USA, a quadrivalent polysaccharide-protein conjugate vaccine against serogroups A, C, Y, and W-135 is used, but it is relatively ineffective in young children (Rosenstein et al., 2001, p. 1385). Furthermore, "in the abscence of an effective vaccine against serogroup B, a more widespread outbreak would result in substantial morbidity and mortality" (Rosenstein et al., 2001, p. 1379). This is why a correct and fast diagnosis of the meningococcal disease is of great importance. Thompson et al. (2006) discuss the difficulty of this task keeping in mind that IMD shares many symptoms

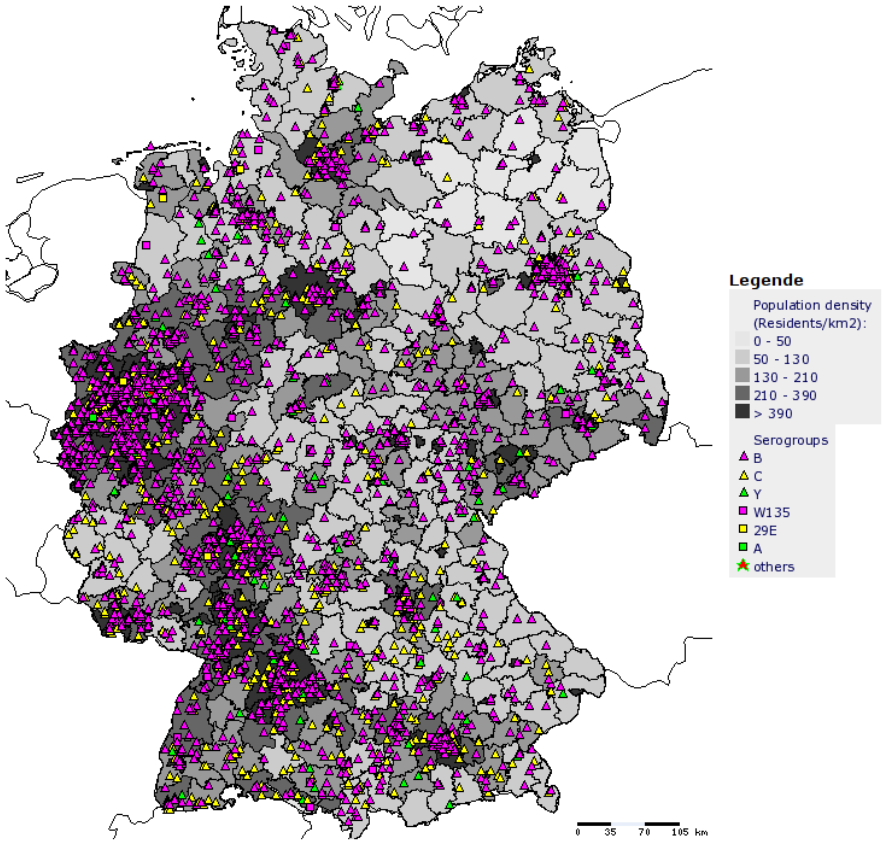


Figure 1.1: Locations of all IMD cases in Germany during the years 2002–2008 registered by the German National Reference Centre for Meningococci. The figure has been generated using EpiScanGIS (Meningococcal Disease Surveillance in Germany, `http://www.episcangis.org`, as of 14 August 2009).

with less serious illnesses. Meningococcal meningitis similar to other forms of meningitis is accompanied by headache, fever, and sometimes stiffness of the neck, nausea, vomiting, photophobia, and an altered mental status. Meningococcal sepsis (Meningococcemia) is characterized by fever and petechial or purpuric rash.

When analysing the pathogenic agents of the cases of IMD, one is concerned with the high diversity of meningococci. Beneath their classification into serogroups, there are actually hundreds of different *finetypes* of Neisseria meningitidis, i.e. different antigenic profiles of meningococcal strains. In Germany the two most common finetypes are *B:P1.7-2,4:F1-5* and *C:P1.5,2:F3-3*, which belong to serogroups B and C, respectively, and are denoted by the antigen sequence types of the outer membrane proteins PorA and FetA. For a detailed description see Elias et al. (2006), who carried out a spatio-temporal analysis of German IMD cases registered and finetyped by the NRZM during the period 1 December 2001 and 1 June 2005 with respect to evolving spatio-temporal disease clusters. Here, "a meningococcal disease cluster is regarded as an aggregation of cases caused by the same bacterial strain

closely grouped in space and time" (Elias et al., 2006, p. 1689). Applying a retrospective spatio-temporal scan statistic using the SaTScan software (Kulldorff, 1997), they detected 26 small and short clusters that included a total of 68 cases corresponding to 4.2% of all cases. The clusters covered an average of 2.6 patients (min: 2, max: 10) and the median cluster duration was four days (min: < 1, max: 24). The median interval between the first and the second case was four days ranging from < 1 to 23 days. The authors mention several other analyses concerning this interval: In an American analysis 73% of secondary cases appeared in the first two weeks after the index case, in France 72% of secondary cases occured in the first week after the index case. In a British survey, the median interval was 1.5, 5 or 23 days in households, schools and universities, respectively.

In the work of Elias et al. (2006), different finetypes are basically assumed as spreading independently of each other in the sense that the finetype of the meningococcal strain does not change during transmission. If there are two cases of different finetypes in temporal and spatial proximity, a transmission between the two cases is out of question. Although it must be noted that potentially "meningococci also have the capacity to [...] switch from serogroup B to C or vice versa" (Rosenstein et al., 2001, p. 1379), this simplifying approach of finetype-specific transmission will also be followed in this thesis.

1.2 Available Data

The following subsections describe the available data and their preparation for analysis. The data retrieved from various sources include geographic information on Germany, registered cases of IMD, weekly numbers of influenza cases, and official population statistics.

1.2.1 Shapefiles

An essential component in spatial statistics is the region of observation, which for the IMD data is entire Germany. A computational representation of its shape enables mapping of the spatial distribution of the events (cases of IMD) and the estimation of spatial point process models. A popular electronic format for geographic data are ESRI shapefiles (Environmental Systems Research Institute, 1998). These files describe so-called vector data, i.e. geometrical objects to which additional information may be attributed. The geometrical objects can be either points, lines or polygons. For instance, the boundary of Germany can be represented as a set of 31 polygons of which the biggest one corresponds to the main region of Germany and the others to small islands in the north. Similarly, a more detailed polygonal representation is obtained by the boundaries of the 413 German administrative districts also called counties

(301 “Landkreise” and 112 “Stadtkreise”, NUTS-3 level, cf. Wikipedia (2009)). These shapefiles are provided by the GeoDataCenter of the German Federal Agency for Cartography and Geodesy and are available online from `http://www.geodatenzentrum.de`.

In `R`, shapefiles can be imported using the function `readOGR` from the package `rgdal` (Keitt, Bivand, Pebesma & Rowlingson, 2009) resulting in an object of the class `SpatialPolygons` (or `SpatialPolygonsDataFrame`) from the package `sp` (Pebesma & Bivand, 2005). The borderlines of the counties from one of the shapefiles can be seen in Figure 1.3. Another class representing sets of polygons is `gpc.poly` from the package `gpclib` (Peng, 2009). I will especially make use of this latter class for the estimation of the spatio-temporal models of Chapter 4 as it provides necessary methods for intersecting polygons.

1.2.2 Cases of IMD

The IMD data I have received from the NRZM consist of 636 cases of the two most common meningococcal finetypes in Germany, B:P1.7-2,4:F1-5 and C:P1.5,2:F3-3, registered during the years 2002–2008. Among the IMD cases finetyped at NRZM in 2008, 13.6% (B:P1.7-2,4:F1-5) and 10.1% (C:P1.5,2:F3-3) corresponded to these finetypes, respectively (cf. German National Reference Centre for Meningococci, 2009). Unfortunately, the supplied data are not complete because reporting of IMD cases to the NRZM is voluntary (it is only obligatory to the RKI). In 2003 the NRZM conducted the finetyping for approximately 65% of all IMD cases in Germany (Schrauder et al., 2007), which could lead to an information bias when some counties consistently report less cases than others. In the analysis of Elias et al. (2006) the underreporting to NRZM by some counties did not substantially affect the results. The NRZM also assumes that the coverage has increased in recent years with an approximate value of 87% in 2008. Table 1.1 lists the available information in the NRZM dataset.

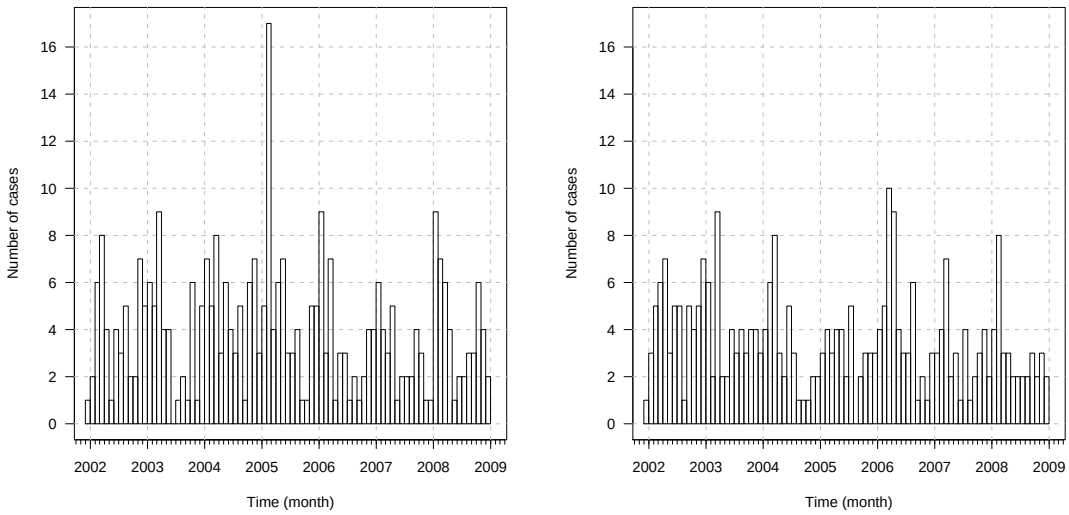
Finetype	Meningococcal antigen sequence type (B:P1.7-2,4:F1-5 or C:P1.5,2:F3-3)
Temporal reference	Date of sampling (e.g. blood culture or central spinal fluid), missing for 14 cases
	Date of receipt of the meningococcal sample for finetyping at NRZM
Spatial reference	Postcode of the patient’s residence
	Geographical coordinates (longitude, latitude) of the centroid of the postcode’s region
	Official key of the county of the patient’s residence
Patient data	Age of the patient at the time of sampling, missing for 1 case
	Gender of the patient, missing for 5 cases

Table 1.1: Information contained in the dataset of the 636 cases of IMD.

Of the total 636 cases, 336 belong to the most common finetype of serogroup B and 300 to the most common finetype of serogroup C. For comparison, I queried the RKI online database

SurvStat@RKI (Robert Koch-Institut, 2009) for registered cases of meningococci and acquired cross tabulars by county and week for the years 2001–2009 and for the serogroups B and C separately. In 2002–2008, the RKI registered 2 401 cases belonging to serogroup B, 900 cases belonging to serogroup C, and a total of 4 186 IMD cases (including many cases with unknown serogroup). Thus, the NRZM data of the specific finetype B:P1.7-2,4:F1-5 make up only 14.0% of all known cases of serogroup B, which also shows the high diversity of the bacterial pathogen. For the finetype C:P1.5,2:F3-3 the corresponding proportion is 33.3%.

As usual with infectious diseases, the actual time point of infection is unknown for the IMD cases. One only knows about the date of specimen sampling, when the infection has already become apparent. As a rough estimate, I define the time of illness and infectivity as being four days before the sampling date. These four days are assumed to be the median incubation period, which actually exhibits large deviations between different cases and may also take more than two weeks (J. Elias, personal communication). Note that this choice of time shift does not affect the sequence of IMD cases, but only its temporal link to the cases of influenza presented in the next subsection. To the 14 missing sampling dates the corresponding reception date minus 5 days is imputed, which is the median lag between sampling and reception.

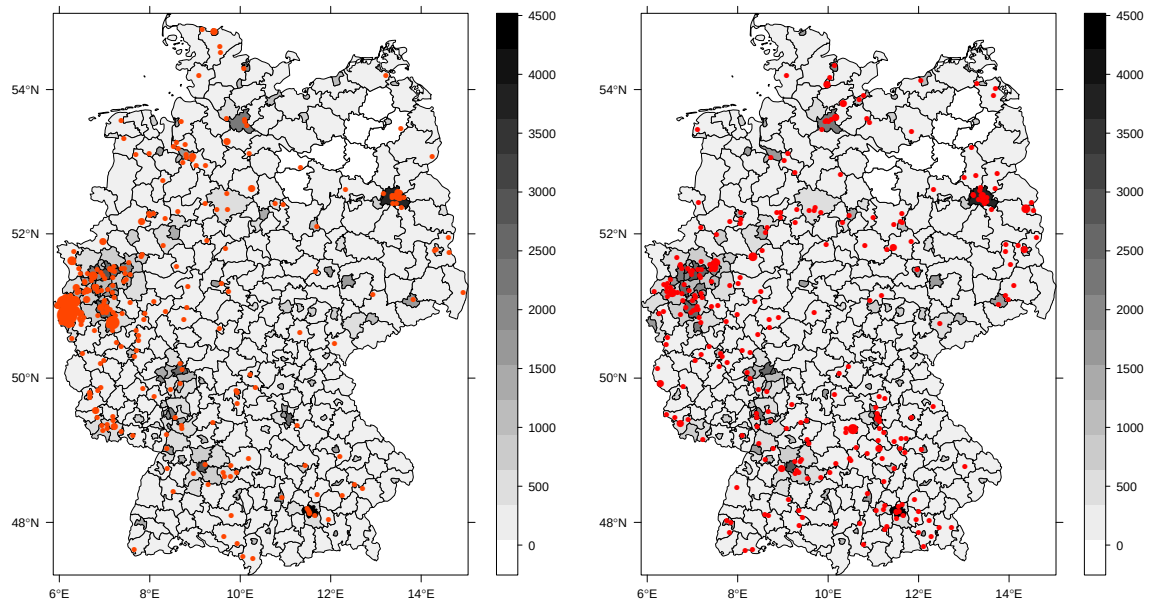


(a) Finetype B:P1.7-2,4:F1-5.

(b) Finetype C:P1.5,2:F3-3.

Figure 1.2: Monthly numbers of IMD cases for both finetypes separately.

Figure 1.2 shows the monthly numbers of IMD cases for each finetype. Cases of IMD predominantly occur during winter and early spring, which can be seen from more or less pronounced peaks in the figure. Except from a large outbreak caused by the serogroup B finetype around February 2005, both finetypes exhibit a comparable amount of cases per month. Observe that for these NRZM data there is no apparent decrease in the number of cases over the years as is the case for the complete meningococci records from the RKI (see Section 1.1).



(a) Finetype B:P1.7-2,4:F1-5.

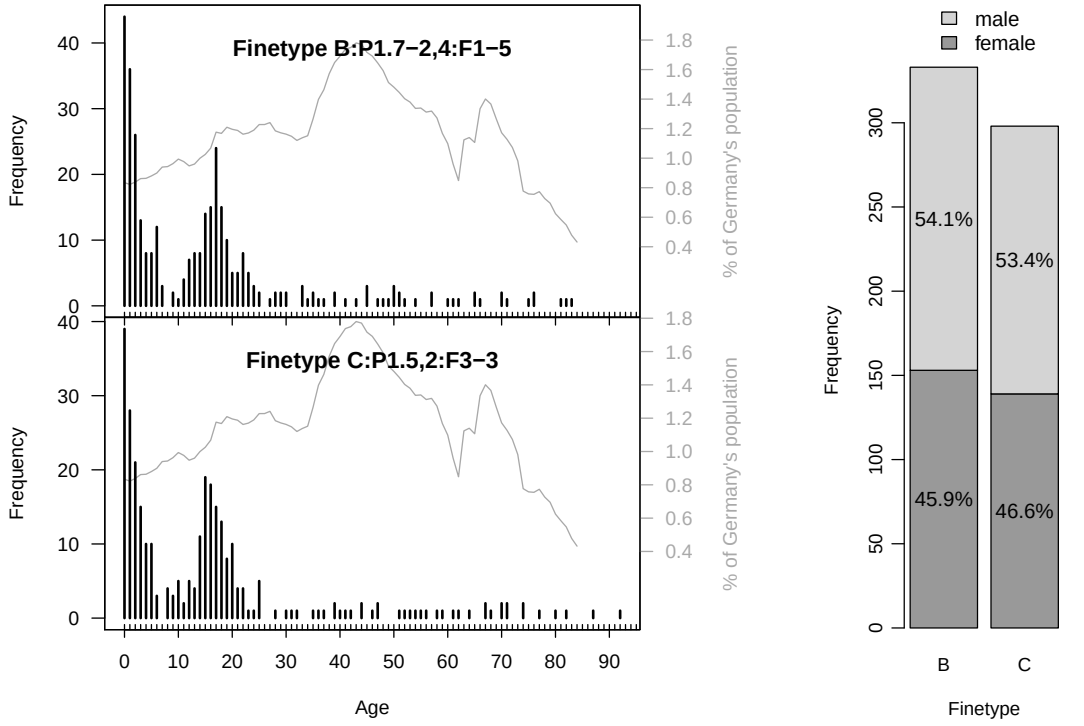
(b) Finetype C:P1.5,2:F3-3.

Figure 1.3: Spatial point patterns of the cases of meningococci by finetype during the years 2002–2008. The area of each dot is proportional to the number of cases at its location. Also shown are the population densities (inhabitants per km^2) of Germany's counties.

Besides being the most common finetypes, a particular interest of the NRZM in these two specific finetypes arises because they seem to have different spatio-temporal distributions. Figure 1.3 presents the spatial distributions of the two finetypes based on the postcodes of the patients' residences. Over the 7-year period some cases shared the same location; therefore, the area of each point in the figure is drawn proportional to the number of cases at its location. For the serogroup B finetype in (a) the highest point multiplicity is 16, whereas for the serogroup C finetype in (b) this number is 4. In connection with the temporal occurence of the events shown in Figure 1.2, the spatial distribution suggests that IMD is an endemic disease, i.e. cases can occur at any time and at any location. The maps also show the population densities (inhabitants per km^2) of the counties, which can be assumed to be roughly proportional to the population at risk of infection. Certainly, spatial heterogeneity of the observed point patterns partially arises from spatial variation in the density of the population at risk. Not surprisingly, the intensity of points in metropolitan areas like Berlin, Munich or the Ruhr area is thus higher. The population data was extracted from the official municipality statistics (Statistisches Bundesamt (DESTATIS), 2009), which also contain the municipal area in km^2 required later.

Overlaying this background intensity, the NRZM observes different transmission dynamics in the hundreds of existing finetypes. The most common finetype B:P1.7-2,4:F1-5 features

a more "stationary" behaviour in the sense that infections cluster more in space and time (especially in western North Rhine-Westphalia), whereas the serogroup C finetype appears more diffuse and does not stay in a region for long. It is supposed that this phenomenon is due to differences in the efficiency of the immune reactions elicited by the two finetypes. For instance, the immunity of the nasopharyngeal mucous membrane elicited by B:P1.7-2,4:F1-5 might not be as efficient as it is against C:P1.5,2:F3-3. Animated graphics of the space-time locations of infections would give more insight into the infectiousness of the finetypes, i.e. of their epidemic character.



(a) Age distributions.

(b) Gender distributions.

Figure 1.4: Distributions of (a) age and (b) sex in the NRZM data for both finetypes separately. Also shown in (a) is the age distribution in Germany's population from 0 to 84 years quantified by the axis on the right-hand side of the plot. For comparison in (b), the overall female proportion in Germany is 51.0%. These population data are as of 31 December 2007.

Concerning the supplied patient data there is not much difference between the finetypes (see Figure 1.4). The graphic (b) shows that there are slightly more male than female patients. The corresponding proportions are similar for both finetypes, but they do not reflect the gender distribution in Germany, which is 49.0% male and 51.0% female (as of 31 December 2007, Statistisches Bundesamt (DESTATIS) (2009)). Graphic (a) highlights for both finetypes a modal age of < 1 year (13.1% of all cases) followed by children of one (10.1%) and two (7.39%) years. A reason for the high impact among infants is that they have not yet developed

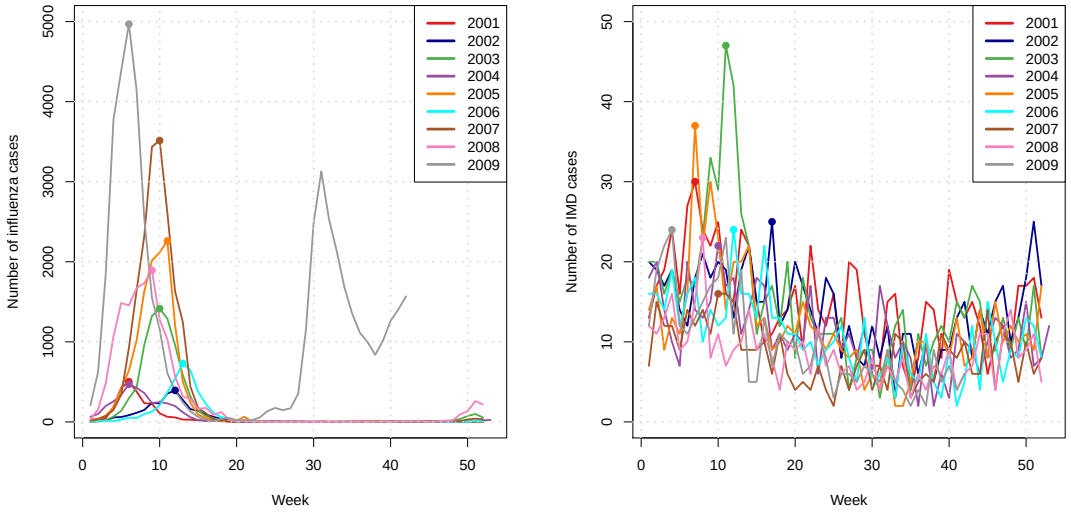
protective antibodies (Rosenstein et al., 2001, p. 1378). The frequency of the disease declines until the age of seven years, where no single infection has been observed by NRZM. The second large group of mainly affected persons is from about 12 to about 22 years of age with a peak at its centre (17 years) for the serogroup B finetype and at 15 years for the serogroup C finetype. As much as about 85.5% of affected persons are younger than 26 years. Other cases of IMD occur occassionally at higher ages where they are distributed rather uniformly. For comparison, the graphic also shows the age distribution of Germany's population as of 31 December 2007 (GENESIS-Online, 2009). The observed age distribution in the IMD data clearly does not represent the distribution in the population.

1.2.3 Cases of Influenza

As mentioned in Subsection 1.1, a concurrent viral infection of the upper respiratory tract is an important risk factor for IMD. Furthermore, as was seen in Figure 1.2, infections with virulent meningococci likely occur during the winter months. This suggests the assumption that outbreaks of influenza-like diseases predispose to the spread of invasive meningococci. Note that there are hundreds of agents causing respiratory infections and only about 10% are actually caused by influenza viruses. The predominant pathogens of influenza-like diseases are rhinoviruses and coronaviruses to which should be drawn more attention than they receive at the moment (Jefferson, 2004). However, because influenza is a notifiable pathogen and the others are not, analysis is restricted to the case numbers of influenza-like diseases due to infections with influenza viruses only. I again queried the RKI online database *SurvStat@RKI* (Robert Koch-Institut, 2009) to obtain cross tabulars for the registered number of cases of influenza by county and week for the years 2001-2009 separately.

Figure 1.5(a) illustrates in the same plot the weekly numbers of cases for each year in Germany. The case numbers of the current year 2009 are of course not complete and one recognises the impact of the recent "swine flu". This tremendous peak relative to the other years' case numbers in this period of the year is most likely to be partially induced by the large public concern addressed to the "swine flu". In "normal" years however, cases of influenza are largely underreported, which is probably also due to ambiguous diagnoses of influenza-like diseases (Jefferson, 2004). This also affects the stochastic modelling of the spread of influenza which is e.g. carried out in Dargatz, Georgescu & Held (2005). Currently, the NRZM monitors if the typical seasonal increase of the IMD cases around late February will occur earlier in 2010 as a consequence of the recent "swine flu" epidemic.

Note in the Figure 1.5(a) that the increase in the final weeks of every year already marks the beginning of the next year's influenza wave reaching its peak around the 10^{th} week (median). In 2006, the increase was very slow and had a late peak in week number 13. Years 2001 (2 489



(a) Weekly numbers of influenza cases.

(b) Weekly numbers of IMD cases (any serogroup).

Figure 1.5: Weekly numbers of influenza (any type) and meningococci (any serogroup) cases for the years 2001–2009. As the data are queried on 4 November 2009, the numbers in late 2009 are not complete. The dot on each curve designates its peak.

cases), 2002 (2 575) and 2004 (3 494) have also been less severe, whereas the other years had tremendous winter outbreaks: 2003 (8 488), 2005 (12 737), 2008 (14 852), 2007 (18 900), 2009 (50 537), where the 2009 number due to the "swine flu" is much higher than the case numbers of previous years.

Part (b) of the Figure 1.5 contains the same type of graphic as in (a), but shows the cases of IMD of any serogroup (this data was also obtained from *SurvStat@RKI*). Although the seasonality is less pronounced and exhibits much smaller peaks than in (a), one can guess a yearly sinusoidal course with a maximum amplitude around week number 10 (the median peak time), which coincides with the median peak of the influenza waves.

It is however difficult to recognise direct connections between outbreaks of both diseases from the above time series. Figure 1.6 rearranges the curves and scales the number of influenza cases to have the same maximum value as the number of IMD cases in the respective year. In the years 2001 and 2003, the influenza wave precedes the IMD peak by about two weeks. Their mutual sequence is inversed in 2005, and it is ambigous in the remaining years.

The connection between outbreaks of meningococci and influenza has already been analysed, e.g. in a project work at the LMU Department of Statistics (cf. Breitenacher, Obermeier & Schomaker, 2004), and by Jensen, Lundbye-Christensen, Samuelsson, Sørensen & Schønheyder (2004). The latter analysis is based on observations in North Jutland County, Denmark, during 1980–1999, and the findings support "the theory that the influenza detection rate is associated with the number of [I]MD cases in the population during the same week". Lags of

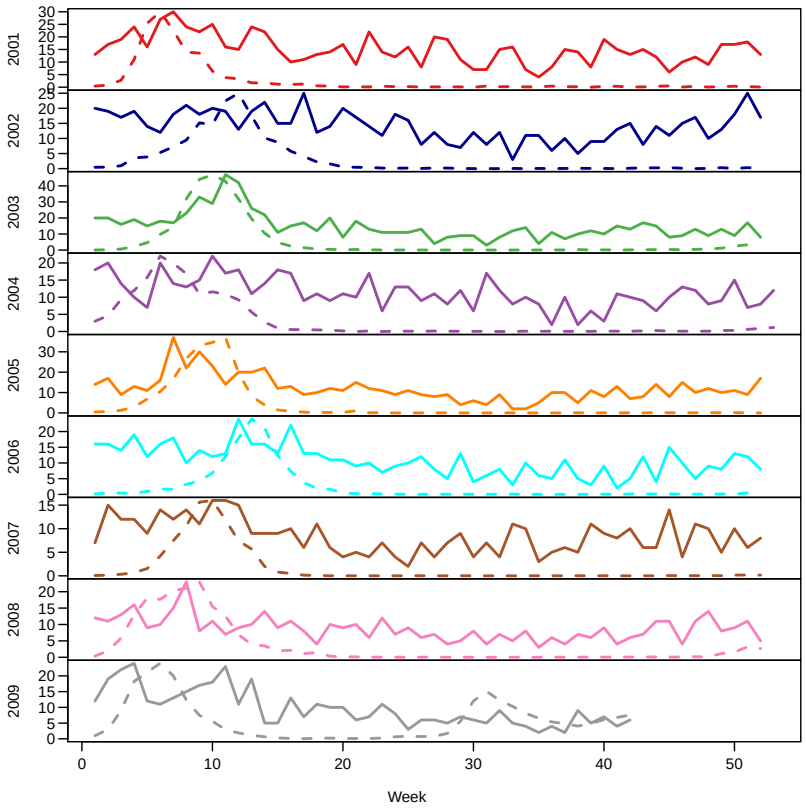


Figure 1.6: Weekly numbers of IMD cases for each year (solid lines) together with the scaled number of influenza cases (dashed). The annotation of the vertical axis corresponds to the former.

one and two weeks of the influenza case numbers were also tested, but they did not show up as being significant. In contrast, Breitenacher et al. (2004) conclude – based on the numbers from RKI in the period January 2001 to August 2004 – that a delay of two weeks would be the best choice. Note however that the reporting delay (infection to notification) is probably different for influenza and meningococci.

As mentioned in Subsection 1.2.2, I use the sampling date (or its imputed value if it was missing) minus four days as temporal reference for the IMD cases of NRZM. In contrast, the available case numbers of influenza are given by the week of notification at RKI. Thus there is already an artificial translation of about one week between both diseases in the data analysed in the following chapters. When using the case numbers of influenza as an explanatory variable for cases of IMD, different lags will be considered during model selection.

1.2.4 Data Preparation

In addition to importing and arranging the above described data from its various sources into `R`, several data preparations are necessary for the point process based analyses.

Projection of geographical into Cartesian coordinates

Doing spatial statistics one is usually concerned with geographical coordinates in longitude and latitude, i.e. coordinates on the surface of an ellipsoid approximating the earth. The most common coordinate frame and reference ellipsoid is called *World Geodetic System* (WGS) with its latest revision being WGS 84, which is also used by the GPS satellite navigation system and for NATO military geodetic surveying (see `http://spatialreference.org/ref/epsg/4326/`). It emerged from and is only a minor refinement of the original reference ellipsoid *Geodetic Reference System 1980* (GRS 80).

The direct use of geographical event coordinates in statistics is hindered by the fact that special distance formulae are then needed to calculate the distance between events, e.g. the great-circle distance, and that the data are actually not 2D which makes the drawing of graphics on the flat surface of a sheet of paper or a screen more complicated. The best way to prevent such issues is to initially project the geographical coordinates onto a plane in a Cartesian system (or rather into a planar coordinate reference system (CRS)) and continue the analysis using only the well known euclidean geometry, which makes a statistician's life much easier. One such pan-European CRS is the *European Terrestrial Reference System 89*, briefly ETRS89, which is recommended by the European Environment Agency for statistical analysis and display, especially where true area representation is required (see `http://spatialreference.org/ref/epsg/3035/`).

For practical cartographic reprojections the library `PROJ.4` can be used and accessed from `R` through the package `rgdal`. Its function `spTransform` reprojects objects of the `sp` classes. For instance, if we want to transform the WGS 84 longitude/latitude representation of Germany's counties from Figure 1.3 – say the corresponding `SpatialPolygonsDataFrame` is called `counties_longlat` – to a Cartesian coordinate system in kilometres, the command is

```
R> spTransform(counties_longlat, CRS("+init=epsg:3035 +units=km"))
```

where the second argument defines the desired CRS through an identifying code, which is 3035 for ETRS89. An application of the spatial capabilities of `R` on spatio-temporal ecological data can be found in Hengl, Loon, Sierdsema & Bouten (2008), who also provide a short overview of `R` packages for spatial data.

All methods that I have implemented to handle the spatial data are available from the attached CD (Appendix A).

Mapping of grid-based covariate information to event locations

Besides handling geographical coordinates, another aspect of spatial data is to attribute covariate information, which is available on some spatial grid, to the event locations in order

to analyse their effects on the spread of the disease. Although this sounds less complicated, the task can hold many difficulties in the details.

For a municipality-based analysis using the population data from DESTATIS (Statistisches Bundesamt (DESTATIS), 2009), it is necessary to attribute to each case location (here: a postcode) its municipality. This is the highest available spatial resolution for the statistical analyses in this thesis. The mapping requires a database (or shapefile) for linking postcodes (or geographic coordinates of postcode centroids) to official municipality keys. Such a database is freely available through the internet open community project *OpenGeoDB* (see `http://opengeodb.giswiki.org/`). However, the mapping of the case's postcode to a specific municipality is in fact not always possible, because postcodes can be part of multiple municipalities (and even counties). In the IMD data, 74 of 509 postcodes do not match a unique municipality. As there are no further hints, one could e.g. choose the municipality representing the mean of the candidates with respect to the covariate information or choose a municipality at random. Other difficulties arose from the fact that there are 12 300 municipalities in Germany and 367 weeks in the observation period for which reason the spatio-temporal analyses applied in this thesis would involve a `data.frame` of 4 514 100 rows or even 20 159 700 rows, depending on the implementation. Because the resulting computational requirements turned out to be too heavy, the maximal spatial resolution of the covariate information will be the county level.

For a county-based analysis, the prerequisites are better because the provided IMD data already contain the official key of the county of the patient's residence. The only corrections concerned out-dated county keys of the federal states Saxony and Saxony-Anhalt due to local government reforms in 2008 and 2007, respectively. This permitted the mapping of population numbers from the DESTATIS data to the case data.

More laborious was the mapping of the influenza case numbers from *SurvStat@RKI* (Robert Koch-Institut, 2009), which are on a county×week grid, to official DESTATIS counties. This is because the 425 "counties" from the RKI online database are only referenced by name, which is not always the current official one used by DESTATIS. This drawback, as well as the subdivision of the official county "Berlin, Stadt" into 13 parts considered by RKI and the above mentioned reforms of Saxony and Saxony-Anhalt made a manual mapping necessary. The temporal reference of the numbers of influenza cases must also be treated with caution when searching for a specific date. The calendar week numbers in *SurvStat@RKI* are provided in the ISO 8601 format (cf. ISO, 2004), which has the advantage that all calendar weeks are complete 7-day-weeks. On the other hand, the number of calendar weeks of a year then either is 52 or 53, and the last calendar week of a year can reach into the next year. Fortunately, this format can be matched in `R` by the function `strftime` using `format="%V"`, and `format="%G"` corresponds to the week-based year.

Breaking up tied event times

As will be explained in the next section, orderly point processes do not allow for tied event times. Nevertheless, real data are likely to contain tied event times due to rounding and temporal resolution. The preferred method to deal with such duplicates is to introduce some random jitter and to perform a sensitivity analysis on whether the random variation is crucial for the results. However, because of the complexity of the analyses in this thesis I content myself with a single re-ordering of the tied events. For each of the finetypes, non-unique event times are identified and the time points of the tied group are adjusted by substracting 0.01 [days] for each of the duplicated event times. For instance, if the event times are $(1, 1, 2, 2, 2, 3)$ the procedure returns $(0.99, 1, 1.98, 1.99, 2, 3)$. This onesided adjustment is reasonable because events on the same day should be attributed with the same endemic covariates, which could change their values at days 1, 2 or 3, respectively. Other approaches for handling ties are known from survival analysis, but these are not applicable in the current spatio-temporal setting, where explicit event times are needed for evaluation (cf. Diggle, Kaimi & Abellana, 2009, Section 2.3).

For the serogroup B finetype in the current data, there are 26 dates with two infections at once, and for the serogroup C finetype this number is 16. For the latter, there is also one day with three cases at once.

2 Point Process Theory and a Preliminary Model

Moving towards modelling of spatio-temporal point processes this chapter provides in Section 2.1 an introduction to the theory of point processes, which is essential for the understanding of the models treated in this thesis. Section 2.2 then sketches a first spatio-temporal point process model corresponding to a finite partition of the observation region.

2.1 Basics of the Theory of Point Processes

This section shall prepare the notation and point process methodology used in the remainder of this thesis. At first informally, a point process is a stochastic model underlying the occurence of *events* (points) in time and/or space. Actually, most point processes in the literature are purely temporal (e.g. renewal/failure time or survival data) or purely spatial (e.g. locations of lightning strikes, tree population, or other events aggregated over time). A comprehensive book for statistical analysis and modelling of the latter was written by Illian, Penttinen, Stoyan & Stoyan (2008), whereas Møller & Waagepetersen (2004) provide a more profound description of their mathematical and measure theoretical aspects. The two parts of Daley & Vere-Jones (2003, 2008) provide a deeper insight into general point process theory, also covering the *spatio-temporal* case in Section 15.4. The Master's Thesis by Fuest (2009, Section 3.2) also contains an introduction to point processes with a focus on marked point processes in time. A brief description of spatio-temporal point processes is Schoenberg, Brillinger & Guttorp (2002), and an overview of related methods and applications is given by Diggle (2007) and Vere-Jones (2009). The rare connection of both space and time in classical statistics is probably due to the facts that spatio-temporal point process data has not been routinely available and that related analyses usually demand heavy computations. Nevertheless, there are of course applications concerning dynamic spatio-temporal behaviour, e.g. in epidemiological surveillance, where the events constitute the time points and locations of infections with a disease. However, because the research in this field of statistics is very recent "the methodology of space-time point processes is still under-developed" (Illian et al., 2008, p. 426).

2.1.1 Basic Definitions

Formally, a point process $\boldsymbol{X}$ is a random countable subset of a complete separable metric space $\mathcal{X}$, usually an Euclidean space. It can equally be represented by its corresponding random counting measure $N(B)$ on the state space $\mathcal{X}$ (silently equipped with the corresponding Borel σ-algebra $\mathcal{B}(\mathcal{X})$), which counts the number of points falling in a measurable subset $B \subset \mathcal{X}$. The full distribution of the process is then defined in principle by prescribing the finite-dimensional joint distributions of all possible families $\{N(B_1), N(B_2), \ldots, N(B_k)\}$, $B_1, B_2, \ldots, B_k \subset \mathcal{X}$, $k \in \mathbb{N}$. In this thesis, $\mathcal{X}$ will either be the positive real line $\mathbb{R}_+$ (temporal point processes), the two-dimensional Euclidean space $\mathbb{R}^2$ (spatial), or the Cartesian product of both, $\mathbb{R}_+ \times \mathbb{R}^2$ (time $\times$ space). For instance, $N(t) := N((0,t])$ denotes a purely temporal point (or counting) process.

Occassionally, I will refer to the so-called *ground process*

$$N_g(t) := \int_0^t \int_{\mathbb{R}^2} \mathrm{d}N(u, \boldsymbol{s}) = \int_{(0,t]\times\mathbb{R}^2} N(\mathrm{d}u \times \mathrm{d}\boldsymbol{s}) = N((0,t] \times \mathbb{R}^2)\,,$$

which for a spatio-temporal process counts the overall number of points up to and including time t, irrespective of their locations. Note in this context that first, counting processes have right-continuous trajectories, and second, integrals like the above are to be understood as stochastic Lebesgue-Stieltjes integrals. In particular, if the integral is with respect to a counting process, it actually equals the sum of the integrand values at the points of the process inside the given limits of integration. Hence if $\{t_1, \ldots, t_{N_g(t)}\}$ denotes the random set of time points where events of the process occur, i.e. where the counting process N jumps one step higher, then the above ground process could be written as $N_g(t) = \sum_{i:t_i \leq t} 1$.

Usually and also in this thesis it is assumed that point processes are *locally finite*, i.e. only a finite number of points falls in any bounded subset of $\mathcal{X}$, and they are required to be *regular* in the sense of Daley & Vere-Jones (2003, Definition 7.1.I.), which enables likelihood functions for point processes. Regularity also implies that the point process is *simple* (or *orderly*), meaning that events cannot occur simultaneously. A point process N on $\mathcal{X} = \mathbb{R}^d$ is regular, if for any bounded Borel set $B \subset \mathcal{X}$ the local Janossy measures $J_n(\cdot|B)$, $n \in \mathbb{N}$, exhibit Lebesgue-densities on $\mathcal{X}$. The local Janossy measure $J_n(\cdot|B)$ satisfies, for $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in B$,

$$J_n(d\boldsymbol{x}_1 \times \cdots \times d\boldsymbol{x}_n|B) = \mathbb{P}(\text{exactly } n \text{ points in } B \text{ at locations } d\boldsymbol{x}_1, \ldots, d\boldsymbol{x}_n)\,,$$

where $d\boldsymbol{x}_i$ here denotes the infinitesimal sphere around $\boldsymbol{x}_i$. Then, the likelihood of a *realisation* $\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ of a regular point process on a bounded Borel set $B \subset \mathbb{R}^d$ is exactly the Lebesgue-density $j_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n|B)$ of the local Janossy measure $J_n(d\boldsymbol{x}_1 \times \cdots \times d\boldsymbol{x}_n|B)$. For the details on point process likelihoods, Janossy densities and Janossy measures, I refer to

Daley & Vere-Jones (2003, Sections 5.3, 5.4, and 7.1).

Although the above formulation of a point process is general, it is often necessary to separate spatial and temporal domains in a space-time setting in order to account for the evolutionary character of time. Thereby, the observed realisation $\boldsymbol{x}$ of $\boldsymbol{X}$ is called *spatio-temporal point pattern*, and the underlying stochastic model *spatio-temporal point process*. Note that although the term "spatio-temporal" is commonly used, I always consider the temporal dimension as the first coordinate, as it implies a natural ordering of the events. Thus, the basic format of the data is

$$\boldsymbol{x} = \{(t_i, \boldsymbol{s}_i) : i = 1, \dots, N\},$$

where t_i is the time of occurence, and $\boldsymbol{s}_i$ is the location of the ith event. These data are assumed to form a complete record of all events occuring in a prespecified observation period $[0, T]$, $T > 0$, and observation region $W \subset \mathbb{R}^2$ (higher dimensions are equally possible, but not of interest in this thesis). Note that I use the uppercase symbol N instead of n for the observed number of events $N_g(T)$ to indicate that this number is actually random in point processes. Although the symbol N is now overloaded by the final number of events and the related counting process $N(\cdot)$, there should be no ambiguity of the intended meaning in the remainder of this thesis.

The reference point process is the *Poisson process*, which can be characterised by the following two properties:

Poisson distribution of counts: the number of events in any bounded subregion $B \subset \mathcal{X}$ follows a Poisson distribution such that

$$N(B) \sim \mathrm{Po}(\Lambda(B)) \qquad \text{where} \qquad \Lambda(B) := \iint_B \lambda(t, \boldsymbol{s}) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s} \, .$$

Here, $\lambda(t, \boldsymbol{s})$ is a non-negative function called the *intensity function*, and Λ is the *intensity measure*.

Independent scattering: the numbers of events $N(B_1), N(B_2), \dots, N(B_k)$ in disjoint subregions $B_1, B_2, \dots, B_k \subset \mathcal{X}$, $k \in \mathbb{N}$, are mutually independent.

The special cases, where $\lambda(t, \boldsymbol{s}) \equiv \lambda(t)$ and $\lambda(t, \boldsymbol{s}) \equiv \lambda(\boldsymbol{s})$ are called spatially *homogeneous* and time-*stationary*, respectively. Although both terms indicate that the statistical properties of the process do not change under translation, the term "homogeneous" is preferred in the spatial setting, whereas "stationary" is predominantly used for temporal point processes. For instance, a temporal point process is stationary, when the number of points occuring in an interval does not depend on the interval's location on the time axis, but only on the length of the interval. In the spatial dimension, besides the similar property of homogeneity, there

is also the property of invariance of the point process distribution under rotations about the origin $\mathbf{0}$, which is called isotropy. A homogeneous Poisson process is isotropic.

For a general spatio-temporal point process on the Euclidean space $\mathcal{X} = \mathbb{R}_+ \times \mathbb{R}^2$, the intensity function is defined as the Lebesgue density of the intensity measure $\Lambda(B) = \mathbb{E}(N(B))$ on the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. If the density exists, which is commonly the case and assumed here, it holds

$$\Lambda(B) = \int_B \lambda(\boldsymbol{u}) \, \mathrm{d}\boldsymbol{u} \qquad (B \in \mathcal{B}(\mathcal{X})) \,,$$

Similarly, the intensity function can be defined as the limiting expectation or probability

$$\lambda(\boldsymbol{u}) = \lim_{|\mathrm{d}\boldsymbol{u}| \to 0} \frac{\mathbb{E}(N(\mathrm{d}\boldsymbol{u}))}{|\mathrm{d}\boldsymbol{u}|} = \lim_{|\mathrm{d}\boldsymbol{u}| \to 0} \frac{\mathbb{P}(N(\mathrm{d}\boldsymbol{u}) = 1)}{|\mathrm{d}\boldsymbol{u}|} \,, \tag{2.1}$$

where $\mathrm{d}\boldsymbol{u}$ denotes an infinitesimal sphere around the point $\boldsymbol{u} = (t, \boldsymbol{s})$ with Lebesgue measure (volume) $|\mathrm{d}\boldsymbol{u}|$. The probability expression above holds because the process is assumed to be simple, such that almost surely $\mathbb{E}(N(\mathrm{d}\boldsymbol{u})) \in \{0, 1\}$.

The interpretation of the intensity function is easiest if the point process is homogeneous and stationary, in which case $\lambda(t, \boldsymbol{s}) = \lambda$ is a constant and the expected number of points in B, $\mathbb{E}(N(B)) = \lambda \, |B|$, is proportional to the space-time volume of B. Thus, λ is the mean number of events per unit area and time interval. As a consequence, conditionally on the number of events in B, the events in B are uniformly distributed. In the more general case, the interpretation of the intensity is restricted to an infinitesimal view derived from (2.1):

$$\mathbb{E}(N(\mathrm{d}t \times \mathrm{d}\boldsymbol{s})) \approx \lambda(t, \boldsymbol{s}) \, |\mathrm{d}t||\mathrm{d}\boldsymbol{s}| \tag{2.2}$$

is the expected number of points in an infinitesimal sphere with center $(t, \boldsymbol{s})$ and volume $|\mathrm{d}t||\mathrm{d}\boldsymbol{s}|$. Because we assume that the point process is simple, this is equivalent to speaking of the probability that a point falls in this infinitesimal ball. Note that in the context of infectious disease epidemiology, point processes of cases can hardly be considered as stationary or homogeneous due to casual outbreaks and the spatial inhomogeneity of the population at risk.

2.1.2 Conditional Intensity Function and Likelihoods

As soon as time appears as a variable of the point process, the process has an "evolutionary" character. This enables its unique representation through its *conditional intensity function* (CIF), also known as *conditional rate process* or *intensity process*. It is the key to likelihood analysis, simulation and prediction of point process models with a temporal dimension. The

following two subsections describe the conditional intensity function and the likelihood first for unmarked then for marked spatio-temporal point processes.

Unmarked Spatio-Temporal Point Process

A spatio-temporal point process is uniquely characterised by its so-called *conditional intensity function* (CIF) $\lambda^*(t, \boldsymbol{s})$. It represents the instantaneous rate or hazard for events at time t and location $\boldsymbol{s}$ given all the observations up to time t. One may think of an instantaneous value of the Poisson rate parameter varying as time proceeds and which always depends on what has happened before. For a rigorous mathematical definition of conditional intensities based on the Doob-Meyer decomposition I refer to Daley & Vere-Jones (2008, Chapter 14).

The CIF is often written as $\lambda(t, \boldsymbol{s}|\mathcal{H}_t)$ with an explicit conditioning on $\mathcal{H}_t$, the σ-algebra generated by the past of the process which represents the information available at time t. I would like to point out that in the literature there is some inconsistency in the notation of $\mathcal{H}_t$. Some authors denote by $\mathcal{H}_t$ the σ-algebra of events up to but *excluding* time t, and others *include* this time point. There are also papers which simply write "up to time t" (as I did above), where the correct understanding is left to the reader. Looking closely at the introduction of the conditional intensity function in Daley & Vere-Jones (2003, Section 7.2), the situation becomes clearer. Denote by $\{t_1, \ldots, t_{N_g(T)}\}$ the ordered set of random time points occuring in the fixed interval $(0, T]$. The conditioning in $\lambda^*(t, \boldsymbol{s})$ is based on the σ-algebra generated by the observed past of the process: $\mathcal{H}_t := \sigma(\{t_1, \ldots, t_{N_g(t)}\})$, which includes a possible event at time point t (this inherits from the more general theory of stochastic processes sketched in Daley & Vere-Jones (2003, Appendix A3.3)). The asterisk notation shall remind that the CIF is actually itself a stochastic process dependent on the random past *history* of the point process: $\lambda^*(t, \boldsymbol{s}) \equiv \lambda^*(t, \boldsymbol{s}, \omega)$ depends on ω through the realisation $\{t_1(\omega), \ldots, t_{N_g(t)}(\omega)\}$ of the history up to time t. The actual ambiguity in the notation is due to a lack of uniqueness in the definition of the conditional intensity function just as the density function of a probability distribution is only determined up to a set of Lebesgue measure zero. Uniqueness may be ensured by taking the *left-continuous* modification of $\lambda^*(t, \boldsymbol{s})$, which guarantees its predictability: "if the conditional intensity has a discontinuity at a point of the process, then its value at that point should be defined by the history before that point, not by what happens at the point itself" (Daley & Vere-Jones, 2003, p. 232). Hence, $\lambda^*(t, \boldsymbol{s}) = \lambda(t, \boldsymbol{s}|\mathcal{H}_{t-})$ is measurable with respect to $\mathcal{H}_{t-}$ and the paths hold $\lambda^*(t, \boldsymbol{s}) = \lambda^*(t-, \boldsymbol{s})$, $\forall t > 0$. Thus, the conditional intensities usually appearing in point process papers are left-continuous versions and can be interpreted as the conditional risk of occurrence of an event at time point t (and location $\boldsymbol{s}$), given the realisation of the process over the interval $[0, t)$. Thus, similar to the unconditional intensity in equation (2.1), the conditional intensity can be defined as the

limit

$$\lambda^*(t, \boldsymbol{s}) = \lim_{\Delta t \to 0,\, |\mathrm{d}s| \to 0} \frac{\mathbb{P}(N([t, t+\Delta t) \times \mathrm{d}\boldsymbol{s}) = 1 \,|\, \mathcal{H}_{t-})}{\Delta t \, |\mathrm{d}\boldsymbol{s}|} \,, \tag{2.3}$$

where $\mathrm{d}\boldsymbol{s}$ is an infinitesimal sphere around location $\boldsymbol{s}$. Note that for an ordinary Poisson process, the independence or memorylessness property yields that its conditional intensity function equals the unconditional intensity function, i.e. the information in $\mathcal{H}_{t-}$ is not needed.

Because of its dependency on the past history, the conditional intensity function is especially useful if interest lies in modelling the evolution of the point process. Consider a parametric model for the CIF with parameter vector $\boldsymbol{\theta}$ and a completely observed realisation $\boldsymbol{x} = \{(t_i, \boldsymbol{s}_i) : i = 1, \ldots, N\} \subset (0, T] \times W$ of a spatio-temporal point process in the time period $[0, T]$ in the spatial region $W \subset \mathbb{R}^2$. Then, the likelihood regarded as a function of $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}; \boldsymbol{x}, T, W) = \left[\prod_{i=1}^{N} \lambda^*_{\boldsymbol{\theta}}(t_i, \boldsymbol{s}_i)\right] \cdot \exp\left(-\int_0^T \int_W \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s}\right)$$

leading to the log-likelihood

$$l(\boldsymbol{\theta}; \boldsymbol{x}, T, W) = \left[\sum_{i=1}^{N} \log \lambda^*_{\boldsymbol{\theta}}(t_i, \boldsymbol{s}_i)\right] - \int_0^T \int_W \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s} \,. \tag{2.4}$$

Note that I follow the convention in statistics to denote the natural logarithm by "log", which is otherwise also written as "ln".

For continuous space-time point processes, the evaluation of the log-likelihood (2.4) thus incorporates integration over time and over the two-dimensional domain W. Depending on the specific model this might be analytically intractable and in fact the "direct evaluation of the integral term [...] is seldom straightforward" (Diggle, 2007, p. 24). Because this is also the case for the models in Chapter 4, statistical inference demands numerical methods of integration (Chapter 3).

Diggle et al. (2009) propose a partial likelihood for spatio-temporal point processes as an adaption of the partial likelihood for proportional hazards models in survival analysis. This approach at least avoids temporal integration by conditioning on the observed event times and considering the likelihood for the observed time-ordering of the events. The probability for observing the ith event at time point i is

$$p_i = \frac{\lambda(t_i, \boldsymbol{s}_i | \mathcal{H}_{t_i-})}{\int_{\mathcal{R}(t_i)} \lambda(t_i, \boldsymbol{s} | \mathcal{H}_{t_i-}) \, \mathrm{d}\boldsymbol{s}} \,. \tag{2.5}$$

Here, $\mathcal{R}(t_i)$ denotes the at-risk set at time t_i. In spatially discrete processes, $|W| < \infty$ and

$$\mathcal{R}(t_i) = \Big\{\boldsymbol{s}_j : t_j \geq t_i, \; j \in \{1, \ldots, N\}\Big\} \cup \Big\{\boldsymbol{s} \in W : \text{no event at } \boldsymbol{s} \text{ by time } T\Big\}$$

such that the integral is actually a sum (for an application see Diggle, 2006), whereas in spatially continuous processes, events may occur anywhere in the region W, hence $\mathcal{R}(t_i) = W$. The partial likelihood is the product of all those probabilities p_i for $i = 1, \ldots, N$, and hence the partial log-likelihood writes

$$l_p(\boldsymbol{\theta}; \boldsymbol{x}, T, W) = \sum_{i=1}^{N} \log(p_i) \,. \tag{2.6}$$

In survival analysis, estimation based on the partial log-likelihood (2.6) is generally accepted because maximum partial log-likelihood estimators inherit the general asymptotic properties of full maximum likelihood estimators, such as asymptotic normality. However, there may be a loss of efficiency and "the exact conditions under which the method gives consistent estimation for spatial point process models have yet to be established" (Diggle, 2007, p. 25). Some parameters might also be unidentifiable from the partial likelihood – just like the baseline hazard in proportional hazard models, which cancels out in the p_i fraction (2.5).

Marked spatio-temporal point process

If the events are stochastically *marked* by additional characteristics, the basic format of the data becomes

$$\boldsymbol{x} = \{(t_i, \boldsymbol{s}_i, \boldsymbol{m}_i) : i = 1, \ldots, N\},$$

where $\boldsymbol{m}_i$ are further *marks* (covariates) related to the ith event. The symbol $\mathcal{M}$ shall denote the *mark space*, i.e. the domain where the marks come from. If e.g. the only mark was the finetype of the IMD case, then $\mathcal{M} = \{1, 2\}$, where the mark value 1 corresponds to finetype B:P1.7-2,4:F1-5, and value 2 corresponds to finetype C:P1.5,2:F3-3. In other applications, marks could also be the altitude where a lightning hits the surface of the earth, the type or the financial amount of damage, or the magnitude of an earthquake. Furthermore, it is even possible to formally consider the spatial location as one component of a multidimensional mark of a temporal point process.

A *marked spatio-temporal point process* can be uniquely described by a CIF of the form $\lambda^*(t, \boldsymbol{s}, \boldsymbol{m})$, where the *internal history* $\boldsymbol{\mathcal{H}} \equiv \{\mathcal{H}_t : t \geq 0\}$ of the process now also collects values of preceding marks (Daley & Vere-Jones, 2003, Proposition 7.3.IV.). For the IMD cases, the length of the infectious periods, the gender and the age can be considered as *unpredictable marks*, i.e. they are independent of the internal history of the process, but they can influence its subsequent evolution. For instance, young people probably spread the disease at a higher rate and the longer they are infective the more infections they can cause. In such

situations it is often convenient to write

$$\lambda^*(t, \boldsymbol{s}, \boldsymbol{m}) = \lambda^*(t, \boldsymbol{s}) \cdot f_{\mathcal{M}}(\boldsymbol{m}|t, \boldsymbol{s}) \,, \tag{2.7}$$

where $f_{\mathcal{M}}$ is the density of the mark distribution at a specific time t and location s, and the $\boldsymbol{\mathcal{H}}$-intensity $\lambda^*(t, \boldsymbol{s})$ implicitly depends on the marks through the internal history $\boldsymbol{\mathcal{H}}$ of the process (cf. Daley & Vere-Jones, 2003, eq. (7.3.3)). More generally, if the process depends on some exogeneous variables (like e.g. the cases of IMD might depend on waves of influenza), the conditional intensity function is actually controlled by some larger history, which also incorporates the pasts of all related jointly evolving processes. If the point process further has initial conditions, these are all gathered in the initial history $\mathcal{H}_0$. It may e.g. contain individuals, which are already infective at the beginning of the observation period.

Using the above decomposition of the marked CIF and omitting an explicit parametrisation, the likelihood of a realisation $\boldsymbol{x}$ is expressible in the form

$$\begin{aligned} L &= \left[\prod_{i=1}^{N} \lambda^*(t_i, \boldsymbol{s}_i, \boldsymbol{m}_i)\right] \cdot \exp\left(-\int_0^T \int_W \int_{\mathcal{M}} \lambda^*(t, \boldsymbol{s}, \boldsymbol{m}) \,\mathrm{d}t \,\mathrm{d}\boldsymbol{s}\, \nu_{\mathcal{M}}(\mathrm{d}\boldsymbol{m})\right) \\ &= \left[\prod_{i=1}^{N} \lambda^*(t_i, \boldsymbol{s}_i)\right] \left[\prod_{i=1}^{N} f_{\mathcal{M}}(\boldsymbol{m}_i|t_i, \boldsymbol{s}_i)\right] \cdot \exp\left(-\int_0^T \int_W \lambda^*(t, \boldsymbol{s}) \,\mathrm{d}t \,\mathrm{d}\boldsymbol{s}\right) \,, \end{aligned} \tag{2.8}$$

where $\nu_{\mathcal{M}}$ is the reference measure on $\mathcal{M}$, and the second line follows from the assumption that the density $f_{\mathcal{M}}$ is proper, i.e. integrates to unity (Daley & Vere-Jones, 2003, Proposition 7.3.III.). The corresponding log-likelihood is

$$l = \left[\sum_{i=1}^{N} \log \lambda^*(t_i, \boldsymbol{s}_i) - \int_0^T \int_W \lambda^*(t, \boldsymbol{s}) \,\mathrm{d}t \,\mathrm{d}\boldsymbol{s}\right] + \left[\sum_{i=1}^{N} \log f_{\mathcal{M}}(\boldsymbol{m}_i|t_i, \boldsymbol{s}_i)\right] \,, \tag{2.9}$$

where the second term is an ordinary i.i.d. log-likelihood for the unpredictable marks, and the first term prima facie looks like the log-likelihood (2.4) of an unmarked spatio-temporal point process. But observe that the history of the process hidden in the asterisk of the CIF here also contains past values of the marks which influence the future evolution of the process.

However, in this thesis no attempt is made to model additional marks like gender and age but they are taken as given predictor variables in models of the CIF. In this case, the first term in (2.9) can still be maximised as a kind of partial likelihood. Actually, "the *full likelihood* would relate to a joint model for all the participating processes, including the predictor variables", and, "such a full representation might be the ultimate goal of the modelling stage, but one not easily realizable in practice, so such partial likelihoods are in widespread use" (Vere-Jones, 2009, p. 180).

In the IMD application, special attention only receives the finetype of the meningococci related

to an infection, if the cases of both finetypes are to be analysed in a common point process model. The finetype does not only influence the future of the process but is itself influenced by the process. For clarification imagine a regional outbreak of a specific finetype. During this outbreak further events ("offspring") are generated according to the degree of infectivity of this finetype, and the instantaneous rate of infections with this same finetype in the region of the outbreak is increased. Therefore, this mark can not be treated as given but it makes up an additional dimension of the point process. Following equation (2.8) the log-likelihood (or rather partial log-likelihood) of the complete point pattern of both finetypes then is of the form

$$l \;=\; \sum_{i=1}^{N} \log \lambda^*(t_i, \boldsymbol{s}_i, \kappa_i) - \int_0^T \int_W \sum_{\kappa \in \mathcal{K}} \lambda^*(t, \boldsymbol{s}, \kappa)\, \mathrm{d}t\, \mathrm{d}\boldsymbol{s}\,, \tag{2.10}$$

where $\mathcal{K}$ here corresponds to the set $\{1, 2\}$ of the two finetypes.

In order to prevent confusions with the partial likelihood of Diggle et al. (2009), where conditioning is with respect to the observed event times, I will reserve the adjective "partial" for this approach and omit it in the widespread case where only unpredictable marks are considered as given.

2.2 A Preliminary Point Process Model

This section briefly overviews some preliminary point process based strategies from the literature, which can be used for analysing infectious disease data. Hereby, an important distinction is to be made concerning the resolution of the data in question: the temporal and the spatial dimension may each be regarded as a discrete or a continuous space leading to different statistical approaches.

A coarse resolution of time would rule out the modelling of the transmission of infectious diseases. In such a case, methods of *disease mapping* for spatial lattice data may be applied, e.g. a Poisson regression model. Region-specific random effects based on a Gaussian Markov random field can be included to account for the neighborhood structure of the spatial units. An example of application is the analysis of the yearly number of infections with tick-borne encephalitis in the counties of the federal state Baden-Württemberg carried out by Meyer & Wimmer (2009).

In this thesis, the data are assumed to be continuous in time, and non-aggregated point-process based analyses are the intended approach. The important case, where both the temporal and the spatial dimensions are continuous, i.e. the events occur in a space-time continuum, will be treated elaborately in the separate Chapter 4 with application in Chapter 5. It is also the framework, in which the IMD data fit best, but which is most rarely found in the

literature. Unfortunately, "generic methods for analysing spatio-temporal data-sets have not yet become well established" (Diggle, 2007, p. 7). Elias et al. (2006, p. 1689) also complain about the lack of objective methods in disease surveillance.

In `R`, purely spatial point patterns are treated by the packages `spatstat` (Baddeley & Turner, 2005), `spatial` (Venables & Ripley, 2002), and `splancs` (Rowlingson, Diggle et al., 2009). Popular methods of spatial statistics like the K-function will not be investigated in this thesis, because considering the 7-year summary of the IMD data serves only little for describing the dynamics of the infectious disease. Purely time-dependent point process models specified through their conditional intensity function can be analysed with the package `PtProcess`, which is part of the *Statistical Seismology Library* (SSLib) concerned with analysing earthquake events. Retrospective temporal modelling and temporal cluster detection using statistical monitoring is provided by the `R` package `surveillance` (Höhle, 2007). A brief exemplification using the IMD data can be found in Höhle (2009b). In fact, only very few open source software is available for the analysis of point patterns in space *and* time. For spatially discrete infectious disease data, the package `DCluster` (Gómez-Rubio, Ferrándiz-Ferragud & Lopez-Quílez, 2005) is concerned with the detection of spatial disease clusters (similar to what is provided by the `SaTScan` software).

The package `RLadyBug` (Höhle & Feldmann, 2007; Höhle, Feldmann & Meyer, 2009) implements the spatio-temporal point process model proposed by Höhle (2009a). Here, the spatial observation region is regarded as a finite and fixed subset $W = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ of $\mathbb{R}^2$. The process is then viewed as a sequence $\{N_i\}_{i=1,\ldots,n}$ of purely temporal point processes which interact with one another. The original application considered in Höhle (2009a) was the stochastic modelling of transmission of the classical swine fever virus among domestic pig farms (which make up the discrete observation "region" W).

The aim of the proposed model for the conditional intensity function is to statistically evaluate retrospectively the spread of the disease. Requiring the spatio-temporal regression model to be capable of modelling multiple outbreak data leads to a decomposition of the risk of infection into two components: an *endemic component*, which models imported cases, and an *epidemic component*, which represents infections by direct transmission inside the population. The CIF then has the form

$$\lambda^*(t, \boldsymbol{s}_i) =: \lambda_i^*(t) = Y_i(t) \cdot \{h_i(t) + e_i^*(t)\} .$$

Here, $Y_i(t)$ is the left-continuous at-risk indicator for individual i, which is 1 if the individual is susceptible at time t, and 0 otherwise. The endemic component $h_i(t)$ is modelled as

$$h_i(t) := \exp(h_0(t) + \boldsymbol{z}_i(t)'\boldsymbol{\beta}) ,$$

which expresses the endemic risk of infection in the form of a Cox model. Possible time-

dependent covariates $\boldsymbol{z}_i(t)$ act multiplicatively on the baseline risk $\exp(h_0(t))$. Conversely, the history-dependent epidemic component $e_i^*(t)$ models the additive disease pressure on individual i caused by the set $I^*(t)$ of infectious individuals just before time t:

$$e_i^*(t) := \sum_{j \in I^*(t)} f(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|) \, .$$

Here, each infectious individual transmits the disease to the susceptible individual i with rate according to some (usually non-increasing) function f of the distance to the susceptible individual. In Höhle (2009a), the epidemic component is actually proposed to also depend on the temporal distance between the susceptible and infectious individuals, and on additional covariates specific to each pair (i, j) of a susceptible and an infectious individual.

Note that the model includes two well known special cases: the Cox model, if no epidemic component is specified, and the compartmental SIR model, if no endemic component is considered. The case $f(u) = \gamma_0$ equals the homogeneous SIR model with transmission parameter $\gamma_0 > 0$, i.e. the degree of infectivity does not depend on the mutual distance between the individuals. The model also looks similar to the additive-multiplicative approach proposed in survival analysis, see e.g. Martinussen & Scheike (2002). However, it differs essentially in that the multivariate counting process of all individuals now has dependent paths.

For the IMD data, this compartmental model framework could be artificially entered by considering Germany's counties as the "individuals" $i = 1, \ldots, 413$ and define a county as being infectious if it contains at least one infectious patient. At any time point, a county can thus be disease-free, i.e. susceptible, or infectious. Constrained maximum likelihood estimation as implemented in the `R` package `RLadyBug` could then be performed using e.g. a SIS (susceptible-infectious-susceptible) model for the counties, where the endemic component contains population and influenza effects. However, such artificial county-based modelling of the IMD data is not attractive, because it likely obscures dynamics of the IMD cases and does not represent the true nature of events. Instead, this additive-multiplicative conditional intensity model, which I call `twinSIR`, will serve as a prototype of the continuous spatio-temporal modelling persued in Chapter 4.

3 Numerical Integration

Approaching the aim of spatio-temporal point processes continuous in space-time, this chapter deals with numerical integration. Specifically, likelihood-based analysis of parametric spatio-temporal CIF models is concerned with the evaluation of the log-likelihood from equation (2.4). If a closed-form of the multidimensional integral term in the log-likelihood is not available, then its calculation involves numerical integration. Therefore, various methods for calculating or approximating numerically the definite integral of functions in one and especially two dimensions will be presented and compared. Readers with less interest in the numerical details of the computation may skip this chapter and proceed with the statistical modelling in Chapter 4.

The well-developed one-dimensional case of numerical integration is also called *quadrature* and is sketched in Section 3.1. The books of Davis & Rabinowitz (1984), Krommer & Ueberhuber (1998) and Press, Teukolsky, Vetterling & Flannery (2007, Chapter 4) serve as main references. For the more interesting two-dimensional case, also called *cubature*, integration over polygons is of special interest as the spatial units, e.g. the borders of Germany or its subregions, are represented as (sets of) polygons. Hence, Section 3.2 focuses on methods dealing with this issue. A number of algorithms are exemplified, compared to each other and their applicability for the inference problem at hand is discussed. This is especially interesting, because most of the literature on spatial point processes do not treat the issue of numerical integration explicitly.

3.1 Quadrature

A large volume of methods has been proposed for numerical integration in the basic one-dimensional case. What they have in common is their form of approximation by adding up weighted values of the integrand evaluated at a number of abscissae:

$$\int_a^b f(x)\,\mathrm{d}x \approx \sum_{j=1}^{n} w_j f(x_j) \tag{3.1}$$

Here, w_j are the weights and x_j the abscissae (aka nodes or evaluation points) of this n-point quadrature formula. Note that we restrict the task to functions f that are bounded and

Riemann-integrable on the finite interval $[a, b]$. Computational integration over an unbounded range or integrand is e.g. described in Krommer & Ueberhuber (1998, Sections 7.2 and 7.3, respectively).

In the following, I will sketch the classical Newton-Cotes quadrature rules, the more sophisticated method of Gaussian quadrature, and Monte Carlo integration, which provides a stochastic estimation of the integral.

3.1.1 Newton-Cotes Rules

Classical quadrature rules boil down to exactly integrating a polynomial interpolation of the integrand over the interval $[a, b]$. For instance, this can be a constant, linear or quadratic interpolation – corresponding to the *midpoint rule*, the *trapezoidal rule* and *Simpson's rule*, respectively. These rules are known as *Newton-Cotes formulae*. Usually they are applied in their *compound* form, i.e. the interval $[a, b]$ is divided into subintervals of equal length and the quadrature rule is applied to each subinterval. In doing so, the overall approximation of the integrand becomes more precise and the whole integral obviously equals the sum of the integral pieces. The advantage of the polynomial interpolation is that its integral can be expressed in a closed form and the equidistant grid results in simple formulae, e.g. in case of the compound midpoint rule with n subdivisions one obtains

$$\int_a^b f(x)\,\mathrm{d}x = \sum_{j=1}^{n} \int_{a+(j-1)h}^{a+jh} f(x)\,\mathrm{d}x \approx \sum_{j=1}^{n} h \cdot f(a + (j - \tfrac{1}{2})\,h) \qquad \text{where } h = \frac{b-a}{n}\,. \tag{3.2}$$

The compound midpoint and trapezoidal rules converge to the true value of the integral at least as fast as n^{-2}, if f is twice continuously differentiable. Simpson's rule converges with rapidity n^{-4} at worst (Davis & Rabinowitz, 1984, pp. 54, 58).

A more sophisticated polynomial interpolation of the integrand that abandons the equidistance of the abscissae is e.g. used in *Clenshaw-Curtis quadrature*, namely the Chebyshev approximation (cf. Press et al., 2007, Section 5.8 and 5.9.1). The next method, Gaussian quadrature, goes even further by determining optimal abscissae and weights to attain the maximum possible degree of accuracy for n interpolatory points.

3.1.2 Gaussian Quadrature

The Gaussian quadrature rule is of the form

$$\int_a^b w(x)\,f(x)\,\mathrm{d}x \approx \sum_{j=1}^{n} w_j f(x_j) \tag{3.3}$$

known as *product formula* (Krommer & Ueberhuber, 1998, p. 119), i.e. it includes a known nonnegative weight function $w(x)$ in the integrand.

The idea of Gaussian quadrature is to optimise not only the weights w_j, but also the evaluation points x_j leading to twice the degrees of freedom compared to Newton-Cotes formulae. The solution is to determine a set $\{P_d^w : d = 0, 1, 2, \ldots\}$ of pairwise orthogonal polynomials P_d^w of degree d with respect to the weight function w over the interval $[a, b]$, i.e.

$$\int_a^b w(x) P_s^w(x) P_t^w(x) \, \mathrm{d}x = 0 \qquad \forall s \neq t \in \mathbb{N}_0 \, .$$

The n zeros of the polynomial P_n^w, which are real, simple, and located in the interior of $[a, b]$ (Krommer & Ueberhuber, 1998, Theorem 5.2.9) are chosen as clever evaluation points in the n-point Gaussian quadrature formula. Because then, the weights w_j can be determined explicitly in such a way that the quadrature (3.3) is exact whenever $f(x)$ is a polynomial of degree $d \leq 2n - 1$ (Davis & Rabinowitz, 1984, p. 97). This is also the maximum degree of accuracy of an n-point quadrature formula of the form (3.3) and attained if and only if the quadrature rule is chosen as the Gaussian quadrature formula with respect to w (Krommer & Ueberhuber, 1998, Theorem 5.2.10). In general, the approximation of the integral is the better the more accurate f can be approximated by a polynomial of maximal degree $2n - 1$. Press et al. (2007, p. 1089) also state that compared to Newton-Cotes rules, "the main advantage of Gaussian integration is that it converges exponentially with n for smooth functions" f.

The calculation of the nodes and weights in Gauss formulae is well studied for the most important weight functions and intervals, see e.g. Press et al. (2007, Section 4.6.1). The function `gauss.quad` of the `R` package `statmod` (Smyth, 2009) provides an implementation. In this thesis I will only be concerned with the so-called *Gauss-Legendre* case, because it will be used as part of a two-dimensional integration method. Here, $w(x) = 1$, $a = -1$, $b = 1$ and hence the orthogonal polynomials in use are the *Legendre polynomials* (cf. Davis & Rabinowitz, 1984, Section 1.12).

In order to make this Gauss-Legendre quadrature formula applicable to other intervals, it has to be transformed as explained in Krommer & Ueberhuber (1998, Section 5.4.1). For instance, if $\tilde{x}_j$ and $\tilde{w}_j$, $j = 1, \ldots, n$, are the original nodes and weights of the n-point Gauss-Legendre rule for the interval $[-1, 1]$, then

$$x_j = \frac{b - a}{2} \, \tilde{x}_j + \frac{a + b}{2} \qquad \text{and} \qquad w_j = \frac{b - a}{2} \, \tilde{w}_j$$

are the corresponding data for the n-point Gauss-Legendre rule on the interval $[a, b]$.

3.1.3 Monte Carlo Integration

Also a product formula (cf. equation (3.3)), but a completely different approach using randomly sampled evaluation points is Monte Carlo integration. The idea is to interpret the integral in question as the expectation of a random variate, which can be estimated by the mean of independent and identically distributed (i.i.d.) replications of this random variate: Let $w(x)$ be a probability density on the domain $[a, b]$, i.e.

$$w(x) \geq 0,\ \forall x \in [a, b] \qquad \text{and} \qquad \int_a^b w(x)\,\mathrm{d}x = 1\,.$$

Hence

$$I_w(f) := \int_a^b w(x)\,f(x)\,\mathrm{d}x = \mathbb{E}(f(X))\,,$$

where $\mathbb{E}$ denotes expectation and X is a random variate following the probability density w. Due to the Strong Law of Large Numbers, the integral can be estimated by the mean of an i.i.d. sample from the distribution $\mathcal{L}(X)$ of X such that the estimate

$$\hat{I}_n = \frac{1}{n}\sum_{j=1}^{n} f(X_j) \qquad \text{with} \qquad X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{L}(X) \tag{3.4}$$

converges almost surely towards the expected value $\mathbb{E}(f(X))$ as $n \to \infty$. As this estimate is unbiased, its root mean square error (RMSE) equals its standard deviation (sd)

$$\mathrm{RMSE}(\hat{I}_n) = \sqrt{MSE(\hat{I}_n)} = \sqrt{\mathbb{E}\Big((\hat{I}_n - I_w(f))^2\Big)} = \sqrt{\mathrm{Var}(\hat{I}_n) + \mathrm{Bias}^2(\hat{I}_n)} = \mathrm{sd}(\hat{I}_n)$$

and by Jensen's inequality, the expected absolute integration error is bounded by the RMSE:

$$\mathbb{E}\Big(|\hat{I}_n - I_w(f)|\Big) = \mathbb{E}\Big(\sqrt{(\hat{I}_n - I_w(f))^2}\Big) \leq \sqrt{\mathbb{E}\Big((\hat{I}_n - I_w(f))^2\Big)} = \mathrm{RMSE}(\hat{I}_n) = \mathrm{sd}(\hat{I}_n)\,. \tag{3.5}$$

Thus, the rate of convergence in Monte Carlo integration can be measured by the standard deviation of the estimate,

$$\mathrm{sd}(\hat{I}_n) = \frac{\mathrm{sd}(f(X))}{\sqrt{n}}\,,$$

where $\mathrm{sd}(f(X))$ is the standard deviation of the (transformed) random variate $f(X)$. This is known as the "$1/\sqrt{n}$ law" meaning that one can expect the integration error to be reduced by a factor of 10 (one decimal place) when increasing the number of function evaluations by a factor of 100. Although this expected convergence rate is slow compared to the deterministic quadrature rules described previously, it has the advantage of being independent of the dimension of X and the integration region, respectively. The Monte Carlo method is thus especially appealing for multidimensional integration, where it can be applied in an analogous

manner.

In contrast to the above quadrature methods the root mean square error is a probabilistic estimate of the integration error and does not permit guaranteed bounds on the actual error. If the standard deviation of $f(X)$ in (3.1.3) is not known, it is usually estimated from the Monte Carlo sample $f(x_1), \ldots, f(x_n)$ resulting in the standard error of the integral estimate

$$\operatorname{se}(\hat{I}_n) = \frac{1}{\sqrt{n}} \cdot \left(\frac{1}{n-1} \sum_{j=1}^{n} (f(x_j) - \hat{I}_n)^2 \right)^{\frac{1}{2}} .$$

The standard error can not only be used as a point estimate of the absolute integration error, but also to provide asymptotic confidence intervals constructed by the Central Limit Theorem. Let $1 - \alpha$ be the desired statistical certainty, e.g. 95%, and z_q be the quantile of the standard normal distribution such that $\Phi(z_q) = q$, where Φ denotes the cumulative distribution function of the standard normal distribution. Then the confidence interval

$$\left[\hat{I}_n - z_{1-\alpha/2} \operatorname{se}(\hat{I}_n) \, , \, \hat{I}_n + z_{1-\alpha/2} \operatorname{se}(\hat{I}_n) \right] \tag{3.6}$$

asymptotically covers the true integral value $\mathbb{E}(f(X))$ with a probability of $1 - \alpha$.

Another way to reduce the expected error besides increasing the sample size n is to introduce an alternative *positive* probability density on $[a, b]$, say $v(x)$, and reformulate the integral in question:

$$I_w(f) = \int_a^b w(x)\, f(x)\, \mathrm{d}x = \int_a^b \frac{w(x)}{v(x)} f(x)\, v(x)\, \mathrm{d}x = I_v\Big(\frac{w}{v} f\Big) . \tag{3.7}$$

As in (3.4) above, the integral can then be estimated by

$$\frac{1}{n} \sum_{j=1}^{n} \frac{w(Y_j)}{v(Y_j)}\, f(Y_j) \qquad \text{with} \qquad Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} v(x) .$$

If v is chosen in such a way that $\frac{w(x)}{v(x)} f(x)$ is approximately constant, i.e. the transformed random variate $(\frac{w}{v} f)(Y)$ exhibits approximately zero standard deviation, then the integral estimate features only a small standard deviation and the integration error generally decreases. This approach is known as *importance sampling*, but note that most authors introduce Monte Carlo integration using uniformly sampled points without an explicit weight function, which they reserve for importance sampling. Note as well that the rearrangement (3.7) is also useful if it is much easier to sample points from the alternative density v than from w. If otherwise the sampling domain poses the difficulty, it is possible to enlarge the integration region and set f to 0 outside $[a, b]$. As only points in $[a, b]$ contribute to the integral, the Monte Carlo sample has to be larger to retain accuracy. This trick is very valuable in multidimensional integration as presented in the next section.

3.2 Cubature

In the present spatio-temporal integration, the vast amount of literature on one-dimensional integration is only of limited use. Recall the integral term in the log-likelihood (2.4):

$$\int_0^T \left(\int_W \lambda_{\boldsymbol{\theta}}^*(t, \boldsymbol{s}) \, \mathrm{d}\boldsymbol{s} \right) \mathrm{d}t \,,$$

where $W \subset \mathbb{R}^2$ is the two-dimensional observation region. If one could evaluate the inner integral for fixed t by some cubature rule, then it would be possible to use one of the univariate quadrature rules of the previous section for the outer integral and repeatedly evaluate the inner one.

However, as stated in Press et al. (2007, Section 4.8), "integrals of functions of several variables, over regions with dimension greater than one, are *not easy*". On the one hand, the product of the number of function evaluations needed for each dimension usually explodes, i.e. the accuracy with n evaluation points in one dimension approximately equals the accuracy with n^3 evaluation points in three dimensions. On the other hand, the region of integration, which in one dimension is a simple interval given by two numbers, might be very complicated in two or more dimensions. For instance, the observation region W for the IMD data is Germany, whose border is computationally represented by a set of polygons (cf. Subsection 1.2.1).

In the face of these difficulties, we are seeking for a cubature rule

$$\int_W f(\boldsymbol{s}) \, \mathrm{d}\boldsymbol{s} \approx \sum_{j=1}^{n} w_j f(\boldsymbol{s}_j)$$

for two-dimensional integration regions W set up by polygons, which may also be non-convex or "multiply-connected" (i.e. contain holes) but they must be *simple*, i.e. not self-intersecting. In general, the domain W may be expressed as the union of the simple polygons $P_1, P_2, \ldots, P_p$ from which the simple polygonal holes $H_1, H_2, \ldots, H_q$ are substracted:

$$W = (P_1 \cup \cdots \cup P_p) \setminus (H_1 \cup \cdots \cup H_q)$$

The *enclosing rectangle* or *bounding box* of W is denoted by $\overline{W} = [x_l, x_r] \times [y_b, y_t]$ and it is the smallest rectangle such that $\overline{W} \supset W$.

As an example, consider the integration of the density

$$f(\boldsymbol{s}; \boldsymbol{\mu}, \sigma^2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2} \left(\frac{\|\boldsymbol{s} - \boldsymbol{\mu}\|}{\sigma} \right)^2 \right) \tag{3.8}$$

of the isotropic (radially symmetric) bivariate normal distribution $\mathrm{N}_2(\boldsymbol{\mu}; \sigma^2 \boldsymbol{I}_2)$, which will be essential for the intensity models in the next chapter. The integration region W might for instance be set up by three polygons P_1 (concave), P_2 (concave) and H (convex) such that $W = (P_1 \cup P_2) \setminus H$. The setting is illustrated in Figure 3.1, where the contour lines belong to the bivariate normal density with $\boldsymbol{\mu} = \mathbf{0}$ and $\sigma^2 = 25$.

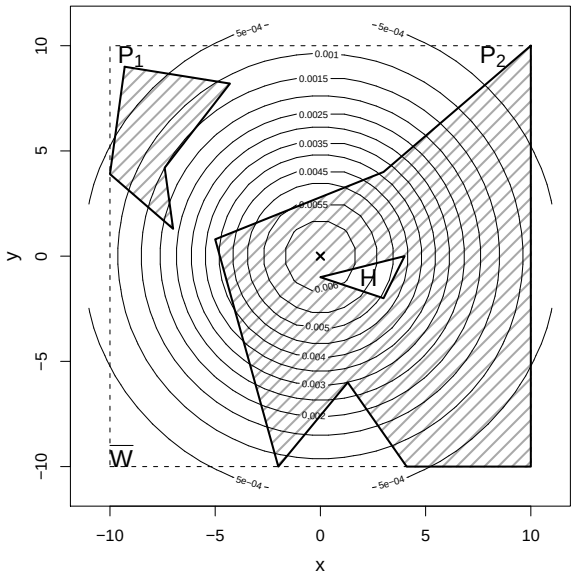


Figure 3.1: Example polygonal region $W = (P_1 \cup P_2) \setminus H$ (shaded), enclosing rectangle $\overline{W} = [-10, 10] \times [-10, 10]$ of W, and contour lines of the bivariate normal distribution with zero mean and covariance matrix $25 \cdot \boldsymbol{I}_2$.

The task is to numerically approximate the integral as fast and as accurate as possible. This will obviously require a trade-off between both aims. The computational cost of the calculation becomes very important during the optimisation of the log-likelihood, where the log-likelihood and hence the integral has to be evaluated for many different parameter configurations.

As possible cubature methods over polygons, I examine two-dimensional generalisations of the midpoint and trapezoidal rules, adaptive cubature, Monte Carlo integration, a two-dimensional Gauss-Legendre rule proposed by Sommariva & Vianello (2007b), and a method specific for the bivariate normal density using formulae found in the Abramowitz & Stegun handbook. In order to find out which of these methods fits best to the purposes of this thesis, I have implemented them in `R`. The following subsections present and exemplify the methods with the setting of Figure 3.1 The names of the corresponding `R` functions available from the attached CD (see Appendix A) are given in parantheses in the titles of the subsections. Afterwards, the cubature methods will be compared with respect to both accuracy and speed.

3.2.1 A Two-Dimensional Midpoint Rule (`polyint.midpoint`)

A generalisation of the compound midpoint rule of formula (3.2) to two dimensions is straightforward. The domain W of the integrand or rather its enclosing rectangle $\overline{W} = [x_l, x_r] \times [y_b, y_t]$ is subdivided into a regular grid of rectangles (pixels) corresponding to the subintervals in the one-dimensional case. The function is then evaluated at the rectangle midpoints, which lie inside W and the weights correspond to the areas of the rectangles. Defining n_x and n_y as the numbers of subintervals in the x and y directions, respectively, with corresponding bandwidths

$$h_x := \frac{x_r - x_l}{n_x} \qquad \text{and} \qquad h_y := \frac{y_t - y_b}{n_y} \,,$$

the set of evaluation points becomes

$$\left\{ \left(x_l + (j - \tfrac{1}{2})\, h_x \,,\; y_b + (k - \tfrac{1}{2})\, h_y \right) : j = 1, \ldots, n_x,\; k = 1, \ldots, n_y \right\} \,\cap\, W \,.$$

Accordingly, the two-dimensional midpoint rule is

$$\int_W f(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s} \approx h_x h_y \sum_{j,k:(x_j,y_k)\in W} \sum f((x_j, y_k)) \,.$$

If the integrand is more or less constant in one direction, but very bumpy in the other, one would choose a smaller bandwidth in the latter. In the case of the isotropic bivariate normal density it is rather appropriate to have $h_x = h_y$ leading to a grid of squares. Obviously, the accuracy of the integral approximation increases with an increasing number of evaluation points.

The only difficulty of this cubature rule is the need for a method, which decides whether a grid point lies inside W. An implementation of the necessary components already exists in the `R` package `spatstat` (Baddeley & Turner, 2005), see `as.im` and `summary.im`, which also works for our example since version 1.17-0 of the package (see the corresponding release notes).

Figure 3.2 illustrates the method using $n_x = n_y = 25$ and hence $h_x = h_y = \frac{20}{25} = 0.8$ resulting in 326 function evaluations and an approximate integral value of 0.5474141.

3.2.2 A Two-Dimensional Trapezoidal Rule (`polyint.linear`)

When constructing a two-dimensional trapezoidal rule the main question is how the polygonal domain W gets *triangulated*, i.e. how it is subdivided into triangles. The subdivision determines the piecewise linear interpolation of the integrand used in the trapezoidal cubature rule. It is plausible that there is no unique solution to this task without further constraints.

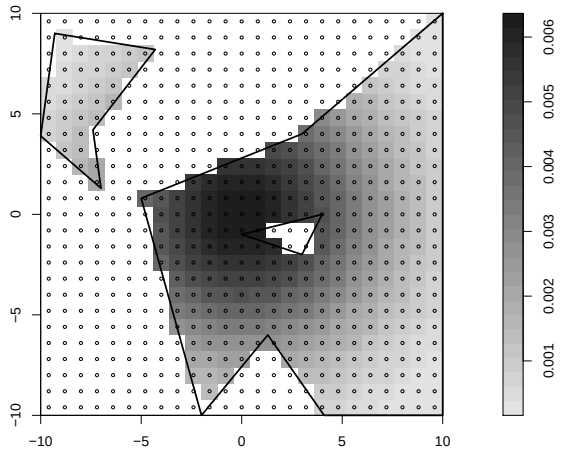
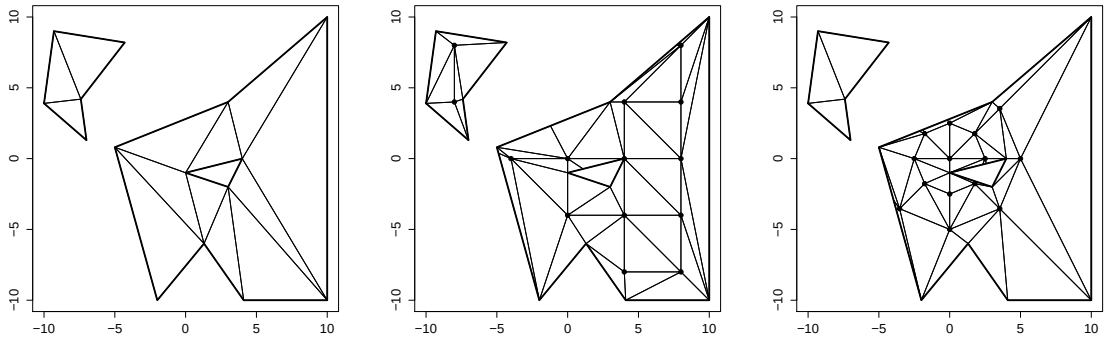


Figure 3.2: Illustration of the two-dimensional midpoint rule on a 25×25 grid.

A famous triangulation type is the *Delaunay triangulation* (Press et al., 2007, Section 21.6.1), where the constraint is that the circumcircle of any triangle contains no other vertices (but there are other definitions, too). "It is the triangulation whose triangles best avoid small angles and long sides" (Press et al., 2007, p. 1131). The Delaunay triangulation is unique if no three points (vertices) are collinear and no four lie on the same circle.

Figure 3.3 shows Delaunay triangulations of the example region using different sets of auxiliary nodes. They have been constructed using the `spatstat` function `delaunay`. In variant (a) the region is triangulated using only the vertices of the polygonal domain, whereas a 5×5 grid of auxiliary nodes is used in (b). In alternative (c) auxiliary nodes are radial from the mean $\boldsymbol{\mu} = \mathbf{0}$ of the bivariate normal density with points at distance $\sigma/2$ and σ. Note that the triangulations are not unique in the latter two cases.



(a) Without auxiliary nodes. (b) 5×5 grid. (c) Radial auxiliary nodes.

Figure 3.3: Delaunay triangulation of the example region using different sets of auxiliary nodes.

Obviously, the accuracy of the approximation highly depends on this choice of auxiliary nodes. Here, the results are (a) 0.5211882, (b) 0.4588546, (c) 0.5487904. Using the same 25×25 grid as in the midpoint cubature rule (see Figure 3.2), the approximation becomes 0.5287144. As stated in Press et al. (2007, p. 1142), "you need a *lot* of triangles to get a reasonable representation of any function with much detailed structure." However, the construction of the Delaunay triangulation needs a significant amount of computation time, which grows as $O(n \log n)$, where n is the number of nodes. The whole procedure would presumably be faster with a more efficient `C` implementation like that in Held (2001), which is not yet available in the public domain.

3.2.3 Adaptive Cubature (`polyint.adapt`)

In iterative, adaptive quadrature or cubature rules, the evaluation grid is subsequently refined until the estimated relative integration error falls below a prespecified bound ε. Adaptive integration routines thus account for the nature of the integrand and distribute the evaluation points predominantly in subdomains where the numerical integration of the integrand is difficult. A popular adaptive cubature rule is that implemented by A. C. Genz & Malik (1980) in the FORTRAN subroutine ADAPT. It integrates functions over d-dimensional hypercubes, where $2 \leq d \leq 15$. The `R` package `adapt` provided an interface to this method (the package has been removed from the online repositories but old versions can still be obtained from the archive `http://cran.at.r-project.org/src/contrib/Archive/adapt/`).

In order to integrate over other domains than the d-dimensional hypercube one needs to incorporate the characteristic function of the domain, i.e.

$$\int_W f(\boldsymbol{s})\,\mathrm{d}s = \int_{\overline{W}} \mathbb{1}_W(\boldsymbol{s})\, f(\boldsymbol{s})\,\mathrm{d}s\,, \tag{3.9}$$

where $\overline{W}$ is the enclosing hypercube. For a two-dimensional polygonal domain this makes again necessary a method deciding if a point lies inside a polygon. In `R`, this functionality is provided by `point.in.polygon` from the package `sp` (Pebesma & Bivand, 2005).

Applying the adaptive cubature method to the example of Figure 3.1 with $\varepsilon = 0.01$ gives an approximate integral value of 0.4829737 using 379 function evaluations. The estimated absolute integration error is 0.002412295.

3.2.4 Monte Carlo Integration (`polyint.mc.uniform`, `polyint.mc.gauss`)

Monte Carlo integration is the preferred method for complicated integrands or complex regions of integration, when the application of other methods becomes too sophisticated. Its big

advantage is the easy implementation: one only needs to sample evaluation points in the integration region W or some enclosing region.

It works in two and more dimensions in the exact same way as in one dimension, which has already been described in Section 3.1 above (see pp. 30ff.).

Choosing the uniform distribution over the enclosing rectangle $\overline{W} \supset W$ for sampling the evaluation points, i.e. $w(\boldsymbol{s}) = \frac{\mathbb{1}_{\overline{W}}(\boldsymbol{s})}{|\overline{W}|}$ the integral can be rewritten as in (3.9) and (3.7):

$$\int_W f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = \int_{\overline{W}} \mathbb{1}_W(\boldsymbol{s})\,\frac{f(\boldsymbol{s})}{w(\boldsymbol{s})}\,w(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = |\overline{W}| \cdot \int_{\overline{W}} \mathbb{1}_W(\boldsymbol{s})\,f(\boldsymbol{s})\,w(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = |\overline{W}| \cdot \mathbb{E}(\overline{f}(\boldsymbol{S}))\,,$$

where the random variate $\boldsymbol{S}$ is uniformly distributed on $\overline{W}$ with density $w(\boldsymbol{s})$ and

$$\overline{f}(\boldsymbol{s}) = \begin{cases} f(\boldsymbol{s}) & \text{if } \boldsymbol{s} \in W, \\ 0 & \text{otherwise.} \end{cases}$$

This yields the estimate

$$\hat{I}_n = |\overline{W}| \cdot \frac{1}{n} \sum_{j=1}^{n} \overline{f}(\boldsymbol{S}_j) = \frac{(x_r - x_l)(y_t - y_b)}{n} \sum_{j:\boldsymbol{S}_j \in W} f(\boldsymbol{S}_j)$$

with n i.i.d. replications $\{\boldsymbol{S}_j : j = 1, \ldots, n\}$ of $\boldsymbol{S}$, i.e. of a binomial process of n points in $\overline{W}$ with density w. Of course, the implementation again depends on the `point.in.polygon` function, which decides whether a sampled point lies inside W.

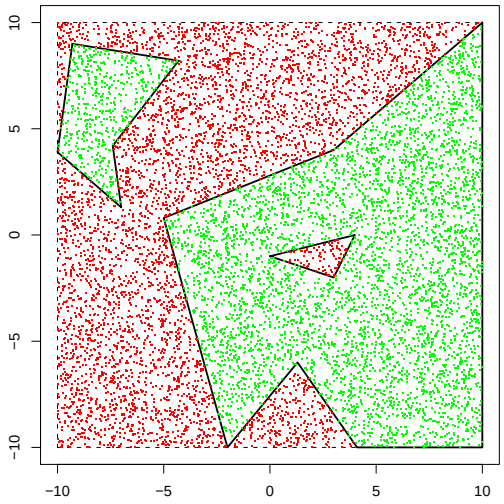
Using Monte Carlo integration with $n = 10\,000$ for the example domain and density results in an approximate integral value of 0.5433528 (with the given seed value). Figure 3.4(a) shows the realised set of uniformly distributed evaluation points, 5 212 of which fall into the domain W and are actually evaluated. On a side note, this number gives a simple Monte Carlo estimate of the area of W: $\frac{5\,212}{10\,000}\,|\overline{W}| = 208.48$ (the true value is 207.895). The RMSE of the integral estimate $\hat{I}_n$ – an upper bound for the expected absolute integration error (cf. equation (3.5)) – equals the standard deviation of $\hat{I}_n$, which is estimated as 0.007461972.

If the integrand f already is a density from which we can generate (pseudo-)random numbers, another idea of Monte Carlo integration is to observe that

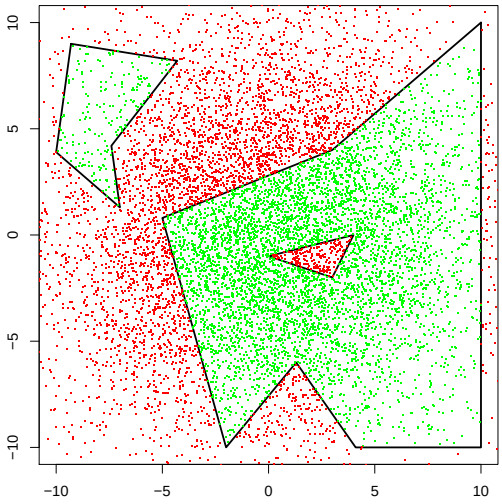
$$I(f) = \int_W f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = P(\boldsymbol{S} \in W) = \mathbb{E}(\mathbb{1}_W(S)) \qquad \text{if} \qquad \boldsymbol{S} \sim f\,.$$

The Monte Carlo estimate of the integral using n i.i.d. replications $\{\boldsymbol{S}_j : j = 1, \ldots, n\}$ of $\boldsymbol{S}$ is thus

$$\tilde{I}_n = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_W(\boldsymbol{S}_j) = \frac{|\{j : \boldsymbol{S}_j \in W\}|}{n}$$



(a) 10 000 uniformly distributed points in $\overline{W}$, of which 5 212 fall inside W.



(b) 10 000 points distributed as $\mathrm{N}_2(\mathbf{0}, 25\boldsymbol{I}_2)$, of which 5 423 fall inside W.

Figure 3.4: Two types of Monte Carlo integration (points are green inside W, otherwise red).

with a standard deviation (RMSE) and a standard error of

$$\mathrm{sd}(\tilde{I}_n) = \left(\frac{I(f)\,(1 - I(f))}{n}\right)^{\frac{1}{2}} \qquad \text{and} \qquad \mathrm{se}(\tilde{I}_n) = \left(\frac{\tilde{I}_n\,(1 - \tilde{I}_n)}{n}\right)^{\frac{1}{2}} \leq \frac{1}{2\sqrt{n}}\,,$$

respectively.

Figure 3.4(b) shows a sample of size $n = 10\,000$ from the corresponding $\mathrm{N}_2(\mathbf{0}, 25\boldsymbol{I}_2)$ distribution forming the point estimate 0.5423 with a standard error of 0.004982075.

However, if other cubature methods are available, the use of Monte Carlo integration during the optimisation procedure is not favourable as it is usually slower. Another problem is that the estimated integral values are subject to chance and so would be the Maximum Likelihood estimates.

3.2.5 Product Gauss Cubature (`polyint.SV`)

Iterated integration (Krommer & Ueberhuber, 1998, Theorem 1.1.5) can be used to reduce the two-dimensional integral to a composition of one-dimensional integrals. For this purpose one has to determine the integration range $(x_l(y), x_r(y))$ in the horizontal direction as a function of the vertical position y (or vice versa). Then,

$$\int_W f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = \int_{y_b}^{y_t} \int_{x_l(y)}^{x_r(y)} f(x, y)\,\mathrm{d}x\,\mathrm{d}y\,,$$

where (y_b, y_t) is the vertical range covered by W and $f((x, y))$ is conveniently written as $f(x, y)$.

A similar approach is taken by Sommariva & Vianello (2007b), who describe a cubature method over arbitrary polygons. It is based on Green's integration formula, which also transforms a two-dimensional into a one-dimensional problem. Let Π denote a polygon – for instance one of the polygonal components of W (holes or not) – given by the anticlockwise sequence of vertices $V_1, \ldots, V_L$ with $V_{L+1} = V_1$. By Green's integration formula,

$$\int_\Pi f(x, y)\,\mathrm{d}(x, y) = \sum_{l=1}^{L} \int_{[V_l, V_{l+1}]} F(x, y)\,\mathrm{d}y$$

where $[V_l, V_{l+1}]$ denotes the lth polygon side between the vertices V_l and V_{l+1}, and $F(x, y)$ is any fixed x-primitive of $f(x, y)$, for instance

$$F(x, y) = \int_{x_0}^{x} f(u, y)\,\mathrm{d}u$$

with fixed x_0.

Figure 3.5 illustrates the procedure for $f(x, y) = 1$, i.e. the integral of f over the polygon Π simply equals the enclosed area. Hence the function $F(x, y) = x - x_0$ has to be integrated along the edges of the polygon. In the Figure, the polygon is a simple triangle described by the vertices V_1, V_2 and V_3, and $x_0 = 2$ was chosen. The value of $F(x, y)$ can be seen from the corresponding length of the shading line, but it is negative if $x < x_0$. Note that the areas outside the triangle do not contribute to the overall integral, as they are always included twice but with different signs: The line integrand $F(x, y)$ contributes with its inverse sign on decreasing edges (in the direction of the arrows).

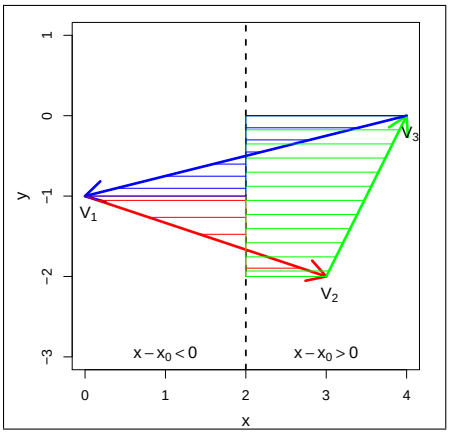


Figure 3.5: Piecewise line integration of $F(x, y) = x - x_0$, $x_0 = 2$.

Now, the idea is to compute the inner integral $F(x, y)$ for fixed y by univariate n-point Gauss-Legendre quadrature and integrate it as a function of y along the sides of the polygon again by univariate Gauss-Legendre quadrature. The cubature rule then has the form

$$\iint_\Pi f(x, y)\,\mathrm{d}x\,\mathrm{d}y \approx \sum_{l=1}^{L} \sum_{i=1}^{n_l} \sum_{j=1}^{n} w_{lij}\, f(\xi_{lij}, \eta_{li}) \tag{3.10}$$

where $n_l = n$ if the side l is parallel to the base-line $x = x_0$ and $n_l = n + 1$ otherwise. Strictly speaking, the sum is only over the sides not being orthogonal to the base-line and not lying

on it, because the horizontal line integral with no change in y equals 0, and $F(x_0, y) = 0, \forall y$, respectively. Explicit expressions for the weights w_{lij} and nodes (ξ_{lij}, η_{li}) as well as the proof of this cubature rule are given in Sommariva & Vianello (2007b, Theorem 2.1). Note that n and n_l denote the numbers of points of the above mentioned univariate Gauss-Legendre rules. Furthermore, the accuracy of the univariate Gauss-Legendre rules transfers to the Gauss-like cubature rule, i.e. it is exact over Π for all bivariate polynomials of degree $2n - 1$ at most. "Thus, when it is applied to a polynomial of degree d, it is sufficient to choose $n \geq (d+1)/2$ to get the integral up to machine precision" (Sommariva & Vianello, 2007b, Remark 2.3). The overall number of cubature nodes resulting from a specific choice of n can be derived from equation (3.10): it is bounded below by $\frac{L}{4}n^2$ – if only every forth side of the polygon is not orthogonal to the base-line and not lying on it (this was mistaken by the authors, who give $\frac{L}{2}n^2$ as the lower bound) – and above by $L\,n\,(n+1)$ – if no side is parallel to the x or y axis. In general, the actual number is rather at the upper bound and not all of the cubature nodes fall inside the polygon Π, but inside its enclosing rectangle. The integrand f must thus be continuous and computable also there.

The authors have implemented their "product Gauss cubature over polygons" in MATLAB as available from their homepage (see Sommariva & Vianello, 2007a). I ported their code to `R`, where I make use of the function `gauss.quad` from the package `statmod` (Smyth, 2009) to compute the weights and nodes of the one-dimensional Gauss-Legendre quadrature rules. In the original implementation it is also possible to rotate the polygon in such a way that more nodes fall inside the polygon and a higher accuracy is achieved (cf. Sommariva & Vianello, 2007b, Remark 2.4). This feature is not included in my `R` port as it is in general not applicable to non-convex polygons. However, I made a small modification concerning the choice of the base-line $x = x_0$, which in the MATLAB code is always set to $x_0 = 0$ when no rotation is applied. In my `R` port, x_0 is set to the horizontal midpoint of the bounding box, which leads to higher accuracy, because more nodes fall into the polygon. The difference is illustrated in Figure 3.6. It would be even worse for a polygon situated far away from $x = 0$.

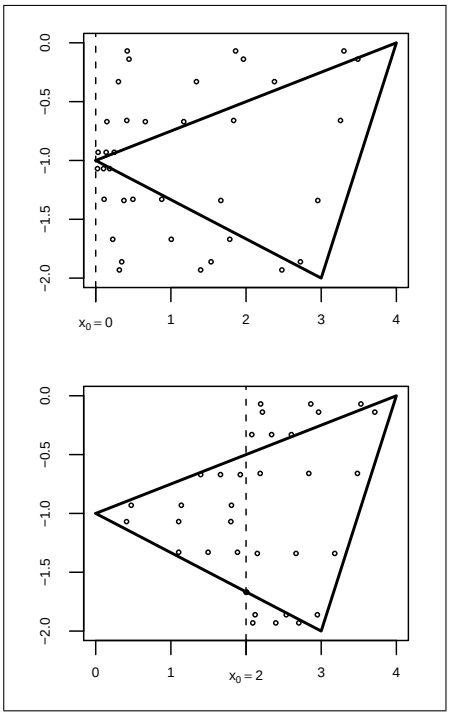


Figure 3.6: Cubature nodes resulting from the original base-line $x_0 = 0$ (top) and the adapted base-line (bottom) with $n = 3$.

Figure 3.7 shows the cubature nodes for the example region resulting from the choice $n = 5$ for every polygon. The overall number of nodes hence function evaluations is 415. Note that

nodes outside W *do* contribute to the cubature rule. The weights are non-zero but they may be negative.

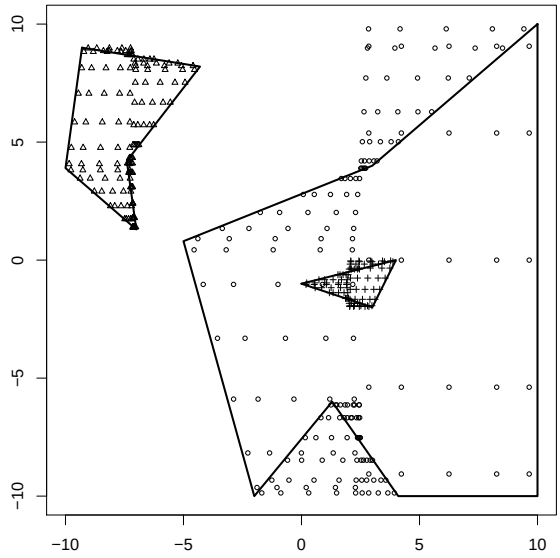


Figure 3.7: Nodes of the product Gauss cubature rule with $n = 5$ applied to each of the three polygons (marked by $\circ$, $\triangle$ and $+$, respectively).

Obviously, the integral has to be solved on a polygon-by-polygon basis, substracting the integral value of the hole H from the integral over the domain $P_1 \cup P_2$, i.e.

$$\int_W f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = \int_{P_1} f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} + \int_{P_2} f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} - \int_H f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\,,$$

where each sub-integral is approximated by the above described product Gauss cubature rule. For the example density f from Figure 3.1 this results in an approximate integral value of 0.5435167.

3.2.6 Specific Methods for the Bivariate Normal Density (`polyint.AS`)

Besides the above cubature rules for the approximate numerical integration of general two-dimensional functions over a polygonal domain, there exist at least two methods for the specific case of the bivariate normal density. DiDonato & Hageman (1982) evaluate the integral over the complement of the polygon and implemented a FORTRAN program with approximately 9-decimal-digit accuracy. Calculating the integral in this way incorporates a polynomial approximation of a scaled primitive and its implementation seems quite cumbersome which is why I did not port their method to `R`. Instead, I have used explicit formulae found in the famous Abramowitz & Stegun handbook in Chapter 26 authored by Zelen &

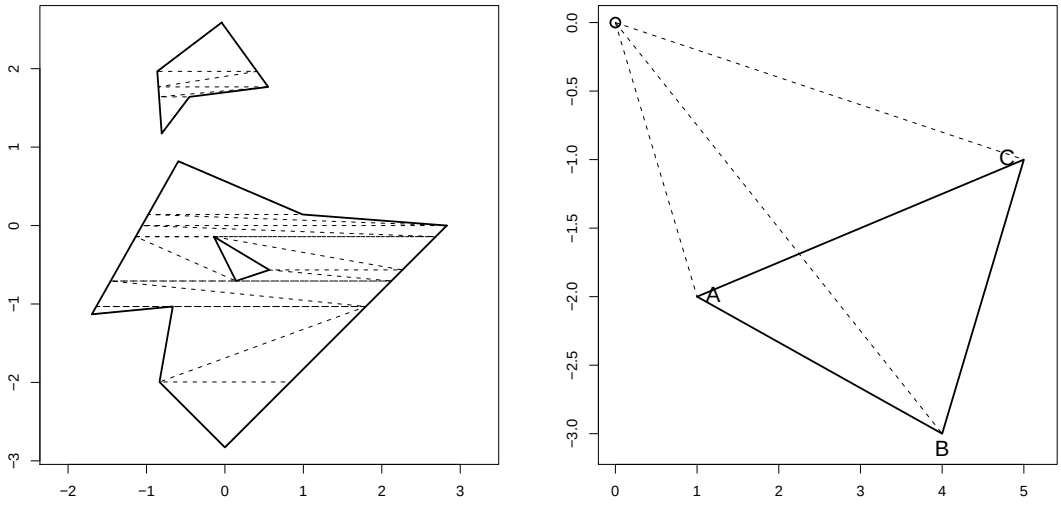
Severo (1970, Section 26.9, Example 9, pp. 956f.). These formulae treat the calculation of the integral of the bivariate normal density over triangles $\triangle AOB$, where one vertex is the origin $O = (0, 0)$. In the following, I will give a brief summary of the approach.

As a first step, the coordinates of the polygon in question are transformed so that we integrate the *standard* bivariate normal density with zero mean and the identity matrix as covariance matrix (cf. Zelen & Severo, 1970, formula 26.3.22). Then, the polygon is divided into triangles by some triangulation procedure. I use the `tristrip` functionality from the General Polygon Clipping `C` library (Murta, 2009) to which the `R` package `gpclib` (Peng, 2009) provides an interface. This is simpler and runs much faster ($> 1\,000$ times) than a Delaunay triangulation with the `spatstat` function `delaunay`. Figure 3.8(a) shows how those two preliminary steps work on the example domain. As the above cited formulae only cover triangles of the type $\triangle AOB$, the integral over each triangle has to be calculated by the decomposition illustrated in Figure 3.8(b). The signs of the three auxiliary integrals depend on the relative position of the origin and can thus be determined automatically. The sign of an auxiliary triangle is positive, if the origin and the remaining point, which is not used in the auxiliary triangle, lie on the same side of the line through the other two points of the auxiliary triangle. Otherwise it is negative, like e.g. for the auxiliary triangle $\triangle COA$, where the origin and the remaining point B are on different sides of the line through A and C. Finally, the core of the formulae and also their computational bottleneck is the calculation of the integral of the bivariate normal density with zero mean, unit variance and some correlation ρ over some infinite rectangle $[h, \infty] \times [0, \infty]$ (cf. Zelen & Severo, 1970, formula 26.3.23). This task can be performed very accurately by the `R` function `pmvnorm` of the package `mvtnorm` (A. Genz et al., 2009).

Applying the outlined method to the example results in an approximated integral value of 0.5434421 with a total of 138 `pmvnorm` evaluations and an estimated absolute error of $1.26 \cdot 10^{-13}$. Unfortunately, it is the slowest of all the implemented methods. The next subsection gives a detailed comparison of the implemented cubature functions concerning both accuracy and speed.

3.2.7 Comparison of the Presented Methods

The preceding descriptions and exemplifications of the implemented cubature methods reveal substantial differences concerning complexity and accuracy. Gathering all the approximated integral values one observes a range from 0.4588546 to 0.5487904. In default of an exact reference value of the integral, one can e.g. consider the quasi exact Abramowitz & Stegun method, which gives the narrow interval $[0.543442134975419, 0.543442134975671]$ for the true value, or take a look at the convergence of the Sommariva & Vianello method when increasing



(a) "tristrip" of the transformed domain.

(b) $\int_{\triangle ABC} = \int_{\triangle BOA} + \int_{\triangle COB} - \int_{\triangle COA}$.

Figure 3.8: Illustration of the "tristrip" triangulation in (a) and the decomposition of the integral over a triangle $\triangle ABC$ using three triangles passing by the origin in (b).

the number n of univariate Gauss-Legendre nodes. This is shown in Table 3.1 using 16 decimal digits (machine epsilon is $1.11 \cdot 10^{-16}$).

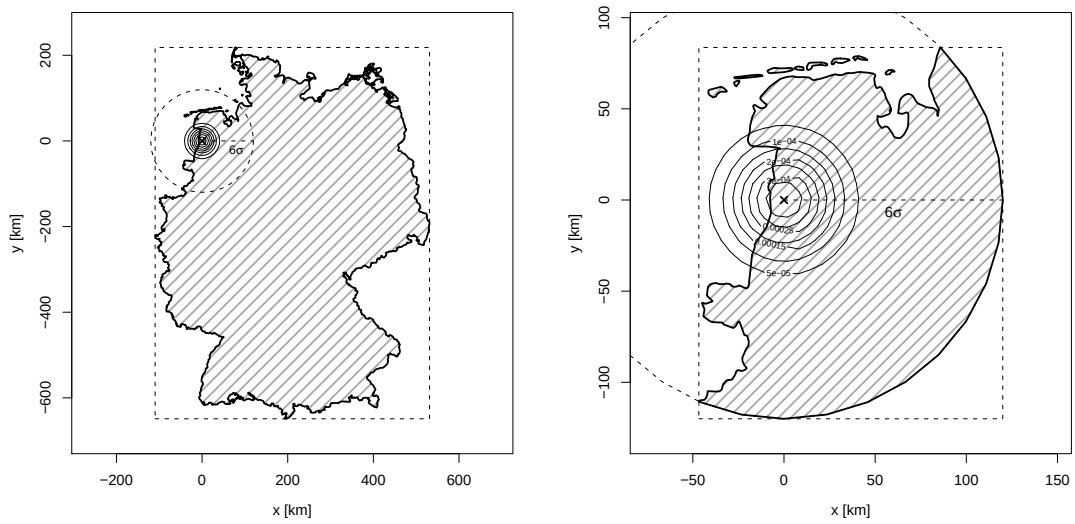
n	N	result	n	N	result	n	N	result
1	27	0.7308917307681466	7	777	0.5434425697317347	13	2535	0.5434421349755513
2	82	0.5049443902231224	8	1000	0.5434421079898190	14	2926	0.5434421349755448
3	165	0.5496474086072402	9	1251	0.5434421364665062	15	3345	0.5434421349755459
4	276	0.5426842459879074	10	1530	0.5434421349013328	16	3792	0.5434421349755448
5	415	0.5435166972422180	11	1837	0.5434421349789060	17	4267	0.5434421349755449
6	582	0.5434359942759850	12	2172	0.5434421349754058	30	12990	0.5434421349755451

Table 3.1: Convergence of the product Gauss cubature rule for an increasing number of nodes. The column "n" denotes the requested number of univariate Gauss-Legendre nodes, whereas the column "N" denotes the resulting total number of two-dimensional nodes.

Thus in this example, as of $n = 12$ the Gaussian cubature is as accurate as the quasi exact method, but much faster. It also seems that machine accuracy has been reached with $n = 14$. It is clearly a useful feature of the Gaussian cubature that one can influence the computation time by requesting a specific degree of accuracy. Apart from the two-dimensional midpoint rule, the other implemented rules seemed to be less accurate in relation to the computational effort. Specifically the `adapt` routine did perform surprisingly bad since it returned the interval $[0.4805614, 0.4853860]$ for the true integral value. However, a detailed comparison of accuracy and runtime requires a realistic and more complicated integration region than used so far.

For this reason, I consider the integration of the isotropic bivariate normal density with

a standard deviation of $\sigma = 20$ [kilometres] centred near the north-western borderline of Germany over the whole region of Germany as shown in Figure 3.9(a). Note that the contour lines in the Figure are shown only up to a density value of $5 \cdot 10^{-5}$.



(a) Whole Germany.

(b) Intersection of Germany and 6σ "circle".

Figure 3.9: Polygonal domains with enclosing rectangle, contour lines of the $\mathrm{N}_2(\mathbf{0}, 400\boldsymbol{I}_2)$ distribution and a 6σ "circle".

One observes that in a predominant part of Germany the density is almost 0. The probability mass over a circular domain around the mean can also be calculated as a function of the radius r (cf. Zelen & Severo, 1970, formula 26.3.24): Let $X, Y \overset{iid}{\sim} N(0, \sigma^2)$, then we search the probability that $X^2 + Y^2 \leq r^2$. Because $(X/\sigma)^2$ and $(Y/\sigma)^2$ are independently distributed following the $\chi^2(1)$-distribution, their sum $Z := (X/\sigma)^2 + (Y/\sigma)^2$ is distributed as $\chi^2(2)$. Thus the probability in question can be calculated as $\mathbb{P}(Z \leq (r/\sigma)^2)$ in `R` by the command `pchisq((r/sigma)^2, 2)`. For instance, 99.999998 percent of the probability mass are concentrated up to a distance 6σ [kilometres] from the mean.

The computation will be faster and often also more accurate if the integration region is narrowed to this range as shown in Figure 3.9(b). Actually, the circular domain is computationally represented as a polygon with 32 vertices and its intersection with Germany is computed by the `intersect` method of the package `gpclib` (Peng, 2009), which again results in a polygon. The intersection procedure took 0.036 seconds.

Table 3.2 contains the cubature results for both integration regions (parts (a) and (b), respectively) using various configurations of the afore presented cubature methods. Hereby, the configuration parameters recorded in the second column have the following meanings:

for `midpoint`: the number of subdivisions in each dimension ($n_x = n_y$).

for `linear`: the number of auxiliary triangulation nodes.

for `adapt`: $\log_{10}(\varepsilon)$, where ε is the upper bound for the relative integration error.

for `mc.*`: the number of sampled points.

for `SV`: the degree of the univariate Gauss quadrature.

for `AS`: there is no parameter choice.

Concerning the linear trapezoidal rule it is actually also of great importance, *where* the auxiliary triangulation nodes are located. For this test, I chose the radial configuration as in Figure 3.3(c) with a point at the center and points at σ and 2σ distance in eight directions. Note that this cubature rule has not been applied to the whole region of Germany, because the Delaunay triangulation would take too long.

The column "♯ eval." basically contains the number of evaluations of the bivariate normal density. However, for the `AS` method it is the number of evaluations of the cumulative distribution function `pmvnorm`, and for `mc.gauss` it is the characteristic function of the polygon, which is evaluated for each sampled point.

As the integral values are (only) listed in 10-digit accuracy, we may take the two results in the last row achieved by the Abramowitz & Stegun method as the exact integral values. They differ slightly on the eighth digit after the decimal point, which is in accordance with the proportion of the probability mass outside the 6σ circle ignored in the setting of Figure 3.9(b). Regarding the runtime of this method it is clear that it can not be used for the log-likelihood evaluations during the optimisation routine. As already mentioned, the same is true for the two Monte Carlo methods, also because of the random nature of their results. However, the table points out that `mc.gauss` would be the better alternative because its estimates exhibit a smaller standard deviation. This is especially noticeable in part (a), where relatively few of the points sampled *uniformly* over whole Germany fall into the main 6σ region of the density peak.

Concerning the other four cubature methods the competition obviously remains between the `midpoint` and the `SV` method. The Delaunay triangulation for the `linear` trapezoidal rule takes much too long, which impedes a dens grid of auxiliary nodes necessary for accurate results. The `adapt` routine scores 2-digit accuracy, but that takes much longer than with the `midpoint` and `SV` methods.

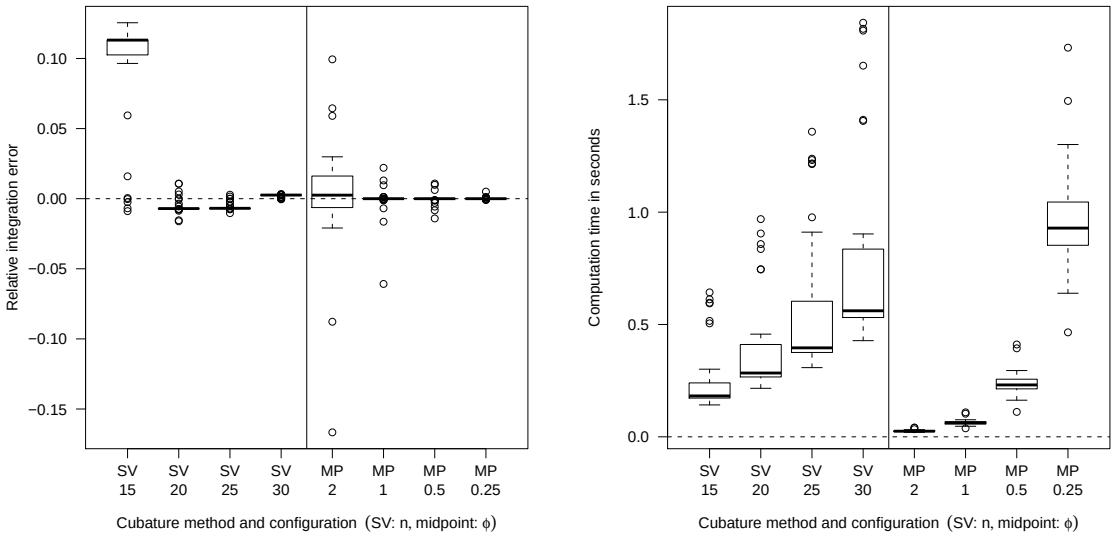
Now comparing the two best performing methods, the two-dimensional `midpoint` rule appears to be less affected by the type of the domain. Accuracy and speed are only slightly worse when integrating over whole Germany in (a) than in (b). The product Gauss cubature (`SV`) needs a significantly higher number of nodes and also more computation time to attain the exact integral value when integrating over whole Germany. This means that the `SV` method is

		(a) Results for Germany			(b) Results for the 6σ-region		
Method	Parameter	♯ eval.	Time	Integral value	♯ eval.	Time	Integral value
midpoint	10	66	0.03	0.0908082261	67	0.01	0.7590965857
midpoint	20	260	0.03	0.7595256939	264	0.01	0.6466609261
midpoint	50	1604	0.07	0.6783876114	1644	0.02	0.6763635916
midpoint	60	2327	0.09	0.6229258021	2372	0.02	0.6804794248
midpoint	100	6435	0.13	0.6899259967	6568	0.03	0.6817244161
midpoint	200	25740	0.25	0.6767119337	26270	0.08	0.6808681436
midpoint	500	160848	2.69	0.6804283121	164143	0.56	0.6810896454
linear	17	-	-	-	929	29.10	0.7882673149
adapt	-1	527	1.81	0.6874490883	561	0.51	0.6927104248
adapt	-2	1779	4.86	0.6714258336	1419	1.38	0.6947242080
adapt	-3	5387	21.41	0.6821925229	5523	6.13	0.6827739466
adapt	-4	14917	50.66	0.6809311801	24395	26.67	0.6782005452
mc.uniform	1000	644	0.29	0.6293671822	647	0.04	0.6612873856
mc.uniform	5000	3224	2.06	0.5506513848	3245	0.17	0.6494419377
mc.uniform	10000	6426	3.78	0.6165596709	6575	0.35	0.6840521611
mc.uniform	50000	32007	25.65	0.6999019599	32725	1.71	0.6922238995
mc.uniform	100000	64202	38.15	0.6595451869	65744	3.43	0.6749424592
mc.gauss	1000	1000	0.35	0.6750000000	1000	0.04	0.6750000000
mc.gauss	5000	5000	1.94	0.6696000000	5000	0.17	0.6696000000
mc.gauss	10000	10000	3.75	0.6817000000	10000	0.35	0.6817000000
mc.gauss	50000	50000	17.96	0.6766400000	50000	1.72	0.6766400000
mc.gauss	100000	100000	28.98	0.6826000000	100000	3.44	0.6826000000
SV	1	15548	0.80	0.0000990873	1836	0.10	0.7428566470
SV	2	46644	0.83	0.4297683105	5508	0.11	0.6639635988
SV	3	93288	0.87	0.9053810992	11016	0.11	0.6810251230
SV	5	233220	1.02	0.6441308961	27540	0.12	0.6810984697
SV	10	855140	2.72	0.6811809816	100980	0.18	0.6810971143
SV	15	1865760	3.97	0.6810970557	220320	0.28	0.6810971143
AS	-	90924	100.76	0.6810971213	10608	11.78	0.6810971139

Table 3.2: Results from the cubature test runs: The parameter column indicates for `SV` the degree of the univariate Gauss quadrature, for `midpoint` the number of subdivisions in each dimension, for `linear` the number of auxiliary nodes, for `adapt` $\log_{10}(\varepsilon)$, and for the Monte Carlo methods the number of sampled points. The column "♯ eval." gives the actual number of function evaluations. The runtime is given in seconds.

much more sensitive with respect to the concentration of the function's mass on a small region of the integration domain like in Figure 3.9(a), and equivalently with respect to varying σ^2 for a fixed polygonal domain. In contrast, when only integrating over the main 6σ region, it basically reaches the exact integral value in 0.18 seconds, whereas the `midpoint` method is only at 4 digits in 0.56 seconds. Depending on the desired minimum accuracy the method of choice (given such a 6σ intersection) would therefore be `midpoint`, which is as fast as 0.02 seconds for 2 digits, and `SV` if 3 or more digits are required. Furthermore, it would be possible to implement the computation of the bivariate Gauss-Legendre nodes and weights in `C`, which would make the `SV` method certainly come off as winner.

However, for the purpose of log-likelihood maximisation, the sensitivity of the Gaussian cubature with respect to σ^2 is a big issue. Because during optimisation a variety of values



(a) Relative integration error.

(b) Computation time in seconds.

Figure 3.10: Benchmark of the Gaussian cubature (SV) and the two-dimensional midpoint rule (MP) on integrating the "narrow" bivariate normal density with $\sigma = 1\,[\text{km}]$ over 35 large domains with a range of up to 300 km. The configuration parameters are for `SV` the number n of univariate Gauss-Legendre nodes and for `midpoint` the factor $\phi = \frac{h}{\sigma}$, which here equals the bandwidth h (in kilometres). The `AS` method was used to compute a reference value for the integral enabling the calculation of relative integration errors.

of σ^2 is tested, it would be either necessary to compute the 6σ intersection as in Figure 3.9(b) for each new value of σ^2, or to adaptively choose a number of knots guaranteeing enough accuracy. Since each computation of an intersection costs about 0.04 seconds, the evaluation of the log-likelihood for a single value of σ^2 would for the IMD data at least take $636 \cdot 0.04 = 25.44$ seconds – only to determine the 6σ intersections for all events. Since it is furthermore not clear how to adaptively choose the degree n as a function of σ^2 and because small σ^2 make necessary a very high and time consuming degree of the Gaussian cubature, the use of the more robust two-dimensional `midpoint` rule is fairly appealing. With this rule it is also straightforward how to retain accuracy when σ^2 varies: just choose the pixel width and height as a multiple of σ, i.e. $h := h_x = h_y = \phi\,\sigma$, where ϕ depends on the desired accuracy.

Figure 3.10 shows results of a benchmark test of the two methods. The task was to integrate the bivariate standard normal density $f(\boldsymbol{s}) = \exp(-\|\boldsymbol{s}\|^2/2)/(2\pi)$, which means having a very small value of $\sigma = 1$ kilometre. The test set consisted of the 35 unique locations of IMD cases, which are closer than 6 km to the border of Germany. For other locations of IMD cases, the 6σ circle completely falls inside Germany and hence there is nothing to do because the integral almost equals 1 anyway. For the selected 35 locations, the intersection of the surrounding circle of 150 km radius with Germany was generated and centred at the respective location

similar to Figure 3.9(b). This truncation has been applied in preview of a similar truncation of the spatial kernel in the next Chapter. For each resulting test region, a reference integral value has been computed by the `AS` method.

The Figure indicates that the product Gauss cubature with $n = 15$ systematically overestimates the integral and the high number of $n = 30$ univariate Gauss-Legendre nodes (yielding a total of 597.060 function evaluations) are necessary for almost exact answers. On the other hand the computational cost of this cubature with $n > 15$ is greater than for the midpoint rule with bandwiths σ ($\phi = 1$) and $\sigma/2$ ($\phi = 0.5$), which for most of the regions hit the true value of the integral very precisely. Furthermore the results show that choosing ϕ around these values yields fast results and sufficient accuracy for the purposes of this thesis. Nevertheless, the main advantage of the midpoint rule actually is that a similar relative accuracy would result if we used another value of σ, because the bandwidth $h = \phi\,\sigma$ is chosen appropriately.

To sum things up, this chapter presented and examined various methods of two-dimensional numerical integration with particular attention drawn to polygonal domains and the isotropic bivariate normal density. In the final comparative study, two methods performed very well: the simple two-dimensional midpoint rule (cf. Subsection 3.2.1) and the rather complex product Gauss cubature by Sommariva & Vianello (cf. Subsection 3.2.5). Although the latter converges faster to the true integral value, it is not suitable for the use during optimisation, because its accuracy strongly depends on the time-consuming clipping of the integration domain to a 6σ region. A possible workaround would be to precalculate these truncated domains for a set of values of σ and recall them as and when required.

Based on the results of this chapter I finally opt for the more robust and σ-adaptive two-dimensional midpoint rule to evaluate the spatial integral in the log-likelihood (2.4) during the optimisation routine. Having thus ensured the capability of evaluating the log-likelihood by numerical integration, the next chapter considers specific spatio-temporal intensity models.

4 Spatio-Temporal Modelling and Inference

This chapter seeks to specify space-time point process models for infectious disease data through the conditional intensity function. Main interest lies in explaining the evolution of the process based on its history, which also includes covariates like the numbers of influenza cases and event marks like the age of the infected person. In the spatio-temporal modelling in Section 2.2, the possible locations of the events constituted a discrete subset of the observation region. Now, the individual events are assumed to occur in a space-time continuum, which certainly better reflects the underlying nature of the IMD point patterns. Due to this generalisation, the estimation of such models is computationally more intensive, but the essential task of evaluating the spatio-temporal integral term in the log-likelihood (2.4) has been solved in the previous Chapter 3.

Various spatio-temporal point processes describing the dynamics of "infective events" are presented in Section 4.1. Here, a novel model class is proposed with a special view to infectious diseases. The estimation of the proposed model class by means of maximum likelihood (ML) will be discussed in Section 4.2. Section 4.3 will touch on the implementation of the data structure and the estimation procedure. The last Section 4.4 presents a different spatio-temporal modelling approach with a tricky estimation technique, which is however not applicable for the data analysis in Chapter 5.

4.1 Modelling Infection Dynamics

As an introduction to spatio-temporal modelling, the research on the dynamics of earthquake occurences boosted by Yosihiko Ogata and David Vere-Jones (cf. Ogata, 1999) will be overviewed in Subsection 4.1.1. Here, the strong relationship between modelling earthquakes and outbreaks of infectious diseases will become obvious. Subsequently, this Master's Thesis proposes a novel extension of the `twinSIR` model of Section 2.2 to continuous space in Subsection 4.1.2 and a further generalisation with marks in Subsection 4.1.3.

4.1.1 A Review on Modelling Earthquakes

Earthquakes and cases of infectious diseases have in common that they feature so-called *self-excitement*, i.e. events influence the future evolution of the point process by producing "offspring" events. Strong earthquakes cause subsequent aftershocks, and an infected individual may transmit the virus to susceptible individuals. Point process modelling of such clustering in space and/or time often inherits in structure from *Hawkes' self-exciting process* (Hawkes, 1971), which is a birth process with immigration. It was introduced as a purely temporal process with conditional intensity function

$$\lambda^*(t) = \psi + \int_{(-\infty,t)} g(t-u)\,\mathrm{d}N(u) = \psi + \sum_{j:t_j<t} g(t-t_j) \tag{4.1}$$

where $\psi > 0$ represents a constant "immigration" rate for new events independently of the evolution of the process, and the function $g(\cdot) \geq 0$ measures the increase in temporal clustering due to events of the past. Note that the range of the stochastic integral above is $(-\infty, t)$ (excluding t) so that the intensity process is left-continuous and hence predictable (g must also be a left-continuous function) (Ogata, 1978, Example 4 on p. 248). As in auto-regressive time series, "the expectation of an event occuring is given by a linear combination of past occurences, where [...] $g(\cdot)$ measures the weights of such combinations" (Ogata, 1999, p. 499). For example, Hawkes considered $g(t) = \alpha e^{-\beta t}$, i.e. an exponential decay of the rate to produce offspring. Thus one obtains the clustering phenomenon that offspring is more likely to occur shortly after the parent event. According to Hawkes & Oakes (1974) the Hawkes process can also be viewed as a cluster process, where the cluster centres equal the immigrants born at rate ψ. Associated with each immigrant is a cluster of subsidiary events formed by the births of all of the descendants of all generations of the immigrant – in contrast to the Neyman-Scott process, where the parent events (cluster centres) are hidden, and only one single generation of offspring is included.

Corresponding to the superposition of rates in (4.1), events may be either immigrants emerging from a homogeneous Poisson process with rate ψ or a descendent of a previous event. With regard to the `twinSIR` model of Section 2.2, one could equally speak of endemic or epidemic sources of infection. Starting from its "birth-time" each event produces offspring according to an inhomogeneous Poisson process with rate $g(t)$, where t is relative to the birth-time. Because this process is independent of the event's own parentage and of other events, $\mu := \int_0^\infty g(t)\,\mathrm{d}t$ can be interpreted as the average number of offspring per event. Clearly, population growth models also fit in this context of Hawkes processes – and both can be interpreted as general *branching processes*. In the theory of branching processes, μ is called the *basic reproduction number*. If $\mu < 1$, the process is called *sub-critical*, because it is of finite total size. A sub-critical Hawkes process *without* immigration (but with a fixed number

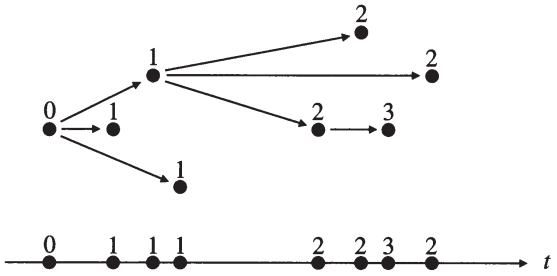


Figure 4.1: The branching structure of the various generations of events in a cluster (top), and the observed ground process on the time axis (bottom). Source: Møller & Rasmussen (2005, Figure 1)

of initial events) will almost surely die out, which is intuitively clear because each event on average produces less than one descendent.

Figure 4.1 illustrates the branching structure of a Hawkes process within a cluster initiated by the single event on the left. This event gives rise to three children one of which itself generates new offspring and so forth. Usually one only observes the occurence of events on the time axis without knowing about their heritage. For instance, infectious disease data in general do not contain information about the source of infection, whether by direct person-to-person contact or otherwise imported. The bottom line in the figure shows the corresponding temporal point pattern.

Inspired by the self-exciting Hawkes process, Ogata (1988) proposed the so-called *epidemic-type aftershock-sequences* (ETAS) model for earthquake occurences. The name of the seismological ETAS model clearly indicates a relationship between earthquakes and cases of infectious diseases so that many ideas of modelling and interpretation can be borrowed from one another. Specifically, the original ETAS model has CIF

$$\lambda^*(t) = \psi + \int_{(0,t)\times[M_0,\infty)} g(t-u)\, e^{\beta(m-M_0)}\, N(\mathrm{d}u \times \mathrm{d}m)\,, \tag{4.2}$$

where the mark m denotes the earthquake magnitude and M_0 is the cutoff magnitude of the data set. The observed marks introduce a weighting of the "epidemic" component of the model (following the *Gutenberg-Richter law* of magnitudes). Stronger earthquakes trigger aftershock-sequences at a higher, exponentially increasing rate (for $\beta > 0$). The function g is the *modified Omori formula* $g(t) = K\,(t+c)^{-p}$, with $K, c, p > 0$, which indicates hyperbolic decay of the aftershock rate. The complete integrand in (4.2) is usually called the *response function* or *triggering function*, because it describes how an earthquake triggers aftershocks as a function of time and magnitude.

Ogata's ETAS model has been extended to a spatio-temporal version also covering the loca-

tions of earthquakes, and a lot of alternatives for the triggering function have been considered in the seismological literature over the years (cf. Ogata, 1998). Especially interesting in the context of this thesis is the spatio-temporal formulation

$$\lambda^*(t, \boldsymbol{s}) = \psi(\boldsymbol{s}) + \sum_{j:t_j<t} \kappa(m_j)\, g(t - t_j)\, f(\boldsymbol{s} - \boldsymbol{s}_j|m_j)$$

in Ogata et al. (2003). Here, the background intensity of new "parent" events (similar to immigration) is represented by a now spatially varying intensity $\psi(\boldsymbol{s})$. The second *epidemic-type* component measures the rate of aftershocks at the space-time coordinate $(t, \boldsymbol{s})$ following an earthquake $(t_j, \boldsymbol{s}_j, m_j)$. The triggering function is decomposed multiplicatively into effects of the magnitude m_j, of the elapsed time $t - t_j$ since the triggering earthquake, and of the relative location $\boldsymbol{s} - \boldsymbol{s}_j$ with respect to the earthquake's epicentre. The magnitude-dependent "impact factor" $\kappa(m_j)$ of the earthquake and the function g are the same as in the original ETAS model. As functional form for the spatial kernel f, an elliptic bivariate normal density for anisotropic clusters is chosen, the parameters of which are determined a priori. Like in the above Hawkes model, "every event with a given magnitude has the same probability structure for generating direct 'offspring', irrespective of its own parentage" (Vere-Jones, 2009, p. 190), which again offers a branching or cluster process interpretation.

4.1.2 Additive-Multiplicative Spatio-Temporal Conditional Intensity Model

In the following, I propose an additive-multiplicative model for the conditional intensity function of an infectious disease (self-exciting) process continuous in space-time. The basic framework is to superimpose an endemic (immigration) and an epidemic (offspring) component just like in the Hawkes, the ETAS, or the discrete space `twinSIR` models discussed so far:

$$\lambda^*(t, \boldsymbol{s}) = h(t, \boldsymbol{s}) + e^*(t, \boldsymbol{s})\,. \tag{4.3}$$

The epidemic component $e^*(t, \boldsymbol{s})$ represents the spread of the disease by person-to-person contact, and is similar to the earthquake aftershock term in the ETAS model. The endemic component $h(t, \boldsymbol{s})$ models otherwise imported cases and is independent of the internal history of the process.

Specification of the endemic component $h(t, s)$

The endemic component is of the multiplicative Cox-type

$$h(t, \boldsymbol{s}) = \exp\left(h_0(t, \boldsymbol{s}) + \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{z}}(t, \boldsymbol{s})\right), \tag{4.4}$$

where $h_0(t, \boldsymbol{s})$ is a parametric or nonparametric spatio-temporal log-baseline intensity and the remainder is a linear predictor of endemic covariates $\tilde{\boldsymbol{z}}(t, \boldsymbol{s})$. These exogeneous covariates actually result from another jointly evolving point process, the history of which extends the internal history of the process in question (cf. Subsection 2.1.2). In the application of the IMD data, the only available endemic covariate is the number of cases of influenza (possibly with a lag of one or two weeks) on a week $\times$ county grid (cf. Section 1.2).

The log-baseline intensity is conveniently separated in its temporal and spatial dimension excluding any baseline space-time interaction:

$$h_0(t, \boldsymbol{s}) = h_0^{\text{temp}}(t) + h_0^{\text{spat}}(\boldsymbol{s}) \,.$$

A common approach in modelling spatial inhomogeneity in epidemiology is to adjust for the population at risk. Therefore, $h_0^{\text{spat}}(\boldsymbol{s})$ is modelled as an offset for the logarithmic population density in the county containing the location $\boldsymbol{s}$, such that the endemic rate of infection is proportional to the population density. The temporal log-baseline $h_0^{\text{temp}}(t)$ is modelled by a step function: this can either be a semiparametric zero-degree B-Spline or some parametric trend (e.g. linear plus sinusoidal with a yearly frequency) *measured at a discrete set of time points*. In the simplest case, $h_0^{\text{temp}}(t) = \beta_0$ is a global intercept. The reason for limiting $h_0^{\text{temp}}(t)$ to be a piecewise constant function is that this greatly simplifies the later integration of the endemic component in the log-likelihood.

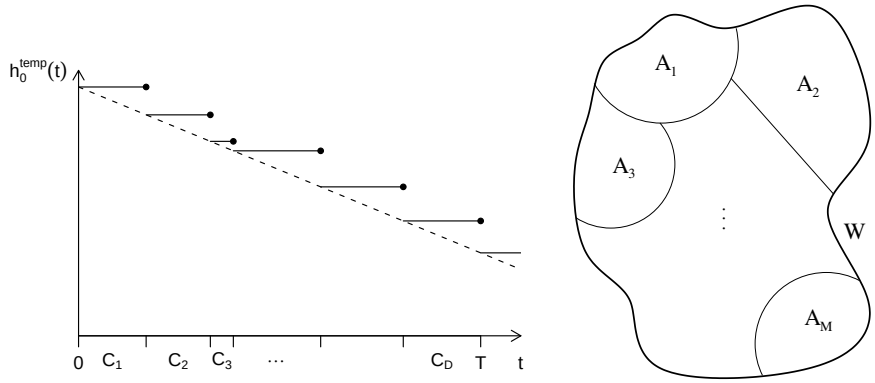
Altogether, the endemic component is modelled as a piecewise constant function on some spatio-temporal grid resulting from a decomposition of the time period $[0, T]$ and the observation region W. The consecutive time intervals of this decomposition are denoted by $C_1, \ldots, C_D \subset (0, T]$, and the spatial tiles (e.g. counties) are denoted by $A_1, \ldots, A_M \subset W$. An illustration is provided in Figure 4.2. Let the functions $\tau(t)$ and $\xi(\boldsymbol{s})$ return the indices of the temporal and spatial grid units containing time point t and coordinate $\boldsymbol{s}$, respectively. Then, the endemic component can be written as

$$h(t, \boldsymbol{s}) = \exp\left(o_{\xi(\boldsymbol{s})} + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right) , \tag{4.5}$$

where $o_{\xi(\boldsymbol{s})}$ is the tile-specific offset (e.g. the log-population density), and

$$\left\{\boldsymbol{z}_{\tau,\xi} : \tau \in \{1, \ldots, D\},\, \xi \in \{1, \ldots, M\}\right\}$$

is a collection of covariates on the spatio-temporal grid $\{C_1, \ldots, C_D\} \times \{A_1, \ldots, A_M\}$. For convenience, $h_0^{\text{temp}}(t)$ is now included in this term of the linear predictor (which is why the tilde was dropped from $\boldsymbol{\beta}$ and $\boldsymbol{z}$), and is treated as ordinary covariates.



(a) Decomposition of $(0, T]$ and (discretised) temporal log-baseline $h_0^{\text{temp}}(t)$.

(b) Decomposition of W.

Figure 4.2: Schematic illustrations of the decompositions of the (a) temporal and (b) spatial observation ranges. The dashed line in (a) represents a linear time trend of which the step function is a discrete left-continuous version over the decomposition $C_1, \ldots, C_D$.

Specification of the epidemic component $e^*(t, s)$

The self-exciting component of the model essentially provides a description of the "infection pressure" at a space-time location $(t, \boldsymbol{s})$ caused by each infectious individual. This infectivity of an infectious individual j, denoted by $e_j(t, \boldsymbol{s})$, corresponds to the inhomogeneous rate of a Poisson process, the realisations of which are the space-time locations of infected individuals. Like in almost all self-exciting spatio-temporal point processes in the literature, this "triggering function" is factorised into separate effects of marks, elapsed time, and relative location:

$$e_j(t, \boldsymbol{s}) = e^{\gamma_0 + \boldsymbol{\gamma}' \boldsymbol{m}_j} \, g(t - t_j) \, f(\boldsymbol{s} - \boldsymbol{s}_j) \,, \qquad (t > t_j) \tag{4.6}$$

where $\boldsymbol{m}_j$ is a vector of unpredictable marks attached to the infected individual, and g and f are positive, Riemann-integrable, temporal and spatial interaction functions, respectively. The linear predictor $\gamma_0 + \boldsymbol{\gamma}' \boldsymbol{m}_j$ of marks shall reflect that different individuals might cause more or less secondary cases, depending on individual characteristics. Young people or infants might trigger more secondary cases than adults, which presumably are more attentive concerning their infectivity. The individual social contact network certainly affects the risk of triggering secondary cases, although this characteristic is hardly measurable in practice. The linear predictor of marks anyway enables modelling the effect of individual-specific covariates on the degree of infectivity.

The interaction functions describe the decay of infectivity with an increasing spatial or temporal distance from the infection source. Note therefore that concerning their multiplicative decomposition in equation (4.6), it would e.g. not be reasonable to split up the components additively: the distance of previously infected individuals j to a location s would then con-

tribute to the total infection pressure at this location even when $g(t - t_j) \approx 0$, i.e. when the individual j actually is no longer infective. In contrast, the interaction of the temporal and spatial components in (4.6) arranges for zero contribution of individuals with $g(t - t_j) = 0$.

In infectious disease applications, f is usually a radially symmetric kernel corresponding to an isotropic spread of the disease, such that $f(\boldsymbol{s} - \boldsymbol{s}_j) \equiv f(\|\boldsymbol{s} - \boldsymbol{s}_j\|)$. A typical example is

$$f(\boldsymbol{s}) = \exp\left(-\frac{\|\boldsymbol{s}\|^2}{2\sigma^2}\right) \qquad (\boldsymbol{s} \in \mathbb{R}^2,\, \sigma > 0)\,, \tag{4.7}$$

i.e. the kernel of a radially symmetric bivariate normal density with zero mean. The temporal interaction function could be chosen similar to the original Hawkes model as

$$g(t) = e^{-\alpha t} \qquad (t > 0,\, \alpha > 0) \tag{4.8}$$

representing an exponential temporal decay of infectivity. If f or g are modelled as constant functions equal to 1, individuals spread the disease homogeneously in space or time, respectively. Other types of interaction functions could be taken from the families of parametric correlation functions known from time series analysis or spatial statistics, e.g. the powered exponential correlation function.

The resulting epidemic component $e^*(t, \boldsymbol{s})$ then is the sum of the contributions (4.6) of all infectious individuals at time t and location $\boldsymbol{s}$. Formally,

$$e^*(t, \boldsymbol{s}) = \int_{(0,t)\times W\times\mathcal{M}} \mathbb{1}_{(0,\varepsilon]}(t - \tilde{t})\, \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s} - \tilde{\boldsymbol{s}}\|)\, e^{\gamma_0 + \boldsymbol{\gamma}'\tilde{\boldsymbol{m}}}\, g(t - \tilde{t})\, f(\boldsymbol{s} - \tilde{\boldsymbol{s}})\, N(\mathrm{d}\tilde{t} \times \mathrm{d}\tilde{\boldsymbol{s}} \times \mathrm{d}\tilde{\boldsymbol{m}})\,,$$

where $\varepsilon, \delta > 0$ are known maximum temporal and spatial interaction ranges. In the infectious disease context, ε is an upper bound for the length of the infectious period, and δ is the maximum spatial cluster size. Assuming infinite ranges is also possible and corresponds to a never ending spread of the disease (temporally or spatially). A past event only influences the process (is infective) at time t and location $\boldsymbol{s}$, if both indicator functions are true, i.e. if it occured at most ε time units ago at a location within distance δ. These indicator functions could alternatively be included in the interaction functions f and g. However, treating ε and δ as hyperparameters suggests this convenient decomposition. Specifically, because the above stochastic integral is actually a sum, a more intuitive notation of $e^*(t, \boldsymbol{s})$ can be introduced: Let $N_g(t)$ be the ground process counting the events irrespective of their location, then define $I^*(t, \boldsymbol{s}; \varepsilon, \delta)$ as the history-dependent set that contains the indices of all past events which are still infective at time t and location $\boldsymbol{s}$:

$$I^*(t, \boldsymbol{s}; \varepsilon, \delta) := \Big\{ j \in \{1, 2, \ldots, N_g(t-)\} : \mathbb{1}_{(0,\varepsilon]}(t - t_j) = 1 \ \wedge\ \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s} - \boldsymbol{s}_j\|) = 1 \Big\}\,.$$

Using this notation, the epidemic component has the suggestive notation

$$e^*(t, \boldsymbol{s}) = \sum_{j \in I^*(t,\boldsymbol{s};\varepsilon,\delta)} e_j(t, \boldsymbol{s}) = \sum_{j \in I^*(t,\boldsymbol{s};\varepsilon,\delta)} e^{\gamma_0 + \boldsymbol{\gamma}' \boldsymbol{m}_j} \, g(t - t_j) \, f(\boldsymbol{s} - \boldsymbol{s}_j) \, . \tag{4.9}$$

For convenience, the explicit conditioning on ε and δ will be dropped from the function I^*.

Definition of `twinstim` and comparison with other models

Altogether, the proposed CIF model for a spatio-temporal point process with components (4.5) and (4.9) is

$$\lambda^*(t, \boldsymbol{s}) = \exp\left(o_{\xi(\boldsymbol{s})} + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right) + \sum_{j \in I^*(t,\boldsymbol{s})} e^{\gamma_0 + \boldsymbol{\gamma}' \boldsymbol{m}_j} \, g(t - t_j) \, f(\boldsymbol{s} - \boldsymbol{s}_j) \, , \tag{4.10}$$

and I assign to the model (4.10) the name `twinstim`. This both suggests the similarity to the additive-multiplicative `twinSIR` model of Section 2.2 as well as it indicates a *s*patio-*t*emporal (conditional) *i*ntensity *m*odel.

The main differences between `twinstim` and `twinSIR` models are:

- In the `twinSIR` model, possible event locations constitute a finite subset of $\mathbb{R}^2$, i.e. $W = \{s_1, \ldots, s_M\}$. This enables compartmental model formulations like SIR models and its extensions, because the "population" W has a fixed size beforehand. Therefore, re-infection of the same unit can in principle be modelled. This is not possible in the spatially continuous setting of `twinstim`. Here, the "population" is considered as a subset $W \subset \mathbb{R}^2$ of infinite size with respect to the counting measure on $\mathbb{R}^2$. This means that the (random) number of events theoretically has no upper bound, that events may occur at any location in W, and that a re-infection has zero probability.
- The spatial and temporal distance kernels in `twinSIR` are modelled as a linear combination of basis functions, whereas in `twinstim` this restriction is dropped and any Riemann-integrable functions with any number of parameters may be supplied.
- Last but not least, the `twinSIR` model is computationally more straightforward, because integrals over W are simple summations and numerical integration is not needed.

It should be mentioned that further similar models have been considered by Peter Diggle. In Diggle (2007, 1.3.2.2) he describes a self-exciting spatio-temporal point process resembling to the epidemic component $e^*(t, \boldsymbol{s})$ from equation (4.9). Specifically, his example is

$$\lambda^*(t, \boldsymbol{s}) = \sum_{j=1}^{N_g(t-)} g(t - t_j) \, f(\boldsymbol{s} - \boldsymbol{s}_j) \, , \tag{4.11}$$

where f is the density of the $\mathrm{N}_2(\mathbf{0}; \sigma^2 \boldsymbol{I}_2)$-distribution and g is the intensity function of an inhomogeneous Poisson process holding

$$\mu := \int_0^\infty g(t)\,\mathrm{d}t < \infty\,.$$

The CIF in (4.11) equals the epidemic component $e^*(t, \boldsymbol{s})$ except that (4.11) contains no individual-specific weighting factor and that there are no maximum interaction ranges ε and δ, which truncate f and g. Hence, Diggle's model is included in the `twinstim` class by neither specifying an endemic component (representable by $h_0(t, \boldsymbol{s}) = -\infty$) nor a weighting of infectives ($\gamma_0 = 0$), and assuming $\varepsilon = \infty = \delta$. Thus, the model basically is a Hawkes process without immigration, spatially enriched by a Thomas process. Starting from a collection of events at time point 0, every event subsequently generates offspring, all offspring subsequently generate further offspring, and so forth. The generation of offspring follows the same rules for every progenitor:

Times of offspring: According to an inhomogeneous Poisson process on $\mathbb{R}_+$ with intensity $g(u - t)$, where t is the "birth-time" of the progenitor.

Locations of offspring: Independently distributed following the isotropic bivariate normal distribution centred at the location of the progenitor with variance matrix $\sigma^2 \boldsymbol{I}_2$.

The point process described by the CIF (4.11) consists of the collection of all resulting offspring $(t_i, \boldsymbol{s}_i)$ such that $\mathcal{H}_t = \{(t_i, \boldsymbol{s}_i) : t_i \leq t\}$ and $\mathcal{H}_0$ contains the initial set of events. The temporal rule above also yields that the total number of offspring generated by an event has a Poisson distribution with mean μ. The number of events as a function of time thus forms a simple branching process, which was illustrated in Figure 4.1 for a single initial event. In the case $\mu < 1$, i.e. each event generates on average less than one offspring event over the whole time, the process is sub-critical and its eventual extinction is almost sure.

However, in a Hawkes process with immigration like in general `twinstim`'s, new parent events may always appear and initiate new outbreaks. Similar to the number μ above, an individual-specific mean number μ_j of infections caused by individual j inside its spatio-temporal range of interaction can equally be established for a `twinstim`:

$$\begin{aligned}
\mu_j &= \int_0^\infty \int_{\mathbb{R}^2} e_j(t, \boldsymbol{s})\, \mathbb{1}_{(0,\varepsilon]}(t - t_j)\, \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s} - \boldsymbol{s}_j\|)\,\mathrm{d}t\,\mathrm{d}\boldsymbol{s} \\
&= e^{\gamma_0 + \boldsymbol{\gamma}'\boldsymbol{m}_j} \cdot \left[\int_0^\infty g(t - t_j)\, \mathbb{1}_{(0,\varepsilon]}(t - t_j)\,\mathrm{d}t\right] \cdot \left[\int_{\mathbb{R}^2} f(\boldsymbol{s} - \boldsymbol{s}_j)\, \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s} - \boldsymbol{s}_j\|)\,\mathrm{d}\boldsymbol{s}\right] \\
&= e^{\gamma_0 + \boldsymbol{\gamma}'\boldsymbol{m}_j} \cdot \left[\int_0^\varepsilon g(t)\,\mathrm{d}t\right] \cdot \left[\int_{b(\mathbf{0},\delta)} f(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\right]\,. \qquad (4.12)
\end{aligned}$$

Here, $b(\mathbf{0}, \delta)$ denotes the disc centred at (0,0) with radius δ. The integration domain $\mathbb{R}_+ \times \mathbb{R}^2$ above stems from the theoretical point of view that the point process has an unlimited state

space. In practice, this is of course not the case, which is why individuals near the border would be attributed a smaller value of μ_j resulting from the integration over W instead of $\mathbb{R}^2$. Similarly, an individual which has been infected just before the end of the observation period at time T would have $\mu_j \approx 0$ if only integrating over $[0, T]$. This is however not the intended interpretation, because it is evident that this individual triggered events which have not been observed. These edge effects are thus overcome by the formulation (4.12), which also simplifies interpretation. Specifically, the number μ_j offers an intuitive way of interpreting the parameters γ_0 and $\boldsymbol{\gamma}$. The "intercept" e^{γ_0} multiplied by the two integrals is the mean number of infections caused by an infective individual whose marks $\boldsymbol{m}_j$ all equal zero. The effects of the marks can then be interpreted as usual in Poisson regression models: a unit positive change in a specific mark m_{jl} (ceteris paribus) multiplies the mean number by the corresponding parameter e^{γ_l}. An alternative interpretation of those parameters works if g and f obey a maximum value of 1 located at the origins 0 and $(0, 0)$, respectively. The epidemic intercept e^{γ_0} then equals the instantaneous rate of infection caused by an infective with $\boldsymbol{m}_j = \mathbf{0}$ just after her infection in an infinitesimal region around her position.

Another paper by Diggle et al. (2009) – pre-published online during the writing of this thesis – discusses a spatio-temporal "population growth model" with respect to partial likelihood inference. In contrast to the above model (4.11), this model also contains a constant intensity for progenitor events (immigrants) corresponding to the endemic component $h(t, \boldsymbol{s})$ from equation (4.4), and a temporally limited range of interaction. Specifically, their model is (in this thesis' notation)

$$\lambda^*(t, \boldsymbol{s}) = e^{\beta_0} + \sum_{j=1}^{N_g(t-)} e^{\gamma_0} f(\boldsymbol{s} - \boldsymbol{s}_j) \mathbb{1}_{(0,\varepsilon]}(t - t_j) , \tag{4.13}$$

and is thus also covered by the `twinstim` class. In their simulation example they used a radially symmetric Gaussian density for f, but they only considered rectangular observation regions over which the integration of f is straightforward (as the product of univariate normal random variates). The model exemplifies that some parameters may not be identifiable from the partial likelihood, which here is the case for β_0 and γ_0: only the ratio e^{β_0}/e^{γ_0} is identifiable (cf. Diggle et al., 2009, Section 3.2).

4.1.3 Type-Specific `twinstim`

The IMD data actually represent a spatio-temporal point pattern marked by the finetype of infection. Although the model of the previous subsection already allows for a finetype-specific weight of infectivity through the vector of marks $\boldsymbol{m}_j$, it is not applicable for a joint modelling of both finetypes. The main issue is that the finetypes do not cause mutual infections but only

the same finetype is transmitted (cf. Subsection 1.1). Both finetypes should however have the same relation to the population at risk and, as I assume for simplicity, to the case numbers of influenza and the time trend. The only finetype-specific element in the endemic component would be the intercept, corresponding to the global background rate. However, the most important distinction concerns their epidemic behaviour represented by the epidemic weight and the two interaction functions. For instance, the serogroup B finetype would have a higher epidemic weight and a relatively narrow spatial interaction function corresponding to more intense but rather local clustering. Taking everything into account, the model of the previous subsection will now be extended to a marked version, which enables the joint modelling of both finetypes.

Denote by $\mathcal{K} = \{1, \dots, K\} \subset \mathbb{N}$ the set of possible event types. Define an indicator matrix

$$\boldsymbol{Q} = (q_{k,l})_{k,l \in \mathcal{K}} \qquad q_{k,l} \in \{0; 1\}$$

which determines the possible ways of transmission. If $q_{k,l}$ equals 1, an infective type k event can cause an event of type l. For instance, the IMD data would require $\boldsymbol{Q} = \boldsymbol{I}_2$, because the transmission is finetype-specific. A marked spatio-temporal point process on $(0, T] \times W \times \mathcal{K}$ is then defined by the following model for the conditional intensity function:

$$\begin{aligned}
\lambda^*(t, \boldsymbol{s}, \kappa) &= h(t, \boldsymbol{s}, \kappa) + e^*(t, \boldsymbol{s}, \kappa) \qquad (4.14)\\
h(t, \boldsymbol{s}, \kappa) &= \exp\left(h_0^{\text{type}}(\kappa) + o_{\xi(\boldsymbol{s})} + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right)\\
e^*(t, \boldsymbol{s}, \kappa) &= \sum_{j \in I^*(t,\boldsymbol{s},\kappa)} e_j(t, \boldsymbol{s})\\
e_j(t, \boldsymbol{s}) &= \exp(\gamma_{0,\kappa_j} + \tilde{\boldsymbol{\gamma}}' \tilde{\boldsymbol{m}}_j) \cdot g(t - t_j | \kappa_j) \cdot f(\boldsymbol{s} - \boldsymbol{s}_j | \kappa_j)\\
I^*(t, \boldsymbol{s}, \kappa) &= \left\{ j \in \{1, \dots, N_g(t-)\} : \ \mathbb{1}_{(0,\varepsilon]}(t - t_j) = 1 \ \wedge \ \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s} - \boldsymbol{s}_j\|) = 1 \ \wedge \ q_{\kappa_j,\kappa} = 1 \right\}
\end{aligned}$$

Here, the transmission indicators from the matrix $\boldsymbol{Q}$ have been integrated into $I^*(t, \boldsymbol{s}, \kappa)$. Note that additional marks $\tilde{\boldsymbol{m}}_j$ of the events are still not being modelled but are part of the history of the process (see the discussion around equation (2.7)). The new endemic baseline component $h_0^{\text{type}}(\kappa)$ either represents a type-specific endemic intercept, i.e.

$$h_0^{\text{type}}(\kappa) = \sum_{k=1}^{K} \beta_{0,k} \, \mathbb{1}_{\{k=\kappa\}}(\kappa) = \beta_{0,\kappa} \, ,$$

or contains only a single global intercept $h_0^{\text{type}}(\kappa) = \beta_0$, corresponding to the hypothesis $\beta_0 = \beta_{0,1} = \cdots = \beta_{0,K}$. In any case, the parameter vector $\boldsymbol{\beta}$ of the third term in the linear predictor of $h(t, \boldsymbol{s}, \kappa)$ should of course no longer contain an intercept. The model further assumes that this term does not depend on κ, which means that the effect of endemic covariates (case numbers of influenza as well as the time trend) is homogeneous over the event types.

In the epidemic component, γ_{0,κ_j} denotes a type-specific intercept. As already mentioned, this is no real extension of the previous model, where the event type could already be treated as an unpredictable mark. However, the history-dependent set $I^*(t, \boldsymbol{s}, \kappa)$ of infective individuals now accounts for the transmission regime $\boldsymbol{Q}$ between the event types, and the interaction functions are allowed to depend on the type of the infective event. For instance,

$$f(\boldsymbol{s}|\kappa) = \exp\left(-\frac{\|\boldsymbol{s}\|^2}{2\sigma_\kappa^2}\right) \qquad (\boldsymbol{s} \in \mathbb{R}^2,\ \sigma_\kappa > 0) \tag{4.15}$$

models a type-specific decay of infectivity with increasing distance from the source of infection. A type-specific exponential decay of infectivity as time goes by is similarly possible for the temporal interaction function g:

$$g(t|\kappa) = e^{-\alpha_\kappa t} \qquad (t > 0,\ \alpha_\kappa > 0,\ \kappa \in \mathcal{K})\,. \tag{4.16}$$

4.2 Inference

The unmarked `twinstim` of Subsection 4.1.2 can be treated as a special case of the marked version of Subsection 4.1.3 by assuming $\mathcal{K} = \{1\}$. Maximum likelihood estimation will thus be presented only for the marked `twinstim`, i.e. we are given a parametric model of the CIF introduced in (4.14):

$$\begin{aligned}\lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) = &\exp\left(h_0^{\text{type}}(\kappa) + o_{\xi(\boldsymbol{s})} + \boldsymbol{\beta}'\boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right) \\ &+ \sum_{j \in I^*(t,\boldsymbol{s},\kappa)} \exp(\gamma_{0,\kappa_j} + \tilde{\boldsymbol{\gamma}}'\tilde{\boldsymbol{m}}_j) \cdot g_{\boldsymbol{\alpha}}(t - t_j|\kappa_j) \cdot f_{\boldsymbol{\sigma}}(\boldsymbol{s} - \boldsymbol{s}_j|\kappa_j)\,,\end{aligned} \tag{4.17}$$

where $h_0^{\text{type}}(\kappa) = \beta_{0,\kappa}$ is the type-specific endemic intercept being equal to β_0 in case of a type-invariant endemic model. For convenience, the parameters $\boldsymbol{\gamma_0} := (\gamma_{0,1}, \ldots, \gamma_{0,K})'$ and $\tilde{\boldsymbol{\gamma}}$ of the epidemic linear predictor will be treated together denoting $\boldsymbol{\gamma} := (\boldsymbol{\gamma_0}', \tilde{\boldsymbol{\gamma}}')'$. Dummy variables for the event types are included in the mark vector $\tilde{\boldsymbol{m}}_j$. For instance, if event j is of type κ_j, then the mark vector of this event is $\boldsymbol{m}_j := (0, \ldots, 0, 1, 0, \ldots, \tilde{\boldsymbol{m}}_j')'$, where the 1 appears at position κ_j. Hence, the epidemic linear predictor is $\boldsymbol{\gamma}'\boldsymbol{m}_j$, and the parameter vector in question is

$$\boldsymbol{\theta} = (\boldsymbol{\beta_0}', \boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\sigma}', \boldsymbol{\alpha}')'\,, \tag{4.18}$$

where $\boldsymbol{\beta_0} = (\beta_{0,1}, \ldots, \beta_{0,K})'$ (type-specific) or $\boldsymbol{\beta_0} = \beta_0$ (type-invariant), and $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}$ are the parameter vectors of the spatial and temporal interaction functions $f_{\boldsymbol{\sigma}}$ and $g_{\boldsymbol{\alpha}}$, respectively.

Parameter estimates can be obtained by maximisation of the log-likelihood (2.10) or the partial log-likelihood (2.6) with respect to $\boldsymbol{\theta}$. Trading the partial likelihood off against the

full likelihood, the results of the simulation study in Diggle et al. (2009) support satisfactory relative efficiency of the partial likelihood. However, I do not see great benefit in using the partial likelihood approach because the need for spatial integration – which is the computational bottleneck of routine statistical inference – remains. Furthermore, as already mentioned, the parameters e^{β_0} and e^{γ_0} would not both be identifiable, but only their ratio. As a consequence of these defaults – although I have implemented the partial likelihood alternative (see Appendix A) – I will concentrate on full maximum likelihood estimation.

Subsections 4.2.1 and 4.2.2 give in turn the log-likelihood and score functions related to the type-specific `twinstim`. An estimation of the expected Fisher information matrix is given in Subsection 4.2.3. Here, I also discuss asymptotic properties of the maximum likelihood estimators, and touch possibilities of model selection. See Fahrmeir, Kneib & Lang (2007, Appendix B.4) for a compendium of likelihood inference.

Note that for the complex (marked) `twinstim` class, analytical maximisation of the log-likelihood is not feasible, and a numerical optimisation routine will be required. This will be further discussed in Section 4.3.

4.2.1 Log-Likelihood Function

According to equation (2.10), the log-likelihood decomposes in a sum over the observed conditional log-intensities $\log \lambda^*_{\boldsymbol{\theta}}(t_i, \boldsymbol{s}_i, \kappa_i)$, $i = 1, \ldots, N$, and a time-space-mark integral (in the general sense of Lebesgue-Stieltjes). The components of the sum can be directly calculated for a specific value of the parameter vector $\boldsymbol{\theta}$ after having determined the set $I^*(t_i, \boldsymbol{s}_i, \kappa_i)$ of potential sources of infection for the ith event. Furthermore, the integrated conditional intensity function in the log-likelihood is

$$\int_0^T \int_W \sum_{\kappa \in \mathcal{K}} \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s} = \int_0^T \int_W \sum_{\kappa \in \mathcal{K}} h_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s} + \int_0^T \int_W \sum_{\kappa \in \mathcal{K}} e^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s}$$

such that the integrations of the endemic and the epidemic component can be performed separately.

Recalling that the endemic component is a piecewise constant function on the spatio-temporal grid $\{C_1, \ldots, C_D\} \times \{A_1, \ldots, A_M\}$, the first integral is in fact a sum over this grid of smallest observed units in space-time:

$$\int_0^T \int_W \sum_{\kappa \in \mathcal{K}} h_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s} = \left(\sum_{\kappa \in \mathcal{K}} \exp\left(h_0^{\mathrm{type}}(\kappa) \right) \right) \cdot \sum_{\tau=1}^{D} \sum_{\xi=1}^{M} |C_\tau||A_\xi| \exp\left(o_\xi + \boldsymbol{\beta}' \boldsymbol{z}_{\tau,\xi} \right) \quad (4.19)$$

The integrated epidemic component can be simplified – moving the indicators of the function $I^*(t, \boldsymbol{s}, \kappa)$ back into the sum – as follows:

$$\begin{aligned}
&\int_0^T \int_W \sum_{\kappa\in\mathcal{K}} \sum_{j=1}^{N} \mathbb{1}_{(0,\varepsilon]}(t-t_j)\, \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s}-\boldsymbol{s}_j\|)\, q_{\kappa_j,\kappa}\, e^{\gamma' \boldsymbol{m}_j}\, g_{\boldsymbol{\alpha}}(t-t_j|\kappa_j)\, f_{\boldsymbol{\sigma}}(\boldsymbol{s}-\boldsymbol{s}_j|\kappa_j)\, \mathrm{d}t\, \mathrm{d}\boldsymbol{s} \\
&= \sum_{j=1}^{N} q_{\kappa_j,\bullet}\, e^{\gamma' \boldsymbol{m}_j} \left[\int_0^T \mathbb{1}_{(0,\varepsilon]}(t-t_j) g_{\boldsymbol{\alpha}}(t-t_j|\kappa_j)\, \mathrm{d}t \right] \left[\int_W \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s}-\boldsymbol{s}_j\|) f_{\boldsymbol{\sigma}}(\boldsymbol{s}-\boldsymbol{s}_j|\kappa_j)\, \mathrm{d}\boldsymbol{s} \right] \\
&= \sum_{j=1}^{N} q_{\kappa_j,\bullet}\, e^{\gamma' \boldsymbol{m}_j} \left[\int_0^{\min\{T-t_j;\varepsilon\}} g_{\boldsymbol{\alpha}}(t|\kappa_j)\, \mathrm{d}t \right] \left[\int_{R_j} f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa_j)\, \mathrm{d}\boldsymbol{s} \right] . \qquad (4.20)
\end{aligned}$$

Here,

$$q_{\kappa_j,\bullet} := \left(\sum_{\kappa\in\mathcal{K}} q_{\kappa_j,\kappa} \right)$$

is the number of different event types which can be triggered by an event of type κ_j, and

$$R_j := [W \cap b(\boldsymbol{s}_j;\delta)] - \boldsymbol{s}_j \qquad (4.21)$$

is the spatial interaction region centred at the location of the jth event, where $b(\boldsymbol{s}_j;\delta)$ denotes the disc centred at $\boldsymbol{s}_j$ with radius δ. Recall that such a type of region R_j was already seen in Figure 3.9(b). In the case of unlimited spatial transmission ($\delta = \infty$), $R_j = W - \boldsymbol{s}_j$ equals the translation of the whole observation region by $\boldsymbol{s}_j$ such that $\boldsymbol{s}_j$ becomes the origin.

The evaluation of the two-dimensional integral over the domains R_j is the most sophisticated task of the log-likelihood evaluation. Following the results from Chapter 3, the integral is computed by the two-dimensional midpoint rule. For the special case of the type-specific Gaussian kernel (4.15), robust accuracy for any value of σ_κ is guaranteed by an adaptive choice of the bandwidth $h = \phi\,\sigma_\kappa$ (see Subsection 3.2.7).

In contrast, the evaluation of the definite integral over the temporal interaction function is analytically accessible for the simple choices of $g_{\boldsymbol{\alpha}}$ assumed in this thesis. Provided $G_{\boldsymbol{\alpha}}(t|\kappa)$ denotes an antiderivative of $g_{\boldsymbol{\alpha}}(t|\kappa)$, the first integral equals

$$\int_0^{\min\{T-t_j;\varepsilon\}} g_{\boldsymbol{\alpha}}(t|\kappa_j)\, \mathrm{d}t = G_{\boldsymbol{\alpha}}\big(\min\{T-t_j;\varepsilon\}|\kappa_j\big) - G_{\boldsymbol{\alpha}}(0|\kappa_j) .$$

For instance, the type-specific exponential decay function $g_{\boldsymbol{\alpha}}$ from equation (4.16) has antiderivative

$$G_{\boldsymbol{\alpha}}(t|\kappa) = -\frac{e^{-\alpha_\kappa t}}{\alpha_\kappa} \qquad (\alpha_\kappa > 0) . \qquad (4.22)$$

The case $\alpha_\kappa = 0$, would correspond to a time-invariant infectivity, i.e. $g_{\boldsymbol{\alpha}}(t) = 1$ with antiderivative $G_{\boldsymbol{\alpha}}(t|\kappa) = t$.

4.2.2 Score Function

Being able to calculate the log-likelihood for a specific set of parameters $\boldsymbol{\theta}$ is the basic prerequisite for numeric maximum likelihood inference. For numerical maximisation of the log-likelihood it is moreover advantageous to have its derivative with respect to $\boldsymbol{\theta}$, i.e. the score function. In what follows, analytical derivations of the score function are given.

Let $\boldsymbol{\vartheta}$ denote any subvector of $\boldsymbol{\theta}$. Then, the partial derivative of the log-likelihood with respect to $\boldsymbol{\vartheta}$ is

$$\boldsymbol{s}_{\boldsymbol{\vartheta}}(\boldsymbol{\theta}) := \frac{\partial}{\partial \boldsymbol{\vartheta}} l(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{\frac{\partial}{\partial \boldsymbol{\vartheta}} \lambda^*_{\boldsymbol{\theta}}(t_i, \boldsymbol{s}_i, \kappa_i)}{\lambda^*_{\boldsymbol{\theta}}(t_i, \boldsymbol{s}_i, \kappa_i)} - \int_0^T \int_W \sum_{\kappa \in \mathcal{K}} \frac{\partial}{\partial \boldsymbol{\vartheta}} \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s} \,, \tag{4.23}$$

and the score function is

$$\boldsymbol{s}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{s}_{\boldsymbol{\beta_0}} \\ \boldsymbol{s}_{\boldsymbol{\beta}} \\ \boldsymbol{s}_{\boldsymbol{\gamma}} \\ \boldsymbol{s}_{\boldsymbol{\sigma}} \\ \boldsymbol{s}_{\boldsymbol{\alpha}} \end{pmatrix} (\boldsymbol{\theta}) \,.$$

The necessary partial derivatives of the CIF with their respective time-space-mark integrals are given in the subsequent paragraphs, and can then be plugged into the equation (4.23).

Endemic intercept(s) $\boldsymbol{\beta_0}$:

Let $\beta_{0,k}$, $k \in \{1, \ldots, K\}$ be one of the type-specific intercepts in $\boldsymbol{\beta_0}$. Then,

$$\frac{\partial}{\partial \beta_{0,k}} \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) = \begin{cases} \exp\left(\beta_{0,k} + o_{\xi(\boldsymbol{s})} + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right) & \text{if } \kappa = k, \\ 0 & \text{otherwise,} \end{cases}$$

since the parameter $\beta_{0,k}$ only appears in the endemic component $h_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa)$ and only if $\kappa = k$. The corresponding integrated value is

$$\int_0^T \int_W \sum_{\kappa \in \mathcal{K}} \frac{\partial}{\partial \beta_{0,k}} \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s} = e^{\beta_{0,k}} \cdot \sum_{\tau=1}^{D} \sum_{\xi=1}^{M} |C_\tau| |A_\xi| \exp\left(o_\xi + \boldsymbol{\beta}' \boldsymbol{z}_{\tau,\xi}\right) \,,$$

cf. the integral of the endemic component in equation (4.19). If the model assumes a type-invariant endemic intercept $\boldsymbol{\beta_0} = \beta_0$, then

$$\frac{\partial}{\partial \beta_0} \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) = \exp\left(\beta_0 + o_{\xi(\boldsymbol{s})} + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right)$$

with integrated value

$$\int_0^T \int_W \sum_{\kappa\in\mathcal{K}} \frac{\partial}{\partial\beta_0}\lambda^*_{\boldsymbol{\theta}}(t,\boldsymbol{s},\kappa)\,\mathrm{d}t\,\mathrm{d}\boldsymbol{s} = K\,e^{\beta_0}\cdot\sum_{\tau=1}^{D}\sum_{\xi=1}^{M}|C_\tau||A_\xi|\exp\left(o_\xi+\boldsymbol{\beta}'\boldsymbol{z}_{\tau,\xi}\right)\ .$$

Endemic covariate effects $\boldsymbol{\beta}$:

$$\frac{\partial}{\partial\boldsymbol{\beta}}\lambda^*_{\boldsymbol{\theta}}(t,\boldsymbol{s},\kappa) = \exp\left(h_0^{\text{type}}(\kappa)+o_{\xi(\boldsymbol{s})}+\boldsymbol{\beta}'\boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right)\cdot\boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}$$

with corresponding integral vector (element-wise integral values)

$$\left(\sum_{\kappa\in\mathcal{K}}\exp\left(h_0^{\text{type}}(\kappa)\right)\right)\cdot\sum_{\tau=1}^{D}\sum_{\xi=1}^{M}|C_\tau||A_\xi|\exp(o_\xi+\boldsymbol{\beta}'\boldsymbol{z}_{\tau,\xi})\,\boldsymbol{z}_{\tau,\xi}\ .$$

Epidemic effects $\boldsymbol{\gamma}=(\boldsymbol{\gamma}_0',\tilde{\boldsymbol{\gamma}}')'$:

$$\frac{\partial}{\partial\boldsymbol{\gamma}}\lambda^*_{\boldsymbol{\theta}}(t,\boldsymbol{s},\kappa) = \sum_{j\in I^*(t,\boldsymbol{s},\kappa)} e^{\boldsymbol{\gamma}'\boldsymbol{m}_j}\,g_{\boldsymbol{\alpha}}(t-t_j|\kappa_j)\,f_{\boldsymbol{\sigma}}(\boldsymbol{s}-\boldsymbol{s}_j|\kappa_j)\,\boldsymbol{m}_j\ ,$$

and the corresponding integral can be deduced similar to equation (4.20) as

$$\sum_{j=1}^{N} q_{\kappa_j,\boldsymbol{\cdot}}\ e^{\boldsymbol{\gamma}'\boldsymbol{m}_j}\boldsymbol{m}_j\left[\int_0^{\min\{T-t_j;\varepsilon\}} g_{\boldsymbol{\alpha}}(t|\kappa_j)\,\mathrm{d}t\right]\left[\int_{R_j} f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa_j)\,\mathrm{d}\boldsymbol{s}\right].$$

Parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}$ of the interaction functions:

For a general spatial kernel $f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa)$,

$$\frac{\partial}{\partial\boldsymbol{\sigma}}\lambda^*_{\boldsymbol{\theta}}(t,\boldsymbol{s},\kappa) = \sum_{j\in I^*(t,\boldsymbol{s},\kappa)} e^{\boldsymbol{\gamma}'\boldsymbol{m}_j}\,g_{\boldsymbol{\alpha}}(t-t_j|\kappa_j)\,\left[\frac{\partial}{\partial\boldsymbol{\sigma}}f_{\boldsymbol{\sigma}}(\boldsymbol{s}-\boldsymbol{s}_j|\kappa_j)\right]$$

with corresponding integral

$$\sum_{j=1}^{N} q_{\kappa_j,\boldsymbol{\cdot}}\ e^{\boldsymbol{\gamma}'\boldsymbol{m}_j}\left[\int_0^{\min\{T-t_j;\varepsilon\}} g_{\boldsymbol{\alpha}}(t|\kappa_j)\,\mathrm{d}t\right]\left[\int_{R_j}\frac{\partial}{\partial\boldsymbol{\sigma}}f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa_j)\,\mathrm{d}\boldsymbol{s}\right].$$

Similarly, for a general temporal kernel $g_{\boldsymbol{\alpha}}(t|\kappa)$,

$$\frac{\partial}{\partial\boldsymbol{\alpha}}\lambda^*_{\boldsymbol{\theta}}(t,\boldsymbol{s},\kappa) = \sum_{j\in I^*(t,\boldsymbol{s},\kappa)} e^{\boldsymbol{\gamma}'\boldsymbol{m}_j}\,\left[\frac{\partial}{\partial\boldsymbol{\alpha}}g_{\boldsymbol{\alpha}}(t-t_j|\kappa_j)\right]\,f_{\boldsymbol{\sigma}}(\boldsymbol{s}-\boldsymbol{s}_j|\kappa_j)$$

with corresponding integral

$$\sum_{j=1}^{N} q_{\kappa_j,\bullet}\, e^{\gamma' \boldsymbol{m}_j} \Big[\int_0^{\min\{T-t_j;\varepsilon\}} \frac{\partial}{\partial \boldsymbol{\alpha}} g_{\boldsymbol{\alpha}}(t|\kappa_j)\,\mathrm{d}t \Big] \Big[\int_{R_j} f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa_j)\,\mathrm{d}\boldsymbol{s} \Big] .$$

The analytic derivatives of f and g with respect to $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}$, respectively, have to be determined for the specific model at hand and plugged into the above formulae. For instance, a type-specific Gaussian kernel with $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_K)'$ as in equation (4.15) has partial derivatives

$$\frac{\partial}{\partial \sigma_k} f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa) = \begin{cases} \exp\left(-\frac{\|\boldsymbol{s}\|^2}{2\sigma_k^2}\right) \frac{\|\boldsymbol{s}\|^2}{\sigma_k^3} & \text{if } k = \kappa, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for any } k \in \mathcal{K} . \tag{4.24}$$

The type-specific temporal exponential decay function with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ from equation (4.16) has partial derivatives

$$\frac{\partial}{\partial \alpha_k} g_{\boldsymbol{\alpha}}(t|\kappa) = \begin{cases} -t\, e^{-\alpha_k t} & \text{if } k = \kappa, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for any } k \in \mathcal{K} . \tag{4.25}$$

While the integral of $\frac{\partial}{\partial \sigma_\kappa} f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa)$ over the region R_j will be approximated by numerical integration, the temporal function $\frac{\partial}{\partial \alpha_\kappa} g_{\boldsymbol{\alpha}}(t|\kappa)$ is assumed to permit analytical integration. Specifically for (4.25), using integration by parts,

$$\int -t\, e^{-\alpha_\kappa t}\,\mathrm{d}t = \left(t + \frac{1}{\alpha_\kappa}\right) \frac{e^{-\alpha_\kappa t}}{\alpha_\kappa} . \tag{4.26}$$

4.2.3 Fisher Information Matrix, Uncertainty, Model Selection

The inverse of the Fisher information matrix (observed or expected) at the maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}_{ML}$ is in general likelihood theory used as an estimate of the variance matrix of $\hat{\boldsymbol{\theta}}_{ML}$. This procedure is well known from classical i.i.d. statistics under Fisher regularity conditions, where the maximum likelihood estimator is consistent, asymptotically efficient, and asymptotically normal. The precise conditions under which asymptotic properties of maximum likelihood estimators hold for spatio-temporal point processes have been established by Rathbun (1996). Specifically, the conditions for existence, consistence and asymptotic normality of a local maximum $\hat{\boldsymbol{\theta}}_{ML}$ as $T \to \infty$ for a fixed observation region W are, roughly spoken:

- Regularity conditions on the CIF, which are "satisfied by all reasonable point proces models" (Rathbun, 1996, p. 62): almost sure positivity almost everywhere, $\mathcal{H}_t$-predictability (which follows from left-continuity), continuity in $\boldsymbol{\theta}$, and almost sure

integrability over $[0, T] \times W$, for $0 < T < \infty$. Note that positivity of the CIF (4.17) is guaranteed without imposing constraints on $\boldsymbol{\theta}$ by only requiring the functions $g_{\boldsymbol{\alpha}}$ and $f_{\boldsymbol{\sigma}}$ to be nonnegative.

- Existence and continuity in $\boldsymbol{\theta}$ of the first and second partial derivatives of the CIF with respect to any parameter. This ensures that the log-likelihood is twice differentiable with respect to $\boldsymbol{\theta}$. The last subsection actually verifies differentiability, and I also wrote down the analytical Hessian.
- A further spatial integrability condition involving the partial derivatives (see Rathbun, 1996, p. 61, condition (A3)).
- Two sophisticated conditions that give appropriate growth, convergence, and continuity of the Hessian matrix as $T \to \infty$. This roughly means that "information" increases and converges with observation length (and thus with the number of observed events).

For the special case of inhomogeneous Poisson point processes in either time or space the conditions are given in Rathbun & Cressie (1994). Since the last two conditions above are very difficult to verify for the (marked) `twinstim` class, I assume them to be satisfied – also for the marked version. This assumption is supported by the provided example of a self-exciting point process in Rathbun (1996, Section 5), where the conditions have been shown to be fulfilled. With $p = 1$ and $\gamma(\boldsymbol{s}) = |W|^{-1}$ therein, his example equals the unmarked `twinstim`

$$\lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}) = e^{\beta_0} + \sum_{j \in I^*(t,\boldsymbol{s})} e^{\gamma_0} f_\sigma(\boldsymbol{s} - \boldsymbol{s}) \, g_\alpha(t - t_j)$$

with $\varepsilon = \delta = \infty$ and the typical choices (4.8) and (4.7) for f_σ and g_α, respectively.

"Alternatively, simulations [e.g. parametric bootstrap] may be useful for obtaining approximate standard errors and for other types of inference" (Schoenberg et al., 2002, p. 7). However, the implementation of a simulation algorithm for the `twinstim` class is beyond the scope of this Master's Thesis. Hence, estimation of the uncertainty of $\hat{\boldsymbol{\theta}}_{ML}$ is restricted to the asymptotic properties based on the Fisher information matrix.

The first possibility to compute the *observed* Fisher information is to analytically determine and implement the negative Hessian of the log-likelihood function. In the face of the complex form of the Hessian for a general `twinstim`, I did not in further detail consider its implementation because there are simpler alternatives. On the one hand, the optimisation procedure described in 4.3 returns a numerical approximation of the Hessian matrix at the MLE. On the other hand, the *expected* Fisher information $\mathcal{I}(\boldsymbol{\theta})$ can be estimated by the "optional variation process" – adapted from Martinussen & Scheike (2002, p. 64) to the marked spatio-temporal setting –

$$\int_0^T \int_W \int_{\mathcal{K}} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) \right)^{\otimes 2} \mathrm{d}N(t, \boldsymbol{s}, \kappa)$$

through its observed realisation

$$\hat{\mathcal{I}}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \left(\frac{\partial}{\partial \tilde{\boldsymbol{\theta}}} \log \lambda^*_{\tilde{\boldsymbol{\theta}}}(t_i, \boldsymbol{s}_i, \kappa_i) \Big|_{\tilde{\boldsymbol{\theta}}=\boldsymbol{\theta}} \right)^{\otimes 2} = \sum_{i=1}^{N} \left(\frac{\frac{\partial}{\partial \tilde{\boldsymbol{\theta}}} \lambda^*_{\tilde{\boldsymbol{\theta}}}(t_i, \boldsymbol{s}_i, \kappa_i)}{\lambda^*_{\tilde{\boldsymbol{\theta}}}(t_i, \boldsymbol{s}_i, \kappa_i)} \Big|_{\tilde{\boldsymbol{\theta}}=\boldsymbol{\theta}} \right)^{\otimes 2}, \tag{4.27}$$

where $\boldsymbol{a}^{\otimes 2} := \boldsymbol{a}\boldsymbol{a}'$ for a vector $\boldsymbol{a}$. The same estimator was also recommended and used in Rathbun (1996, equation (4.7), note that here the inverse is given). Given the implementation of the score function, the calculation of $\hat{\mathcal{I}}(\hat{\boldsymbol{\theta}}_{ML})$ poses no difficulties because a similar quantity already appeared in equation (4.23). In the application, the numerical approximation of the *observed* Fisher information and the estimation of the *expected* Fisher information at $\hat{\boldsymbol{\theta}}_{ML}$ were sufficiently close to deduce that the approach and my implementation are valid. Only if "information" is very limited, e.g. for badly identified epidemic parameters in case of few observed person-to-person infections, the numerical approximation could either fail or result in a (numerically) singular matrix, which can not be inverted. Fortunately, this is not the case for $\hat{\mathcal{I}}(\hat{\boldsymbol{\theta}}_{ML})$.

Uncertainty of the parameter estimates is thus deduced from the diagonal of $\hat{\mathcal{I}}^{-1/2}(\hat{\boldsymbol{\theta}}_{ML})$, which contains the standard errors of the parameters. Considering the asymptotic normality of $\hat{\boldsymbol{\theta}}_{ML}$, ordinary component-wise Wald confidence intervals of the type (3.6) can be constructed. In principle, of course, likelihood-based confidence intervals dominate Wald confidence intervals, but they require heavy computations of likelihood profiles and demand numeric root-finding. Such computations were not further investigated in this Master's Thesis, because the Wald approach was regarded as the only feasible approach with respect to the computational cost.

However, likelihood ratio tests and AIC procedures "apply to point processes as to any other class of stochastic models" (Daley & Vere-Jones, 2008, p. 502) and are main ingredients of model selection. The former will e.g. be used to test whether the null hypothesis of a type-invariant endemic intercept $\beta_0 = \beta_{0,1} = \cdots = \beta_{0,K}$ can be rejected. Generally, if the linear null hypothesis $H_0 : \boldsymbol{C}\boldsymbol{\theta} = \boldsymbol{d}$ is tested against $H_1 : \boldsymbol{C}\boldsymbol{\theta} \neq \boldsymbol{d}$, where the $(r \times p)$-matrix $\boldsymbol{C}$ has $\operatorname{rank}(\boldsymbol{C}) = r \leq p$, and $p := \dim(\boldsymbol{\theta})$ is the number of parameters, then the (log-)likelihood ratio test statistic is

$$lr := 2 \log \left(\frac{L_1(\hat{\boldsymbol{\theta}}_1)}{L_0(\hat{\boldsymbol{\theta}}_0)} \right) = 2 \left(l_1(\hat{\boldsymbol{\theta}}_1) - l_0(\hat{\boldsymbol{\theta}}_0) \right) \overset{a}{\sim} \chi^2(r) \,.$$

Here, $L_1(\hat{\boldsymbol{\theta}}_1)$ denotes the maximum likelihood value of the unrestricted model, and $L_0(\hat{\boldsymbol{\theta}}_0)$ denotes the maximum likelihood value of the restricted model under the null hypothesis H_0. If the former is "sufficiently" larger than the latter, then H_0 will be rejected in favour of H_1. The meaning of "sufficiently" depends on the number r of hypotheses and the confidence level, and is determined by the asymptotic $\chi^2(r)$-distribution of the test statistic.

The AIC (Akaike's information criterion) is defined as

$$AIC := -2\,l(\hat{\boldsymbol{\theta}}_{ML}) + 2\,p\,,$$

and represents a trade-off between goodness of fit and model complexity. Models exhibiting smaller AIC values are generally more favourable. It is especially helpful for the comparison of non-nested models and could also be used to examine hyperparameters like ε and δ by "brute force" trying a range of values, although uncertainty of these parameters is not assessable.

4.3 Implementation

The implementation of all necessary computations for fitting models of the (marked) `twinstim` class in `R` made up a considerable amount of time during my Master's Thesis period. All the code is available from the CD attached as Appendix A. In this section, I will give an overview of the usage of the two main `R` functions. Further details are found in the code on the CD. It is worth pointing out that the provided implementation is not limited to the analysis of the IMD data or infectious diseases, but is applicable in the quite general context of spatio-temporal self-exciting point processes. It could therefore serve as a basis for future research and other applications.

The first step towards fitting a model of the `twinstim` class is to convert the data about occurrences of events and covariate information in space-time into a standardized structure. The available data usually consists of a dataset, say `events`, where each row represents an observed event. More precisely, `events` is required to be a `SpatialPointsDataFrame` (defined by the package `sp`). Besides the location and the time point, there may also be a number of marks attached to the events. Furthermore, it must contain a specification of the assumed maximum temporal and spatial ranges of interaction ε and δ. In order to fit a `twinstim` with an inhomogeneous endemic component to the observed point pattern, a second dataset (`data.frame`), say `stgrid`, provides endemic covariate information on some spatio-temporal decomposition of the observation region as described in the text of Figure 4.2. These two objects are the main ingredients of an object of the `S3` class `"epidata"`. Such an object can be constructed by the call

```
R> epi <- as.epidata(events, stgrid, W, qmatrix, nCircle2Poly)
```

where additionally, `W` is the observation region of the point pattern (inheriting from the `sp` class `SpatialPolygons`), `qmatrix` is the $|\mathcal{K}| \times |\mathcal{K}|$ indicator matrix $\boldsymbol{Q}$ in case of a multitype epidemic, and `nCircle2Poly` determines the number of edges for the polygonal approximation to a circle. For the IMD data, `W` represents Germany's borders, `qmatrix=diag(2)` because transmission is finetype-specific, and I always used `nCircle2Poly=32` in the analyses of the

next chapter. The converter function `as.epidata` checks the consistency of the supplied data and prepares everything for the model estimation. For example it computes the individual influence regions R_j defined in equation (4.21). After coercion of the present IMD data into an object of the class `"epidata"`, this object prints in `R` as shown in Figure 4.3.

```
R> imd
History of an epidemic
Observation period: 0 -- 2569
Observation window (bounding box): [4031.295, 4672.253] x [2684.102, 3551.343]
Spatio-temporal grid (not shown): 367 time blocks, 413 tiles
Types of events: 'B' 'C'
Overall number of events: 636

          coordinates ID  time  tile type eps.t eps.s pnr age    sex BLOCK
12 (4112.19, 3202.79)  1  3.99 05554    B    30   150 103  17   male     1
50 (4122.51, 3076.97)  2  4.00 05382    C    30   150 402   3   male     1
39 (4412.47, 2915.94)  3  9.00 09574    B    30   150 312  34 female     2
40  (4202.64, 2879.7)  4 11.00 08212    B    30   150 314  15 female     2
78 (4128.33, 3223.31)  5 26.00 05554    C    30   150 629  15   male     4
9     (4089.92, 3178)  6 28.00 05170    C    30   150  86  16   male     4
   start popdensity influenza0 influenza1 influenza2 influenza3
12     0   260.8612          0          0          0          0
50     0   519.3570          0          0          0          0
39     7   209.4464          0          0          0          0
40     7  1665.6117          0          0          0          0
78    21   260.8612          0          0          0          0
9     21   454.7456          0          0          0          0
[....]
```

Figure 4.3: Textual representation of the IMD data as an object of the class `"epidata"`.

Here, only the first six rows of the event table are printed. The `coordinates` are in the ETRS89 coordinate reference system (cf. Subsection 1.2.4). By this projection, the events are located in a Cartesian coordinate system, where the unit is one kilometre. The `time` is counted in days since 24 December 2001, which is the Monday of the week of the first case. Note the adjusted time point 3.99 resulting from the tie breaking procedure described in Subsection 1.2.4. The spatio-temporal grid has a week $\times$ county resolution, thus there are a total $D = 367$ time intervals and $M = 413$ tiles. The column `tile` equals the index $\xi(\boldsymbol{s}_i)$ of the spatio-temporal grid `stgrid`, here referenced by the official key of the county. The column `BLOCK` ($\tau(t_i)$) indexes the time interval (i.e. week) where the event occured, and its lower bound is given in the column `start`. The column `eps.t` equals $\varepsilon = 30$ [days], and `eps.s` equals $\delta = 150$ [km] (these choices will be discussed in the next chapter). While the columns `type`, `pnr` (patient number), `age`, and `sex` are marks of the events (i.e. the mark vector $\boldsymbol{m}_j$), the columns `popdensity` (population density), and `influenza[0-3]` contain the values of the endemic covariates at the time and location of the events (mapped from `stgrid`). The numbers of influenza cases for the corresponding week of `time` and county `tile` are given

as `influenza0`. The other influenza columns represent lagged versions of these numbers, for instance, `influenza1` is the case number of influenza in the county `tile` one week before the infection.

Based on such `epidata` objects, the function `twinstim` can perform maximum likelihood estimation for the model class of the same name. An example call is:

```
fit <- twinstim(
    endemic = ~ 0 + offset(log(popdensity)) + I(start/365) + influenza0,
    typeSpecificEndemicIntercept = TRUE,
    epidemic = ~ 0 + type, siaf = siaf_log, tiaf = tiaf,
    data = imd,
    optim.args = list(par = rep(0,9), method = "nlminb", hessian=TRUE),
    finetune = TRUE, nCub = 12, partial = FALSE
)
```

This specifies an endemic component with a finetype-specific intercept, an offset for the log-population density, a linear time trend and a linear effect of the case numbers of influenza. For numerical reasons and easier interpretation, the time in the linear trend is rescaled to years. The epidemic component also has a type-specific intercept (the `type` variable is coded as a `factor` variable in `R`). The arguments `siaf` and `tiaf` specify the spatial and temporal interaction functions, respectively. In fact, the interaction functions have to be provided with all necessary analytical derivatives and integrals. The type-invariant exponential decay function g_α from equation (4.8) is e.g. implemented as

```
tiaf <- list(
    g = function(t, alpha) exp(-alpha*t),
    G = function(t, alpha) if (alpha == 0) t else -exp(-alpha*t)/alpha,
    deriv = function(t, alpha) as.matrix(-t*exp(-alpha*t)),
    Deriv = function(t, alpha) if (alpha == 0) as.matrix(-t^2/2) else
        as.matrix((t+1/alpha)*exp(-alpha*t)/alpha),
    npars = 1
)
```

Here, $\texttt{G}_\alpha(t) = \int \texttt{g}_\alpha(t)\,\mathrm{d}t$, $\texttt{deriv}_\alpha(t) = \frac{\partial}{\partial\alpha}\texttt{g}_\alpha(t)$, and $\texttt{Deriv}_\alpha(t) = \int \texttt{deriv}_\alpha(t)\,\mathrm{d}t$ (see equations (4.22), (4.25), and (4.26)). Note that the special case $\alpha = 0$ is also accounted for, and no efforts are made to artificially restrict the parameter to the domain $\mathbb{R}_+$.

The specification of the spatial interaction function will be exemplified by a type-specific Gaussian kernel like in equation (4.15), but the parameters σ_κ are re-parametrised to the log-scale in order to avoid constrained optimisation and inference:

$$f_{\boldsymbol{\varphi}}(\boldsymbol{s}|\kappa) = \exp\left(-\frac{\|\boldsymbol{s}\|^2}{2\exp(\varphi_\kappa)^2}\right) \qquad (\boldsymbol{s} \in \mathbb{R}^2,\ \varphi_\kappa := \log(\sigma_\kappa))\,, \tag{4.28}$$

where $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_K)'$. For the evaluation of the score function and the Fisher information

matrix of Subsections 4.2.2 and 4.2.3, respectively, we further need the partial derivatives of f with respect to the elements of $\boldsymbol{\varphi}$:

$$\frac{\partial}{\partial \varphi_k} f_{\boldsymbol{\varphi}}(\boldsymbol{s}|\kappa) = \begin{cases} \exp\left(-\frac{\|\boldsymbol{s}\|^2}{2\exp(\varphi_k)^2}\right) \frac{\|\boldsymbol{s}\|^2}{\exp(\varphi_k)^2} & \text{if } k = \kappa, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for any } k \in \mathcal{K}\,. \tag{4.29}$$

The basic form of specification of the spatial interaction function then is:

```
siaf_log <- list(
    f = flog,
    deriv = deriv_flog,
    Fcircle = intCircle_flog,
    effRange = function (logsd) 6*exp(logsd),
    npars = K
)
```

Here, `flog` is the implementation of the type-specific Gaussian kernel from equation (4.28), and `deriv_flog` its derivative with respect to $\boldsymbol{\varphi}$ from equation (4.29):

```
flog <- function (s, logsd, types) {
    sLengthSquared <- rowSums(s^2)
    sd <- exp(logsd)[types]
    exp(-sLengthSquared/2/sd^2)
}


deriv_flog <- function (s, logsd, types) {
    sLengthSquared <- rowSums(s^2)
    L <- length(sLengthSquared)
    deriv <- matrix(0, L, length(logsd))
    sd <- exp(logsd)[types]
    deriv[cbind(1:L,types)] <- exp(-sLengthSquared/2/sd^2) * sLengthSquared / sd^2
    deriv
}
```

Note that the argument `s` above is a matrix of coordinates, and `logsd` is the vector $\boldsymbol{\varphi}$. Moreover, the implementation is designed to take special care concerning the integration over those influence regions R_j, which are actually discs completely inside W. This occurs if the minimal distance from $\boldsymbol{s}_j$ to the boundary of W is smaller than or equal to δ (the minimal distance is calculated by the `spatstat` function `bdist.points`). In this case, the Gaussian kernel can be integrated quickly through the distribution function of a χ^2-distribution as explained in Subsection 3.2.7. The calculation is implemented by the function given as `Fcircle = intCircle_flog`, which computes the integral of `f` over a disc with centre (0,0) as a function of its radius r:

```
intCircle_flog <- function(r, logsd, type) {
    sd <- exp(logsd)[type]
    pchisq((r/sd)^2, 2) * 2*pi*sd^2
}
```

Another application of this quick integration is motivated by computational convenience. The function `effRange` returns the "effective range" for the cubature for a specific value of $\varphi_\kappa = \log(\sigma_\kappa)$, which is chosen as $6\,\sigma_\kappa$ (cf. Subsection 3.2.7). Hence, if $R_j \cap b(\mathbf{0}, 6\,\sigma_\kappa)$ is a circular domain completely inside W, then `f` can be integrated quickly by `Fcircle` using $r = \min(6\,\sigma_\kappa, \delta)$ as the radius. However, if these tricks are not applicable, then the type-specific Gaussian kernel is integrated numerically by the adaptive two-dimensional midpoint rule. In the `twinstim` call on page 70, the parameter `nCub` determines the desired bandwidth $h = 0.5\sigma_\kappa =$ `effRange/nCub`, i.e. $\phi = 0.5$ (cf. Subsection 3.2.7).

The remaining arguments to the `twinstim` call on page 70 are control arguments for the optimisation procedure. The logical value `partial` determines whether to maximise the full or partial likelihood, and `optim.args` defines the start value for $\boldsymbol{\theta}$ (here a vector of zeros) and the type of algorithm. For instance, the Nelder-Mead simplex algorithm (Nelder & Mead, 1965) could be used as implemented in the `R` function `optim`. This is a general-purpose optimiser, which only uses function values (no derivatives), and – according to the documentation of `optim` – is "robust" and "works reasonably well for non-differentiable functions". It also provides a numerical approximation of the Hessian matrix at the optimum (if `hessian=TRUE`), which is the negative observed Fisher information matrix. The only drawback is that it is slow compared to Newton-methods which make use of function derivatives. In quasi-Newton methods, the Hessian matrix is iteratively updated numerically, a common example being the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (described in, e.g., Press et al., 2007, Section 10.9), which is as well implemented in `optim`. Another implementation can be found in the *PORT Mathematical Subroutine Library* (Fox, Hall & Schryer, 1978), a collection of FORTRAN 77 routines developed by Bell Labs. Its publicly available general minimisation routines (see `http://www.netlib.org/port/readme`) are accessible from within `R` through the function `nlminb`. In contrast to `optim`, it also accepts an analytical Hessian on input, which is used during optimisation and then usually converges faster than BFGS. This is why I found `method="nlminb"` to be the best choice for optimisation. Note that I used the estimation of the expected Fisher information matrix from equation (4.27) for the (negative) Hessian provided to `nlminb`.

To guard against false convergence, maximisation is commonly repeated from a range of starting values. Additionally, I always applied Nelder-Mead optimisation as a second step to validate and `finetune` the point of maximum likelihood found by `nlminb`. This did never reveal any substantial changes of the MLE. Furthermore, for numerical reasons, convergence is usually improved by scaling the log-likelihood function to have values around 1.

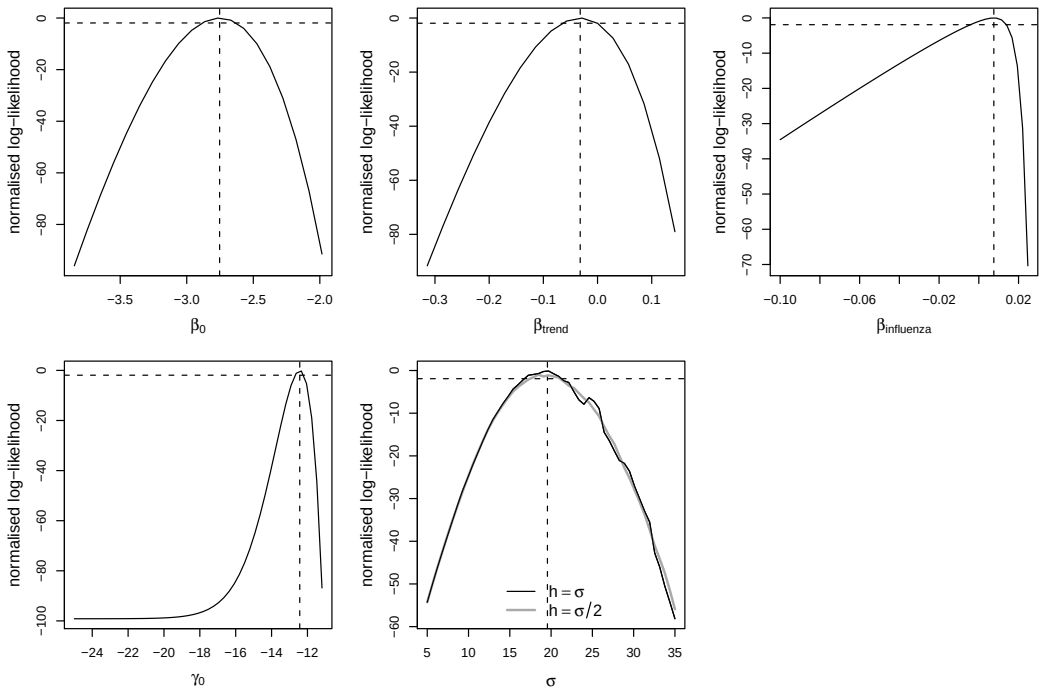


Figure 4.4: Shape of the normalised log-likelihood function around the MLE (indicated by the dashed vertical lines) of an example model for the serogroup B point pattern. The horizontal line is always at -1.92 such that a 95% confidence interval based on the estimated likelihood is bounded by the two intersections of the line with the normalised estimated log-likelihood. The parameters of the endemic component are in the top row, epidemic parameters are in the bottom row.

Figure 4.4 illustrates a typical quality of the shape of the log-likelihood function. The central point around which the partial shapes are computed is $(\hat{\beta}_0, \hat{\beta}_{\text{trend}}, \hat{\beta}_{\text{influenza}}, \hat{\gamma}_0, \hat{\sigma}) = (-2.8, -0.032, 0.0076, -12, 20)$, which is the MLE of the model. Each plot shows the normalised log-likelihood as a function of the parameter on the horizontal axis, when fixing the other parameters at their MLE value. This is known as the *estimated likelihood* approach. The bottom-right plot for the parameter σ of the spatial interaction function shows the log-likelihood shape using two different choices of accuracy for the two-dimensional midpoint rule. The evaluation using $\phi = 1$, i.e. an adaptive bandwidth of $h = \sigma$ is of course more noisy. In this case, trying different starting values would be especially important for the parameter σ. However, for reliability and reasonable computation time, I used $\phi = 0.5$ in the analyses of the IMD data presented in the next chapter. Note that there was no apparent difference in runtime between the two choices of ϕ, because in the plot, $\sigma > 5$: in relation to the affected integration regions R_j with $\delta = 150$, the small implied numbers of subdivisions for $h = \sigma$ ($n_x = n_y < 60$ subdivisions) and $h = 0.5\sigma$ ($n_x = n_y < 120$ subdivisions) have no crucial impact on computation time (cf. Table 3.1(b)). Concerning the intercept γ_0 of the epidemic component, maximisation must also be driven with care: If the algorithm happens to jump

to far to the left, i.e. the weight e^{γ_0} of the epidemic component becomes numerically equal to zero, then the derivative in γ_0 is approximately zero, the algorithm gets stuck, and σ is no longer identifiable.

The shape of the log-likelihood around its maximum suggests that symmetric Wald confidence intervals for the MLEs based on the quadratic approximation of the log-likelihood are applicable. The sharp decline of the log-likelihood for high values of $\beta_{\text{influenza}}$ is due to some high numbers of influenza cases (range: $0-314$). Rescaling the numbers of cases by e.g. considering `influenza/10` could improve the numerical conditioning here. In general, however, further development using efficient compiled `C` code for the likelihood evaluations would be welcome to enable the routine use of likelihood-based confidence intervals. Note that the generation of each of the plots in the figure took about 3.4 minutes. Because there are 50 evaluation points in each plot, a single evaluation of the log-likelihood cost about 4.0 seconds.

4.4 A Tricky but Insufficient Alternative

With regard to the sophisticated modelling and corresponding estimation presented in the previous sections, the question arises whether there is no simpler point process model for the analysis of the IMD data at hand. Indeed, in line with the Cox model, an appealing (unmarked) form for the conditional intensity function would be a log-linear model

$$\lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}) = \exp\left(\eta^*_{\boldsymbol{\theta}}(t, \boldsymbol{s})\right), \tag{4.30}$$

where

$$\eta^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}) = \boldsymbol{\theta}' \boldsymbol{z}^*(t, \boldsymbol{s})$$

is a linear predictor. Here, $\boldsymbol{z}^*(t, \boldsymbol{s})$ incorporates influence factors on the conditional intensity that may vary both in space and time, and may also depend on the history of the process. As for `twinstim`'s no constraints on $\boldsymbol{\theta}$ have to be imposed to guarantee nonnegativity of the conditional intensity. However, the major advantage of such modelling is that the corresponding log-likelihood function is almost surely concave (cf. Ogata, 1978, example on p. 255) so that, if it exists, its maximum is unique.

For the IMD data, possible influence factors would include

$$\boldsymbol{z}^*(t, \boldsymbol{s}) = \left(h_0^{\text{spat}}(\boldsymbol{s}), h_0^{\text{temp}}(t), z^{(0)}(t, \boldsymbol{s}), e^*(t, \boldsymbol{s})\right)',$$

where the baseline components $h_0^{\text{spat}}(\boldsymbol{s})$ and $h_0^{\text{temp}}(t)$ denote the population density and a time trend, respectively. The covariate $z^{(0)}(t, \boldsymbol{s}) \equiv z^{(0)}_{\tau(t),\xi(\boldsymbol{s})}$ denotes the number of cases of influenza during the week $\tau(t)$ of the time point t in the county $\xi(\boldsymbol{s})$ containing coordinate s.

The history-dependent term $e^*(t, \boldsymbol{s})$ is the same as in equation (4.9). Usually, the population density will be treated as an offset term, such that the conditional intensity function

$$\lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}) = h_0^{\text{spat}}(\boldsymbol{s}) \cdot \exp\left(\eta^*_{\boldsymbol{\theta}}(t, \boldsymbol{s})\right)$$

is proportional to the population density.

The charm of the log-linear model (4.30) for the conditional intensity lies in a tricky estimation procedure discovered by Berman & Turner (1992) for purely temporal point processes and the spatial Poisson process. Specifically, the log-likelihood related to the log-linear CIF model can easily be transformed into the log-likelihood of a weighted Poisson GLM. Thus, the broad range of well designed standard implementations can be used to obtain estimates and standard errors of the parameters. Basically, the trick is to include the observed event locations into the set of evaluation points for the numerical approximation of the integral term in the log-likelihood (2.4):

$$\int_0^T \int_W \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s} \approx \sum_{k=1}^{n_T} \tilde{w}_k \sum_{l=1}^{n_W} \tilde{v}_l \lambda^*_{\boldsymbol{\theta}}(\tilde{t}_k, \tilde{\boldsymbol{s}}_l) \,, \tag{4.31}$$

where $\tilde{w}_k$ and $\tilde{v}_l$ are quadrature and cubature weights, respectively, and

$$\begin{array}{rcl} \text{“data points”} & \subset & \text{“design points”} \\ \boldsymbol{x} := \{(t_i, \boldsymbol{s}_i) : i = 1, \ldots, N\} & \subset & \{(\tilde{t}_k, \tilde{\boldsymbol{s}}_l) : k = 1, \ldots, n_T, \, l = 1, \ldots, n_W\} =: \boldsymbol{G} \,. \end{array}$$

Hence, the log-likelihood can be written as

$$\begin{aligned} l(\boldsymbol{\theta}) &\approx \sum_{i=1}^{N} \log(\lambda^*_{\boldsymbol{\theta}}(t_i, \boldsymbol{s}_i)) - \sum_{k=1}^{n_T} \sum_{l=1}^{n_W} \tilde{w}_k \, \tilde{v}_l \lambda^*_{\boldsymbol{\theta}}(\tilde{t}_k, \tilde{\boldsymbol{s}}_l) \\ &= \sum_{k=1}^{n_T} \sum_{l=1}^{n_W} w_{kl} \left(\frac{\delta_{kl}}{w_{kl}} \log \lambda^*_{kl}(\boldsymbol{\theta}) - \lambda^*_{kl}(\boldsymbol{\theta}) \right) , \end{aligned}$$

where $w_{kl} := \tilde{w}_k \, \tilde{v}_l$ are “observation” weights,

$$\lambda^*_{kl}(\boldsymbol{\theta}) := \lambda^*_{\boldsymbol{\theta}}(\tilde{t}_k, \tilde{\boldsymbol{s}}_l) = \exp\left(\eta^*_{kl}(\boldsymbol{\theta})\right) \qquad \text{with} \qquad \eta^*_{kl}(\boldsymbol{\theta}) := \eta^*_{\boldsymbol{\theta}}(\tilde{t}_k, \tilde{\boldsymbol{s}}_l) \,,$$

and $\delta_{kl} = \mathbb{1}_{\boldsymbol{x}}((\tilde{t}_k, \tilde{\boldsymbol{s}}_l))$, which indicates if the grid point $(\tilde{t}_k, \tilde{\boldsymbol{s}}_l)$ is a data point. Up to a constant not depending on the parameters, this log-likelihood is known from Poisson GLMs with pseudo reponse variables $\frac{\delta_{kl}}{w_{kl}}$, weights w_{kl}, log-link, and linear predictor terms $\eta^*_{kl}(\boldsymbol{\theta})$. Note that the linear predictor must be known not only at the locations $\boldsymbol{x}$ of observed events, but at every point of the grid $\boldsymbol{G}$.

The extension of the Berman & Turner device to a wider class of spatial point processes as described in Baddeley & Turner (2000) is e.g. used in `spatstat` for the estimation of such

models. However, the approach is insufficient for the modelling of self-exciting point processes like applied in this thesis. The log-linear formulation of the CIF (4.30) restricts estimation to the intercept γ_0 of the triggering function, which can be factorised from the history-dependent sum $e^*(t, \boldsymbol{s})$. Further parameters like the effects of marks, or the parameters of the interaction functions $f_{\boldsymbol{\sigma}}$ and $g_{\boldsymbol{\alpha}}$ are not log-linear. Thus, their estimation is not possible by the Berman & Turner device, which actually means that the whole triggering function must be given in advance.

As a consequence, I only investigate in the next chapter the applicability of the presented `twinstim` class on the IMD data.

5 Application

The previous chapter introduced the `twinstim` class for (marked) point processes in continuous space-time with a self-exciting character. This chapter will now apply the theoretical considerations on the IMD data presented in Section 1.2 which motivated the development of `twinstim`. Model selection and interpretation will in turn be carried out for the serogroup B finetype (Section 5.1), for the serogroup C finetype (Section 5.2), and then for the joint point pattern of both finetypes (Section 5.3).

For all models, I will use the logarithmic population density as an offset in the linear predictor of the endemic component. Hence, the endemic risk of observing an infection in a specific county is always proportional to its population density. Furthermore, I assume the maximum interaction ranges $\varepsilon = 30\,[\text{days}]$ and $\delta = 150\,[\text{km}]$. These choices are both conservative, because the true ranges are expected to be much smaller. The temporal cluster size is chosen based on the results in Elias et al. (2006) already mentioned in Subsection 1.1. For their space-time cluster detection, they also used a maximum range of 30 days and found that "the time between the first and the second case may rarely exceed this temporal limit". The longest detected interaction then in fact was 24 days. Imposing a spatial bound δ on the distance over which transmission occurs is mainly motivated by a speed-up of the estimation procedure and improved accuracy of the numerical integration for narrow kernels. Furthermore, it has also a subject matter motivation because the occurence of two cases at a large distance away from each other is unlikely caused by direct transmission of the disease. In Elias et al. (2006), the spatial cluster size was bounded by the population included in the clusters rather than their spatial extent in kilometres. This imposes smaller cluster sizes in metropolitan areas like Berlin, and permits larger ones in rural areas. Specifically, they chose the maximum spatial cluster size to correspond to 7% of the German population (5 777 219).

Concerning the interaction functions, besides from the constant versions $g(t) = 1$ and $f(\boldsymbol{s}) = 1$, only the Gaussian kernel (4.7) (type-invariant) or (4.15) (type-specific), and the exponential decay function (4.8) (type-invariant) or (4.16) (type-specific) are considered. The Gaussian kernels are reparametrised to depend on the standard deviation on the log-scale to avoid constrained inference, see equation (4.28).

Selecting a model in the regression context usually begins with the smallest or the largest possible model, followed by forward, backward or stepwise selection of model components. In

principle, this procedure is also advisable for selecting a point process model through its CIF. However, the computational cost of model estimation is tremendous once a spatial interaction function is included in the model. Thus, as a pragmatic approach, I always estimate some preliminary models with assumed homogeneous mixing, i.e. $g(t) = 1$, $f(\boldsymbol{s}) = 1$ – also to determine sensible start values for the subsequent estimation of models with inhomogeneous triggering functions. In order to illustrate `twinstim`, I will provide brief interpretations along with the intermediate models during model selection.

Table 5.1 defines the notations used throughout this chapter.

Symbol	`R` variable	Description
$(t, \boldsymbol{s}, \kappa)$	-	Days since 24 December 2001, coordinate in ETRS89 (kilometre scale), finetype (B=1, C=2).
$(\tau(t), \xi(\boldsymbol{s}))$	-	Week number of time point t, county of coordinate $\boldsymbol{s}$.
$\lfloor t \rfloor$	`start`	Monday of week $\tau(t)$, i.e. lower bound of time intervals $C_1, \ldots, C_D$.
$\rho_{\xi(\boldsymbol{s})}$	`popdensity`	County-specific population density (inhabitants per km^2).
$z^{(l)}_{\tau(t),\xi(\boldsymbol{s})}$	`influenza`l	Number of influenza cases registered at the RKI in the county $\xi(\boldsymbol{s})$, l weeks before the week $\tau(t)$, $l \in \{0, 1, 2, 3\}$.
$m_{j,\text{sex}}$	`sex`	Indicator variable for the patient's gender (0=female, 1=male)
$m_{j,\text{age}}$	`age`	Age of the patient
$\varphi_\kappa = \log(\sigma_\kappa)$	`siaf.`κ	Standard deviation of the Gaussian kernel on the log-scale.
α_κ	`tiaf.`κ	Parameter of the temporal interaction function $g_{\boldsymbol{\alpha}}$.

Table 5.1: Notation of variables used throughout this chapter.

5.1 Finetype B:P1.7-2,4:F1-5

The first model fitted to the observed point pattern of 336 cases of the serogroup B finetype (cf. Figures 1.2(a) and 1.3(a)) is

$$\begin{aligned}\lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}) = \rho_{\xi(\boldsymbol{s})} \cdot \exp\Bigg(\beta_0 + \beta_{\text{trend}}\frac{\lfloor t \rfloor}{365} + \beta_{\sin}\sin\left(\lfloor t \rfloor \cdot \frac{2\pi}{365}\right) + \beta_{\cos}\cos\left(\lfloor t \rfloor \cdot \frac{2\pi}{365}\right) \\ + \sum_{l=0}^{3} \beta_{\text{influenza},l}\, z^{(l)}_{\tau(t),\xi(\boldsymbol{s})}\Bigg) \;+ \sum_{j \in I^*(t,\boldsymbol{s})} \exp(\gamma_0)\,.\end{aligned} \tag{5.1}$$

The epidemic component models a completely homogeneous triggering function, i.e. it simply equals e^{γ_0} times the number $|I^*(t, \boldsymbol{s})|$ of potential sources of infection. In other words, every infectious individual spreads the disease at the homogeneous rate e^{γ_0} during her infectious period of 30 days inside her spatial influence range of 150 kilometres around her position. The

endemic linear predictor consists of an intercept, a linear time trend scaled to years (leaving aside leap years), a sine-cosine wave at frequency $\frac{2\pi}{365}$ corresponding to a time-of-year effect, and effects for the case numbers of influenza with lags from 0 to 3 weeks. The estimation took 1.5 minutes and the results are shown in Table 5.2. Its first part contains the parameter estimates, where endemic and epidemic parameters are separated by a horizontal line, and the second part shows the AIC and the log-likelihood of the model.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	-20.402892	0.135465	-150.614	$< 2 \cdot 10^{-16}$
`I(start/365)`	-0.027941	0.032837	-0.851	0.394839
`sin(start*2*pi/365)`	0.276376	0.099496	2.778	0.005473
`cos(start*2*pi/365)`	0.369904	0.095454	3.875	0.000107
`influenza0`	0.015920	0.012306	1.294	0.195793
`influenza1`	-0.018402	0.040532	-0.454	0.649817
`influenza2`	0.006092	0.039275	0.155	0.876741
`influenza3`	-0.008332	0.025397	-0.328	0.742851
`(Intercept)`	-15.9431	0.2468	-64.6	$< 2 \cdot 10^{-16}$
AIC:	9986.4			
Log-likelihood:	-4984			

Table 5.2: Results from the initial fit. The statistics of endemic and epidemic parameters are separated by a horizontal line. The first `(Intercept)` row thus belongs to the endemic intercept β_0, whereas the second one belongs to $\hat{\gamma}_0$.

Both intercepts are of course significantly different from zero because they scale the effects to the level of the spatio-temporal intensity. For a better understanding, just assume that the pattern stems from a stationary and homogeneous Poisson process. Then, the maximum likelihood estimate of the rate of this process would be

$$\hat{\lambda}_{ML} = \frac{N}{T\,|W|} = \frac{336}{2569 \cdot 357\,603} \approx 3.66 \cdot 10^{-07} \approx \exp(-14.8)\,,$$

which equals the mean number of events per day per km^2. The endemic intercept additionally adjusts for the level of the population density, which ranges up to 4 225 inhabitants per km^2 in Munich.

To give a first idea of the appealing interpretation of the epidemic intercept γ_0 through equation (4.12): Would the above model be true, then each infected individual caused on average $\hat{\mu}_j = \hat{\mu} = \exp(\hat{\gamma}_0) \cdot 30 \cdot 150^2\,\pi \approx 0.25$ secondary cases, which epidemiologists would call the basic reproduction number R_0 of the disease. The `twinstim` class thus offers a practical way to estimate this number while adjusting for endemic effects like the population density. A 95% Wald confidence interval (CI) for $\hat{\mu}$ can be constructed as a transformation of the confidence interval for $\hat{\gamma}_0$ (its bounds are plugged into the above formula for $\hat{\mu}$), which yields the range $[0.16, 0.41]$.

The time trend is slightly negative and its value would mean that the endemic background risk for infections decreased about $(1 - e^{\hat{\beta}_{\text{trend}}}) \cdot 100\% \approx 2.8\%$ per year. The sine-cosine wave for time-of-year effects seems to be an important explanatory covariate, whereas none of the case numbers of influenza is significant. Certainly, they partially explain a similar characteristic through their seasonality as the deterministic wave. However, the correlations between the sine-cosine terms and the influenza covariates do not exceed the value 0.184. As an alternative to linear modelling of the influenza case numbers I considered the transformations $\log(1 + \texttt{influenza})$ and $\sqrt{\texttt{influenza}}$, but they yielded smaller likelihoods and would also complicate interpretation.

Starting from the above model, I next removed the most ambigious effects (according to the p-value) in turn and checked the decrease of the AIC value. The results are listed in Appendix B. Observe that the time trend as well as all influenza effects (in the order of lags 2, 3, 1, 0) were dropped from the model. I checked the sensitivity of the model selection with regard to exclusion of the epidemic component. Carrying out the selection procedure without the epidemic component, the above mentioned terms were equally dropped from the model. It should moreover be noted that the AIC was worse compared to the models with epidemic component. For instance, fitting the initial model without the epidemic component yields an AIC of 10 007 (compared to 9 986.4 when it is included).

This preliminary model selection suggests that a deterministic trigonometric oscillation sufficiently explains the seasonal variations of the serogroup B finetype. On top of that, there seems not to be a specific effect of influenza in the sense that larger waves of influenza cause more cases of this IMD finetype. Trying to improve the time-of-year effect by adding its first harmonic (half-year period) $\sin(2\omega\lfloor t \rfloor) + \cos(2\omega\lfloor t \rfloor)$, where $\omega := \frac{2\pi}{365}$, worsens the AIC (see Table B.6). It could be supposed that the deterministic sine-cosine wave obscures the effect of influenza. However, running the selection procedure without this deterministic wave, all endemic effects but the intercept are in turn removed from the model, and the order of deselection of the different influenza lags is the same as well.

As a next step, I modelled inhomogeneous spatial and temporal interaction functions. To be sure, I also included the influenza effects with lags 0 and 1 in the model. The parameter estimates from a previous model have been used as start values for the optimisation algorithm, but several runs using different start values for the new parameter $\varphi = \log(\sigma)$ were necessary for convergence to the point of maximum likelihood. This is related to the discussion on Figure 4.4. Note that each run now costs considerably more computation time due to the necessary bivariate numerical integration of $f_\varphi(\boldsymbol{s})$. On average the log-likelihood maximisation of the following models took 56 minutes (range: $27 - 87$). The results of the converged model estimation are shown in Table 5.3.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.56630	0.07100	−289.683	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.24048	0.09779	2.459	0.013922
`cos(start*2*pi/365)`	0.30964	0.09397	3.295	0.000984
`influenza0`	0.01622	0.01088	1.491	0.136032
`influenza1`	−0.01726	0.01722	−1.003	0.316069
`(Intercept)`	−12.13582	0.29143	−41.643	$< 2 \cdot 10^{-16}$
`siaf.1`	2.91783	0.09636	30.281	$< 2 \cdot 10^{-16}$
`tiaf.1`	0.01886	0.01723	1.095	0.274
AIC:	9857.4			
Log-likelihood:	−4921			

Table 5.3: Results for the model with inhomogeneous spatial *and* temporal interaction function.

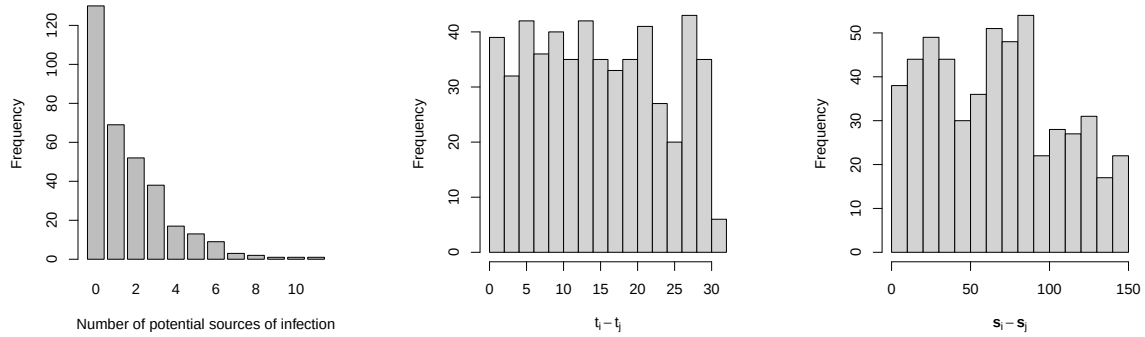
The estimates of the endemic component do not differ substantially from the corresponding model with homogeneous mixing (see Table B.3). Higher values of `influenza0` – the numbers of influenza cases reported to RKI during the actual week and county of the IMD infections – seem to increase the endemic risk of infection. If there occur 10 more cases of influenza in a particular week and county, given the same internal history of the process (i.e. the same past cases of IMD) and the same number of influenza cases in the previous week, the model would predict a local increase of the endemic risk of infection with the IMD finetype B by about 18%. The effect is however not significant on the 5% level and the 95% Wald confidence interval is $[-0.005106, 0.037550]$. It is also questionable whether the effect of influenza on the endemic predictor is linear at all. The lagged number of influenza cases `influenza1` would have an inverse interpretation, because the point estimate of the effect is negative.

Compared to the models with homogeneous mixing, the epidemic intercept γ_0 increased in the current model. This is because it compensates for the narrowed spatial range of interaction induced by the spatio-temporally decreasing triggering function. However, the basic reproduction number implicated by this model is $\hat{\mu} \approx 0.26$ and is thus almost the same as with homogeneous mixing. Note that providing a confidence interval for $\hat{\mu}$, which is in fact a complex transformation of the basic estimators $\hat{\gamma}_0$, $\hat{\alpha}$, and $\hat{\varphi}$, is no longer straightforward by means of just transforming basic confidence intervals. If a simulation algorithm for `twinstim`'s was available, then parametric bootstrap could be used to compute approximate confidence intervals for $\hat{\mu}$ and any other statistics.

Concerning the difficulties of convergence, the parameter α of the temporal interaction function might be the reason for hard identifiability of the model. Observe that endemic effects generally receive more information from the data than epidemic effects. This is because the endemic effects concern all observed realisations $\lambda^*(t_i, \boldsymbol{s}_i)$ of the conditional intensity, whereas epidemic effects only appear in a limited number, depending on the set $I^*(t_i, \boldsymbol{s}_i)$ of potential sources of infection for individual i. For instance, individuals that attracted the disease far away from infective individuals will have $I^*(t_i, \boldsymbol{s}_i) = \emptyset$, which means that epidemic effects can

not be identified from such observations.

The distribution of the number $|I^*(t_i, \boldsymbol{s}_i)|$ of potential causative events for the 336 observed infections is shown in Figure 5.1(a). As much as 130 infections (i.e. 38.7%) do not have any progenitors 30 days before and 150 km around them. Most other infections have 1 to 3 potential vectors. The largest cluster contains 11 possible vectors of the disease. Figure 5.1(b) shows a histogram of the temporal distances $t_i - t_j$ between possible sources and infections. These distances are well uniformly distributed over the range $(0, 30]$. The same type of histogram for the spatial distances $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ in (c) shows a similar broad distribution.



(a) Observed numbers $|I^*(t_i, \boldsymbol{s}_i)|$ of potential sources of infection.

(b) Histogram of the temporal distance to the potential sources of infection.

(c) Histogram of the spatial distance to potential sources of infection.

Figure 5.1: Figures related to the potential sources of infection $I(t_i, \boldsymbol{s}_i)$ for finetype B:P1.7-2,4:F1-5.

In Table 5.3 the parameter estimate $\hat{\alpha}$ (`tiaf.1`) of the temporal interaction function is not significantly different from zero, i.e. from a temporally constant infectivity. Figure 5.2 shows the fitted exponential decay with $\hat{\alpha} = 0.01886$ together with its lower and upper bound corresponding to the 95% Wald confidence interval $[-0.01491, 0.05264]$ for $\hat{\alpha}$. Because this confidence interval includes the value $\alpha = 0$, the hypothesis of a constant temporal interaction function $g(t) = 1$ can not be rejected.

As a consequence, and because α seemed to be hardly identifiable, the temporal interaction function is again assumed to be constant. This yields an AIC of 9 856.7 (see Table B.7), which is smaller than in the previous model. With removed α, convergence of the optimisation procedure was no longer a problem. Because the numbers of influenza cases still do not explain much of the IMD dynamics, they are finally dropped from the model (see Table B.8 for the intermediate drop of `influenza1` only) yielding a further decrease of AIC and the estimates in Table 5.4.

Removing any of the remaining components would re-increase AIC. Hence, we have found

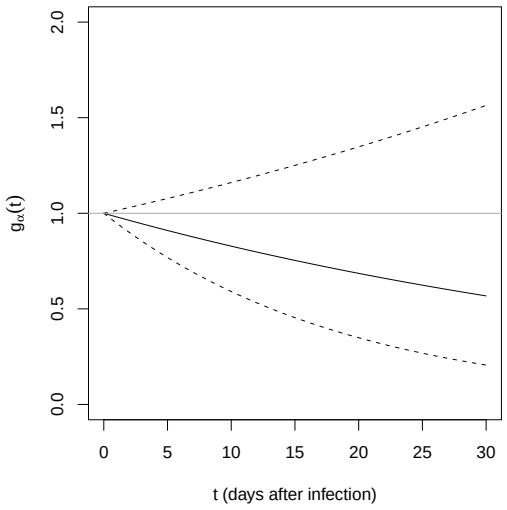


Figure 5.2: Fitted temporal interaction function (solid) and lower and upper bounds (dashed) corresponding to 95%-confidence limits of $\hat{\alpha}$. The null hypothesis of time-invariant infectivity $g(t) = 1$ is symbolised by the grey line.

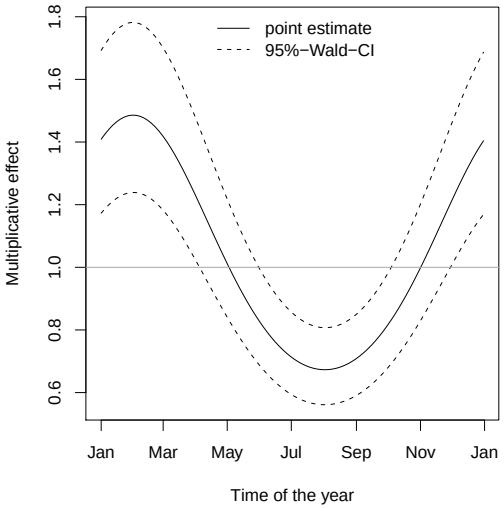
	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.56740	0.06951	−295.892	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.24343	0.09346	2.605	0.009200
`cos(start*2*pi/365)`	0.31240	0.09385	3.329	0.000873
`(Intercept)`	−12.39911	0.18914	−65.55	$< 2 \cdot 10^{-16}$
`siaf.1`	2.91739	0.09662	30.19	$< 2 \cdot 10^{-16}$
AIC:	9854.4			
Log-likelihood:	−4922			

Table 5.4: Results for the selected model with spatial interaction function.

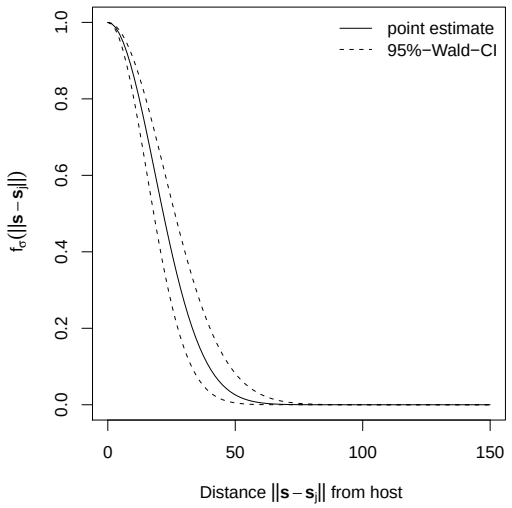
a simple, but well-identified model, which explains the spatio-temporal dynamics of the serogroup B finetype by baseline seasonality and a distance-dependent triggering function

$$\hat{e}_j(t, \boldsymbol{s}) = \exp(\hat{\gamma}_0) \cdot f_\varphi(\boldsymbol{s} - \boldsymbol{s}_j) = \exp(\hat{\gamma}_0) \cdot \exp\left(-\frac{\|\boldsymbol{s} - \boldsymbol{s}_j\|^2}{2\,\exp(\hat{\varphi})^2}\right) = \exp\left(-12.4 - \frac{\|\boldsymbol{s} - \boldsymbol{s}_j\|^2}{2 \cdot 18.49^2}\right) .$$

A point estimate for the corresponding basic reproduction number is $\hat{\mu} \approx 0.27$, i.e. about one out of four infectives causes a secondary case. Figure 5.3(a) shows the estimated time-of-year effect $\exp\left(\hat{\beta}_{\sin} \sin(t\,\omega) + \hat{\beta}_{\cos} \cos(t\,\omega)\right)$ which multiplies the global endemic baseline rate $\exp(\hat{\beta}_0) \approx 1.17 \cdot 10^{-09}$. Note that the discretisation of time on the weekly grid has been ignored for this illustration. The typical increase of IMD cases around late February (see Figure 1.5(b)) is caught by this function. The estimated spatial interaction function, which is an isotropic Gaussian kernel, is plotted as a function of the distance from the infective individual in Figure 5.3(b). It suggests that transmission of the finetype B:P1.7-2,4:F1-5 is effectively limited to about 50 km from the residence of the infected individual. This



(a) Time-of-year effect.



(b) Spatial kernel as a function of the distance $\|\boldsymbol{s} - \boldsymbol{s}_j\|$ from the infective individual.

Figure 5.3: Fitted seasonality and isotropic spatial interaction function $f(\boldsymbol{s}) \equiv f(\|\boldsymbol{s}\|)$.

represents, in part, the natural boundedness of individuals' movements.

The analysis carried out so far assumed homogeneous individuals, i.e. the available marks of gender and age have not been used to discriminate infectivity. Starting from the above selected model, I investigated a gender effect, a linear effect of the age (on the predictor level) and an effect for infants. The age effects both were not significant (see Tables B.9 and B.10). Note that the goodness-of-fit of these models can not be compared to the previous models without marks, because age or gender were missing for 4 individuals and estimation could only be performed on the subset of complete observations. Further investigations trying to identify different levels of infectivity for specific age groups could be valuable, but are not considered in this thesis. In contrast to the age effect, male and female infectives seem to have a different ability to transmit the disease. Table 5.5 shows the estimation results of the selected model enriched by an effect of gender.

Again, the (log-)likelihood of this model is not comparable with above results due to 3 missing values for the gender of the IMD cases of the serogroup B finetype. However, the parameter estimates of the endemic component and also the estimated log-standard deviation of the Gaussian kernel are very similar to the estimates of the unmarked model in Table 5.4. The mean number of infections entailed by a female infective is $\hat{\mu}_{\text{female}} \approx 0.38$, whereas male cases only cause $\exp(\hat{\gamma}_{\texttt{sexmale}}) \cdot \hat{\mu}_{\text{female}} \approx 0.17$ secondary cases on average, i.e. less than half as much as females. The 95% Wald confidence interval for the proportionality factor $\exp(\hat{\gamma}_{\texttt{sexmale}})$ for males is $[0.233, 0.888]$.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.57285	0.06964	−295.428	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.27003	0.09373	2.881	0.003967
`cos(start*2*pi/365)`	0.32228	0.09421	3.421	0.000624
`(Intercept)`	−12.0516	0.2236	−53.907	$< 2 \cdot 10^{-16}$
`sexmale`	−0.7883	0.3417	−2.307	0.0210
`siaf.1`	2.9178	0.0955	30.552	$< 2 \cdot 10^{-16}$
AIC:	9774			
Log-likelihood:	−4881			

Table 5.5: Results for the model with gender-specific infectivity.

5.2 Finetype C:P1.5,2:F3-3

For the observed point pattern of 300 cases of the serogroup C finetype, this section presents a similar analysis as for the serogroup B finetype in the previous section. The results are provided more concisely as above, while similarities and differences with respect to the serogroup B finetype are discussed. Model selection begins with the model of equation 5.1, i.e. a rich endemic component and the simplest epidemic component of constant interaction. Estimation results of this initial model are given in Table 5.6 below.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.324542	0.124533	−163.206	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.057714	0.031967	−1.805	0.07101
`sin(start*2*pi/365)`	0.357065	0.098528	3.624	0.00029
`cos(start*2*pi/365)`	0.170707	0.095221	1.793	0.07301
`influenza0`	−0.026165	0.045310	−0.577	0.56363
`influenza1`	0.004043	0.045546	0.089	0.92927
`influenza2`	0.010041	0.037261	0.269	0.78756
`influenza3`	−0.008592	0.023066	−0.372	0.70954
`(Intercept)`	−16.4509	0.3135	−52.48	$< 2 \cdot 10^{-16}$
AIC:	9198.8			
Log-likelihood:	−4590			

Table 5.6: Results of the initial model for the point pattern of finetype C:P1.5,2:F3-3.

This model implies the point estimate $\hat{\mu} \approx 0.15$ for the basic reproduction number, which is smaller than in the same model for the serogroup B finetype (see Table 5.2). A 95% confidence interval is $[0.082, 0.281]$, which overlaps with the corresponding confidence interval from the serogroup B finetype. Although the negative time trend is more pronounced for the serogroup C finetype, it is not significant on the 5% level in this model. The sine-cosine wave for time-of-year effects is as important as for the serogroup B finetype, while again, none of the lags of the numbers of influenza cases has a significant impact.

Appendix C contains the intermediate results of the subsequently performed model selection.

Following the AIC, the influenza effects with lags 1, 3 and 2 weeks are removed from the model. Removing `influenza0` leads to a small increase of the AIC. However, a likelihood ratio test of the hypothesis $\beta_{\text{influenza},0} = 0$ has p-value 0.1402, and thus there is no statistical evidence for an effect of the numbers of influenza cases. Alternatively, modelling an effect of `influenza0` without a sine-cosine wave yields a worsened AIC of 9205.8 and a p-value of 0.8134 for the Wald test on `influenza0`. Adding a half-year period sine-cosine wave did not improve the goodness-of-fit. Fitting the initial model without the epidemic component yields a worse AIC of 9 210.3.

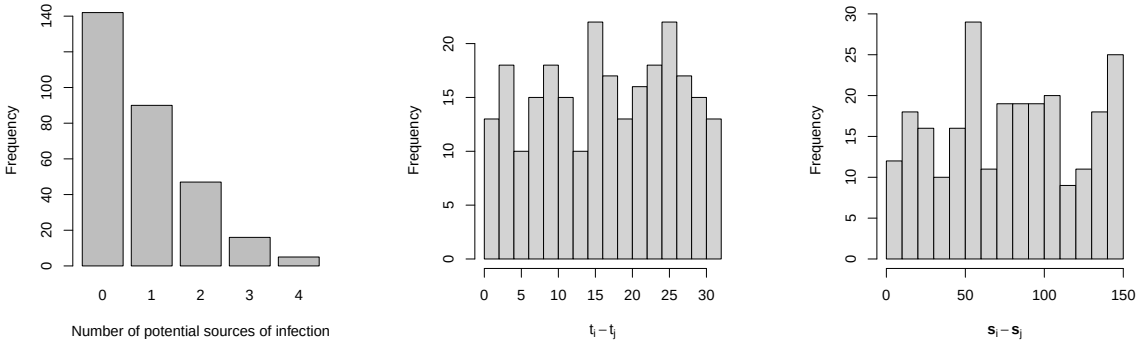
As a next step, inhomogeneous interaction functions were included into the epidemic component. To be sure about its role an effect of `influenza0` was also considered. Estimating the model was equally problematic as for the serogroup B finetype, and different start values were tried for convergence of the optimisation routine. The runtime of likelihood maximisation for the models with an inhomogeneous spatial interaction function was 47 minutes on average (range: $22 - 87$). The results of the highest attained likelihood are shown in Table 5.7.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.28084	0.11986	−169.206	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.06961	0.03194	−2.180	0.02929
`sin(start*2*pi/365)`	0.35357	0.09599	3.683	0.00023
`cos(start*2*pi/365)`	0.20535	0.09203	2.231	0.02565
`influenza0`	−0.02216	0.03176	−0.698	0.48537
`(Intercept)`	−13.176888	0.387036	−34.05	$< 2 \cdot 10^{-16}$
`siaf.1`	2.957433	0.160761	18.40	$< 2 \cdot 10^{-16}$
`tiaf.1`	0.007578	0.022308	0.34	0.734
AIC:	9158.5			
Log-likelihood:	−4571			

Table 5.7: Results for the model with inhomogeneous spatial *and* temporal interaction function.

Although these estimation results are not reliable (as we will see with the next model), I will provide some brief comments. Concerning the estimates of the endemic component, the time trend is significantly negative, whereas the effect of `influenza0` remains ambigious. A point estimate for the basic reproduction number implicated by this model is $\hat{\mu} \approx 0.12$, which is smaller than with assumed homogeneous mixing.

Figure 5.4 shows the distributions related to the observed potential sources of infection. The numbers $|I^*(t_i, \boldsymbol{s}_i)|$ are even smaller than for the serogoup B finetype, which indicates, that most infections of the serogroup C finetype did not occur by direct person-to-person transmission, and is in concordance with the lower basic reproduction number of this finetype. Specifically, as much as 142 infections (i.e. 47.3%) do not have any progenitors 30 days before and 150 km around them. The largest clusters contain only four possible vectors of



(a) Observed numbers $|I^*(t_i, \boldsymbol{s}_i)|$ of potential sources of infection.

(b) Histogram of the temporal distance to the potential sources of infection.

(c) Histogram of the spatial distance to potential sources of infection.

Figure 5.4: Figures related to the potential sources of infection $I(t_i, \boldsymbol{s}_i)$ for finetype C:P1.5,2:F3-3.

the disease. With regard to this characteristic, it is plausible that the parameter α of the temporal interaction function is hardly identifiable from the data.

Nevertheless, Figure 5.5 shows the fitted exponential decay with $\hat{\alpha} = 0.007578$ and its lower and upper bound corresponding to the large 95% Wald confidence interval $[-0.03615, 0.05130]$ for $\hat{\alpha}$. The Figures 5.4(b) and 5.4(c) show histograms of the temporal and spatial distances between possible sources and infections. As for the point pattern of the serogroup B finetype, these distances have a broad distribution over the ranges $(0, 30]$ and $[0, 150]$, respectively.

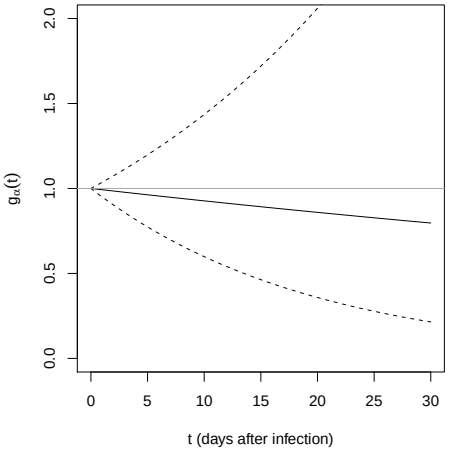


Figure 5.5: Fitted temporal interaction function (solid) and lower and upper bounds (dashed) corresponding to the 95%-confidence limits of $\hat{\alpha}$. The null hypothesis of time-invariant infectivity $g(t) = 1$ is symbolised by the grey line.

Because the parameter α seemed to be hardly identifiable, the temporal interaction function is henceforth assumed to be constant. By this simplification the optimisation procedure

converges without problems. It yields an AIC of 9 156.4 (see Table C.5), which is smaller than in the previous model. Note in particular that the new maximum log-likelihood value -4 571.19 is slightly higher than *without* the restriction $\alpha = 0$ in Table 5.7 (-4 571.23), which theoretically is not possible, and thus shows that the estimation of the last model actually did not converge to the global maximum of the log-likelihood. Because the effect of `influenza0` still seems ambigious (the corresponding Wald test has p-value 0.5198), I estimated the model without this effect the results of which are shown in Table 5.8 below.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.27226	0.11827	−171.400	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.07522	0.03097	−2.429	0.01514
`sin(start*2*pi/365)`	0.30717	0.09083	3.382	0.00072
`cos(start*2*pi/365)`	0.20038	0.08978	2.232	0.02562
`(Intercept)`	−13.0936	0.3114	−42.04	$< 2 \cdot 10^{-16}$
`siaf.1`	2.8362	0.1493	18.99	$< 2 \cdot 10^{-16}$
AIC:	9156.9			
Log-likelihood:	−4572			

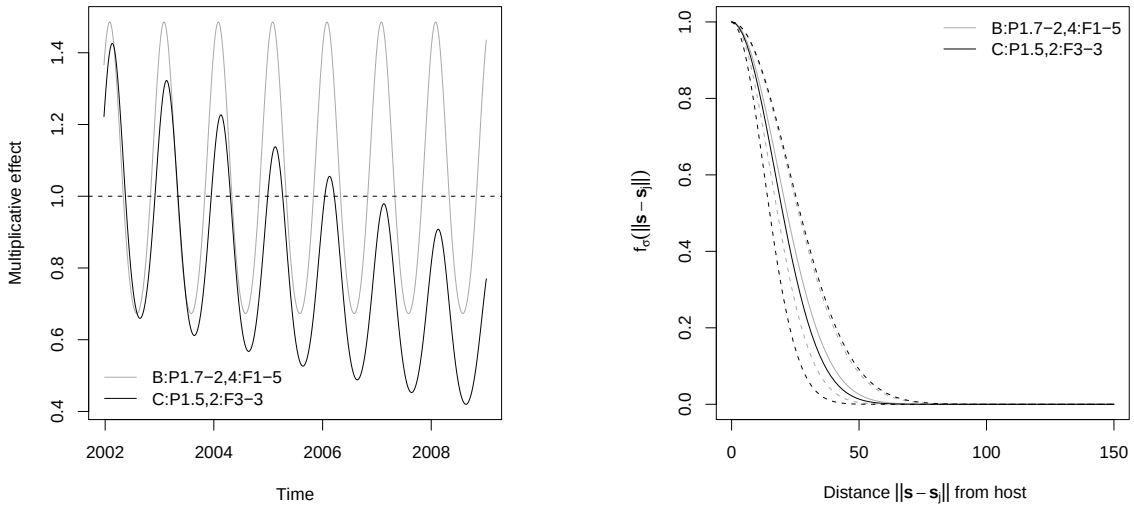
Table 5.8: Results for the selected model with inhomogeneous spatial interaction function.

There is a very small increase of the AIC, but the likelihood ratio test of the hypothesis $\beta_{\text{influenza},0} = 0$ has p-value 0.1148. Hence, there is still no evidence that the numbers of influenza cases explain the endemic risk of infection. However, in contrast to the finally selected model for the serogroup B point pattern, there remains a significant negative effect of the time. The endemic background risk for infections with the serogroup C finetype discreased about 7.2% per year (95%-CI: $[1.4, 12.7]$%). This decrease was not apparent from the plot of the monthly case numbers in Figure 1.2(b), but observe that the model adjusts for the clustering of cases due to direct transmission of the disease. The described temporal decrease refers to the *endemic* background risk only, which is not observable from the aggregated time series. The decreasing endemic risk could e.g. be induced by the new vaccination policy against serogroup C introduced by the STIKO in 2006 (see Section 1.1). The recommendation of vaccination is assumed to have entailed an increased coverage rate of vaccination in the population since 2006, but detailed data are not yet available (Hellenbrand et al., 2008, p. 269). However, more data and detailed analyses are necessary to evaluate the effect of the vaccination coverage. Changepoint models or a flexible modelling of the time trend could be of interest in this context, but are not further investigated in this thesis.

The basic reproduction number in the selected model for finetype C:P1.5,2:F3-3 is estimated to be $\hat{\mu} \approx 0.11$, i.e. about only one out of nine infectives causes a secondary case. Figure 5.6(a) shows the temporal function

$$\exp\left(\hat{\beta}_{\text{trend}}\frac{t}{365} + \hat{\beta}_{\sin}\sin\left(t\,\frac{2\pi}{365}\right) + \hat{\beta}_{\cos}\cos\left(t\,\frac{2\pi}{365}\right)\right)$$

which multiplies the global endemic baseline rate $\exp(\hat{\beta}_0) \approx 1.57 \cdot 10^{-09}$ (95%-CI: $[1.25 \cdot 10^{-09}, 1.98 \cdot 10^{-09}]$). For comparison, the time-of-year effect of the serogroup B finetype is also shown. Note that the graphic ignores the discretisation of time on a weekly grid used for model estimation. Besides the temporally decreasing endemic risk for the serogroup C finetype, the shapes of the sine-cosine waves are similar between the finetypes. The fitted spatial interaction functions of both finetypes are plotted in Figure 5.6(b) as function of the distance from the infective individual. The plot indicates no substantial differences between the finetypes, but the 95%-confidence interval for the serogroup C finetype is wider. This is a consequence of the fewer information attributed to the epidemic component because of the small numbers of observed potential sources of infection shown in Figure 5.4(a). However, the smaller basic reproduction number of the serogroup C finetype is due to smaller estimates of the epidemic intercept $\hat{\gamma}_0 \approx 2.06 \cdot 10^{-06}$ (95%-CI: $[1.12 \cdot 10^{-06}, 3.79 \cdot 10^{-06}]$) and the standard deviation $\hat{\sigma} = \exp(\hat{\varphi}) \approx 17.1$ (95%-CI: $[12.7, 22.8]$) of the spatial kernel.



(a) Temporal effects for both finetypes.

(b) Spatial kernels for both finetypes as a function of the distance $\|\boldsymbol{s}-\boldsymbol{s}_j\|$ from the infective individual.

Figure 5.6: Fitted temporal effects and isotropic spatial interaction functions $f(\boldsymbol{s}) \equiv f(\|\boldsymbol{s}\|)$ for both finetypes from the models in Tables 5.4 (for B) and 5.8 (for C).

Estimating different infectivities depending on the age or gender of the patient did not reveal significant effects. In particular, female infectives of the serogroup C finetype do not have a significantly higher potential of causing secondary cases as was the case for the point pattern of the serogroup B finetype. The results are shown in Tables C.6, C.7, and C.8 in Appendix C.

5.3 Joint Modelling of Both Finetypes

Up to now the point patterns of the two finetypes have been analysed independently of each other. Although visual and heuristic comparisons of the estimated statistics between the finetypes are possible, the differences can as yet not be assessed statistically. For instance we would like to test whether the weight of the epidemic component and thus the basic reproduction number is significantly higher for the finetype B:P1.7-2,4:F1-5. Similarly, testing for a different global endemic baseline risk $\exp(\beta_0)$ is approached.

For this purpose, the `twinstim` class has been extended in Section 4.1.3 in order to model a univariate discrete mark, which here corresponds to the finetype of the IMD case. As initial model for the complete point pattern of 636 cases of IMD, I consider the CIF

$$\lambda^*_{\boldsymbol{\theta}}(t,\boldsymbol{s},\kappa) = \rho_{\xi(\boldsymbol{s})} \cdot \exp\left(\beta_{0,\kappa} + \beta_{\text{trend}}\frac{\lfloor t\rfloor}{365} + \beta_{\sin}\sin\left(\lfloor t\rfloor\cdot\frac{2\pi}{365}\right) + \beta_{\cos}\cos\left(\lfloor t\rfloor\cdot\frac{2\pi}{365}\right) + \sum_{l=0}^{1}\beta_{\text{influenza},l}\, z^{(l)}_{\tau(t),\xi(\boldsymbol{s})}\right) + \sum_{j\in I^*(t,\boldsymbol{s},\kappa)}\exp(\gamma_{0,\kappa}) \tag{5.2}$$

assuming a type-specific endemic log-baseline risk $\beta_{0,\kappa}$ and a type-specific log-rate $\gamma_{0,\kappa}$ of transmission, which is homogeneous in space-time and limited to the same finetype by setting $\boldsymbol{Q} = I_2$. Thus, at a specific point in time and location in space, the intensity for an event of type κ is determined by $\beta_{0,\kappa}$, $\gamma_{0,\kappa}$, and the current number of infectives of type κ. Note that the model imposes the same effect of time and of the numbers of influenza cases for both finetypes, which is a questionable assumption. In the previous analyses, the serogroup C finetype had a more pronounced and significant decreasing time trend, whereas the decrease was not significant for the serogroup B finetype. Table 5.9 presents the estimation results of the initial model. The computation took only 3.7 minutes because an inhomogeneous spatial transmission kernel is not yet included.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`typeB`	−20.332130	0.107217	−189.636	$< 2\cdot10^{-16}$
`typeC`	−20.392231	0.102255	−199.424	$< 2\cdot10^{-16}$
`I(start/365)`	−0.043354	0.022577	−1.920	0.0548
`sin(start*2*pi/365)`	0.308931	0.068179	4.531	$5.86\cdot10^{-06}$
`cos(start*2*pi/365)`	0.270002	0.066548	4.057	$4.97\cdot10^{-05}$
`influenza0`	0.008077	0.010299	0.784	0.4329
`influenza1`	−0.013046	0.015254	−0.855	0.3924
`typeB`	−15.9722	0.2515	−63.51	$< 2\cdot10^{-16}$
`typeC`	−16.4259	0.3020	−54.39	$< 2\cdot10^{-16}$
AIC:	19175			
Log-likelihood:	−9578			

Table 5.9: Results of the initial joint model of both point patterns.

As could already be expected from the previous analyses, the effects of influenza are not significant. Removing them subsequently from the model improves the AIC and the corresponding likelihood ratio tests also do not reject the hypotheses $\beta_{\text{influenza},0} = 0$ ($p = 0.4711$) and $\beta_{\text{influenza},1} = 0$ ($p = 0.3126$). The estimated model without influenza effects is shown in Table 5.10.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`typeB`	−20.33038	0.10658	−190.758	$< 2 \cdot 10^{-16}$
`typeC`	−20.39042	0.10223	−199.465	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.04472	0.02249	−1.989	0.0467
`sin(start*2*pi/365)`	0.30023	0.06541	4.590	$4.43 \cdot 10^{-06}$
`cos(start*2*pi/365)`	0.27076	0.06632	4.082	$4.46 \cdot 10^{-05}$
`typeB`	−15.9721	0.2502	−63.83	$< 2 \cdot 10^{-16}$
`typeC`	−16.4317	0.3034	−54.16	$< 2 \cdot 10^{-16}$
AIC:	19172			
Log-likelihood:	−9579			

Table 5.10: Dropped influenza effects from the initial model.

The effect of the time trend represents a trade off between both finetypes and is significant on the 5% level. It corresponds to a yearly decrease of the endemic risk by about 4.4% (95%-CI: $[0.065, 8.497]$%). The implied basic reproduction numbers are 0.25 (B) and 0.15 (C) with corresponding 95% Wald confidence intervals $[0.15, 0.40]$ (B) and $[0.086, 0.281]$ (C), respectively. These estimates are similar to the initial separate models for the finetypes. Likelihood ratio tests of the hypotheses of common endemic and epidemic intercepts can be performed by fitting the corresponding model under the null hypothesis. This yields a p-value of 0.5538 for the null hypothesis $\beta_{0,1} = \beta_{0,2}$, and a p-value of 0.2392 for the null hypothesis $\gamma_{0,1} = \gamma_{0,2}$. Hence, there is no statistical evidence for any type-specific effects in this model with assumed constant interaction functions.

This assumption is now dropped for the spatial kernel, which is additionally modelled with type-specific standard deviations. However, as a consequence of the experiences of the previous sections, no further attempts are made to model an inhomogeneous temporal interaction function. Table 5.11 shows the estimation results for the model with a type-specific Gaussian kernel. Unfortunately, maximisation did not converge up to required standard accuracy. However, the AIC decreased importantly compared to the model with homogeneous spatial transmission.

Supported by the corresponding likelihood ratio test of the previous model a common endemic intercept is estimated next. Log-likelihood maximisation did again not converge, neither when additionally assuming $\varphi_1 = \varphi_2$. Convergence was only reached when assuming both intercepts to be type-invariant, but permitting type-specific Gaussian kernels. Estimation results for this model are shown in Table 5.12. Because other models were not identifiable, likelihood ratio

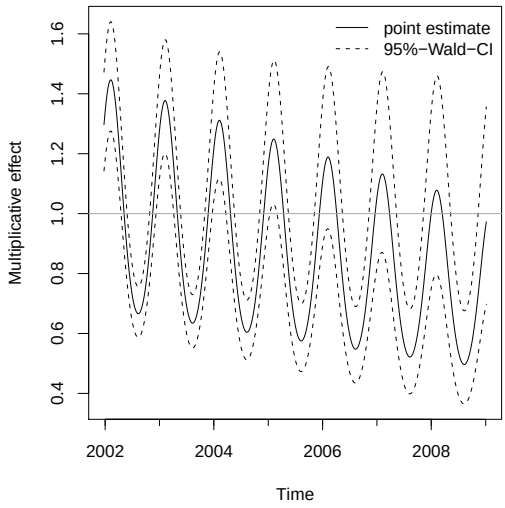
	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`typeB`	−20.40179	0.10350	−197.126	$< 2 \cdot 10^{-16}$
`typeC`	−20.37031	0.09827	−207.284	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.04678	0.02259	−2.070	0.0384
`sin(start*2*pi/365)`	0.27554	0.06574	4.191	$2.77 \cdot 10^{-05}$
`cos(start*2*pi/365)`	0.25834	0.06529	3.957	$7.60 \cdot 10^{-05}$
`typeB`	−12.42354	0.19100	−65.05	$< 2 \cdot 10^{-16}$
`typeC`	−13.45309	0.29437	−45.70	$< 2 \cdot 10^{-16}$
`siaf.1`	2.91644	0.09902	29.45	$< 2 \cdot 10^{-16}$
`siaf.2`	3.07297	0.15116	20.33	$< 2 \cdot 10^{-16}$
AIC:	19010			
Log-likelihood:	−9496			

Table 5.11: Result of the joint model with a type-specific Gaussian kernel.

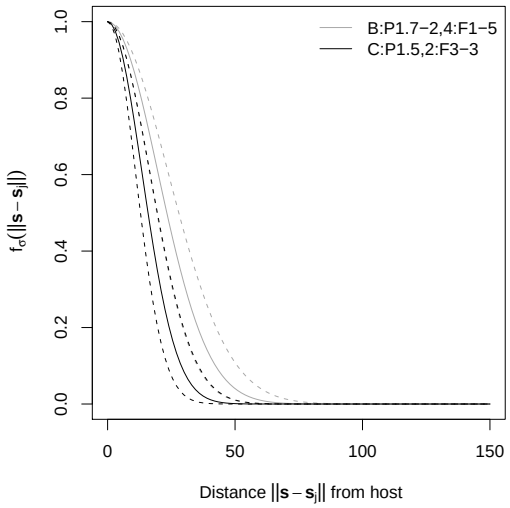
	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.37333	0.08769	−232.325	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.04901	0.02231	−2.197	0.0281
`sin(start*2*pi/365)`	0.27104	0.06502	4.169	$3.06 \cdot 10^{-05}$
`cos(start*2*pi/365)`	0.26007	0.06475	4.016	$5.91 \cdot 10^{-05}$
`(Intercept)`	−12.57470	0.16407	−76.64	$< 2 \cdot 10^{-16}$
`siaf.1`	2.98210	0.09383	31.78	$< 2 \cdot 10^{-16}$
`siaf.2`	2.60018	0.10577	24.58	$< 2 \cdot 10^{-16}$
AIC:	19008			
Log-likelihood:	−9497			

Table 5.12: Result of the joint model with type-invariant intercepts and a type-specific Gaussian kernel.

tests are not applicable. As also indicated by the separate model fits of the previous sections, this joint model suggests a wider kernel for the serogroup B finetype ($\hat{\sigma}_1 = 19.7$, 95%-CI: $[16.4, 23.7]$) than for the serogroup C finetype ($\hat{\sigma}_2 = 13.5$, 95%-CI: $[10.9, 16.6]$). Figure 5.7(b) shows the type-specific spatial interaction functions. They differ from the respective curves of the separate model estimation because of the imposed common endemic effects and the imposed common epidemic intercept. The common temporal decrease averaged over both finetypes is shown in Figure 5.7(a).



(a) Common temporal effect of the finetypes.



(b) Spatial kernels for both finetypes as a function of the distance $\|\boldsymbol{s}-\boldsymbol{s}_j\|$ from the infective individual.

Figure 5.7: Imposed common temporal effect and isotropic spatial interaction functions $f(\boldsymbol{s}|\kappa) \equiv f(\|\boldsymbol{s}\||\kappa)$ from the model in Table 5.12.

6 Discussion and Outlook

This Master's Thesis is a contribution towards a regression framework for self-exciting spatio-temporal point processes modelled through their conditional intensity function. The particular application of interest was infectious disease epidemiology, specifically, the stochastic modelling of the spreading dynamics of two different strains of meningococci.

For this purpose, basic characteristics of invasive meningococcal disease were considered. The corresponding marked spatio-temporal point patterns observed in Germany during the years 2002–2008 were described as well as their potential relation to waves of influenza was investigated. Before considering any stochastic models, I provided a short introduction into the basics of the necessary theory of point processes. As a prerequisite for the evaluation of the likelihood for complex spatial point process models for non-standard observation regions, two-dimensional numerical integration over polygonal domains was investigated in detail. A brief overview of existing approaches to analysing spatio-temporal point process data was provided.

In the main part of this thesis – motivated by the available high resolution IMD data – a conditional intensity model for self-exciting point processes in continuous space-time with an optional discrete mark was proposed. This model called `twinstim` was inspired on the one hand by the additive-multiplicative modelling for self-exciting temporal point processes in discrete space proposed by Höhle (2009a), and on the other hand by the spatio-temporal ETAS model for the occurences of earthquakes. Model estimation by means of maximum likelihood was derived and thoroughly implemented in `R`.

The proposed model was finally applied to the observed point patterns of the two IMD finetypes, first separately, then jointly using the time-space-mark formulation of `twinstim`. Estimation results indicated a time decreasing endemic risk of infection with finetype C:P1.5,2:F3-3, and a higher infectivity of female infectives of finetype B:P1.7-2,4:F1-5. The considered age effects on infectivity (linear and infant indicator) were not significant for any of the finetypes. A sine-cosine wave fitted time-of-year effects on the endemic baseline risk while there was no evidence for a superimposed effect of the numbers of influenza cases or its lagged versions. An exponential decay of infectivity during the infectious period could not be identified from the data. In contrast, decreasing infectivity in space according to an isotropic Gaussian kernel around an infective's location was estimated to have a standard deviation of 18.5 km

(95%-CI: [15.3, 22.3]) for finetype B:P1.7-2,4:F1-5, and of 17.1 km (95%-CI: [12.7, 22.8]) for finetype C:P1.5,2:F3-3. Different estimated weights of the infectious nature of the finetypes yielded basic reproduction numbers of 0.27 (B) and 0.11 (C), respectively, which attributes a higher degree of clustering to the finetype B:P1.7-2,4:F1-5. The joint modelling of both point patterns using an event type mark was intended to permit comparative likelihood ratio tests on type-specific effects. However, the maximisation procedure did not converge for most of the joint models with a spatially inhomogeneous triggering function. Hence the desired tests could not be performed.

The application showed that the proposed model and its possible estimation by the provided `R` implementation are applicable and valuable for the analysis of spatio-temporal infectious disease data. The basic reproduction number, accessible through the estimated triggering function, is an important characteristic of an infectious disease. By the `twinstim` formulation its estimation is adjusted for endemic effects like the population density and localised waves of influenza. Another attractive option provided by `twinstim` (but not used in this thesis) is its ability to estimate, for each observed event, the probability that that event is an initiating event (i.e. not the offspring of some earlier event). Specifically, this probability equals the proportion of the endemic component in the value of the observed conditional intensity at the event in question.

The current analysis pointed out that the epidemic component should be modelled in the philosophy of KIS (keep it simple), because this component receives less information from the data, which did e.g. not permit identification of a time-dependent triggering function. Furthermore, a simple but well-identified model is often more attractive than a model containing more parameters than can be reasonably interpreted and estimated. However, the sine-cosine wave provides no causative interpretation of the modelled seasonality, which is why the endemic part of the model is open for further modelling. Spatio-temporal environmental covariates that potentially explain the heterogeneity of the IMD cases could be investigated for inclusion in the endemic component. Moreover, flexible semiparametric modelling of the temporal baseline would be attractive, because the parametric specifications of the temporal trend and the seasonality impose a limited shape of the temporal effect as seen in Figure 5.6(a).

Common problems of the analysis of point patterns include *tied observations* and *edge effects*. This thesis did only consider a single artificial permutation of tied event times for the analysis (cf. Subsection 1.2.4). More accurate analyses must carry out sensitivity analyses on the specific choices made with respect to breaking ties (e.g. use different randomisations). Edge (or boundary) effects may lead to seriously biased parameter estimates obtained via likelihood maximisation. Even Vere-Jones (2009) does "not know of any simple way of dealing with these". According to him, the simplest workaround is the definition of a buffer zone around

an inner observation region in both time (preceding) and space (surrounding), and build the likelihood just based on the data points lying within the inner observation region. Points of the buffer zone then provide information needed to compute the conditional intensity within the inner observation region. However, a sufficient amount of observations is needed to afford a genuine buffer zone. This requirement does especially not hold for the IMD data, although the performed estimation certainly suffers from edge effects because many IMD cases are at the border of the spatial observation region Germany. In this context Elias et al. (2010) elucidate, whether an increase of disease activity in the region of Aachen (Germany) represented local emergence or cross-border spread from the Netherlands. Hence, the actual disease clusters are wider than observed with a restriction to Germany, which potentially causes underestimation of the epidemic weight in `twinstim`. Another form of potential bias is also related to data sampling: location is given as centroid of the postcode boundary of the patient's residence, which may not always appropriatly reflect the area of the social contact network. However, an operationalisation of the social contact network is hardly accessible.

Other spatial interaction functions than the isotropic Gaussian kernel could be considered. For example in the context of modelling dynamics of forest fires, an anisotropic kernel accounting for the main wind direction could be supposed. The characteristic function of the maximum spatial interaction range $[0, \delta]$ could then take some form of Mahalanobis distance if applicable. The temporal interaction function $g(t)$ might obey as alternative parametric form a trapezoidal shape, where infectivity first increases to 1 and then slowly decreases back to 0 (cf. Lawson & Leimich, 2000, Figure 3). Both interaction functions could also be modelled as semiparametric zero-degree B-splines (step functions), which on the one hand simplifies integration but on the other hand poses the questions of the choice of knots and parameter penalisation for sufficient smoothness.

However, full space-time variation of the conditional intensity function in `twinstim`'s can actually only result from the interaction functions of the epidemic component, because in the endemic component I only allowed for a piecewise constant spatio-temporal log-baseline. Otherwise, the implementation of the integrated CIF in the log-likelihood and its derivatives with respect to the parameters of a continuous parametric trend in the score function would have been much more cumbersome. Although this would be a possible extension of the model, the results of the IMD analysis were probably not affected crucially by the discretisation of the temporal baseline on a weekly grid with regard to a total observation length of 367 weeks.

Note that in this thesis, the interaction functions were chosen to be normalised to take a maximum value of 1 rather than integrating to 1 over $\mathbb{R}^2$ or $\mathbb{R}_+$, respectively. In my opinion using densities is not advantageous since the implied unit integral over $\mathbb{R}^2$ or $\mathbb{R}_+$, respectively, does not appear in the estimation formulae because integration is always limited to the domains $(0, T]$ and W. In contrast, requiring the interaction function to have a maximum

value of 1 (usually at their origin) simplifies the interpretation of $\boldsymbol{\gamma}$ as the unique weighting factor. Presumably this also improves parameter identifiability.

Theoretically, the values ε and δ of maximum interaction ranges could also be individual-specific marks like age and gender. In particular, ε_i would be the length of the individual's infectious period. A full point process model would then assume a distribution for the length of the infectious period (e.g. exponential), the density of which multiplies the CIF as in equation 2.7. Although the implementation allows for observed individual specific values of ε and δ, this approach is generally not applicable for infectious disease data, because lengths of infectious periods and spatial ranges of infectivity are hardly observable.

In the marked `twinstim` of Section 4.1.3, the triggering function $e_j(t, \boldsymbol{s})$ did actually not account for the type κ of the *triggered* event, but only for the types κ_j of the *triggering* events. Similar to multitype epidemics in the context of the compartmental SIR model in Andersson & Britton (2000, Chapter 6), an extension would be to also model different intensities for the mutual transmission between the event types through a generalised transmission matrix $\boldsymbol{Q} = (q_{k,l})$, $q_{k,l} \in \mathbb{R}$. The original epidemic intercepts γ_{0,κ_j} would be replaced by $q_{\kappa_j,\kappa}$, i.e. the log-baseline intensity of transmission from type κ_j to κ, and γ_{0,κ_j} then equals $\sum_{\kappa \in \mathcal{K}} q_{\kappa_j,\kappa}$.

Besides type-specific transmission rates, the model may be extended by tile-specific (county-specific) endemic effects. In the current formulation the deterministic offset for the population density substitutes such effects, which could be modelled stochastically by introducing frailties like in extended Cox models. Spatial random effects following a Gaussian Markov random field which assumes similar effects of neighboring tiles are a typical tool (Rue & Held, 2005).

Moving into the direction of random effects leads to different strategies of estimation. Specifically, a Bayesian framework could be aspired if the `twinstim` class is extended to incorporate frailties, or if the temporal baseline or the interaction functions are modelled as zero-degree B-Splines with a smoothness penalty. The corresponding penalised likelihood inference can already be regarded as an *empirical Bayesian approach*, thus an extension to full Bayesian inference would not be far. On the one hand, this would also have the advantages that inference must not rely on asymptotic properties (e.g. the estimation of parameter uncertainty by the estimated expected Fisher information matrix), that inference accounts for uncertainty of hyperparameters (e.g. by assuming probability distributions for ε and δ), and that credibility intervals for transformed statistics like the basic reproduction number are directly available from the MCMC output. On the other hand, careful tuning and diagnosis is always required to assure convergence of the MCMC samples to a stationary distribution.

Other variants of estimation are Monte Carlo maximum likelihood known from purely spatial point processes (see Møller & Waagepetersen, 2004, Chapter 8) or the Expectation-Maximisation (EM) algorithm. For the former, Diggle (2007) states that "in practice these

methods often need careful tuning to each application and the associated cost of developing and running reliable code can be an obstacle to their routine use". The use of the EM algorithm for spatio-temporal branching processes was explored by Veen & Schoenberg (2006), who considered the space-time ETAS model for exemplification. While observing convergence problems of numerical log-likelihood maximisation, they found the EM estimation to be "extremely robust and accurate".

Perhaps the most important directions for future work on the `twinstim` class are the implementation of a simulation algorithm and investigations of residual processes for assessing the goodness-of-fit proposed by Ogata (1988) for temporal point processes. Simulation of the marked spatio-temporal `twinstim` is possible through the temporal ground process with CIF $\lambda_g^*(t)$ using Ogata's modified thinning algorithm as e.g. described in Daley & Vere-Jones (2003, Algorithm 7.5.V.). If the temporal interaction function is constant, then no rejection sampling step is required. The hyperparameters ε and δ must either be given as deterministic values or could be drawn from a specified distribution. If the triggering function of the epidemic component depends on unpredictable marks, an additional sampling scheme must be specified for these: either a distribution for random generation or e.g. sampling from the set of observed marks as in the bootstrap.

Bibliography

Abramowitz, M. & Stegun, I. A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (9th ed.; M. Abramowitz & I. A. Stegun, Eds.). New York: Dover Publications.

Andersson, H. & Britton, T. (2000). *Stochastic epidemic models and their statistical analysis* (Vol. 151). New York: Springer-Verlag.

Baddeley, A. & Turner, R. (2005, 26th January). `Spatstat`: an `R` package for analyzing spatial point patterns. *Journal of Statistical Software*, *12*(6), 1–42. Available from `http://www.jstatsoft.org/v12/i06`

Baddeley, A. & Turner, T. R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian & New Zealand Journal of Statistics*, *42*(3), 283–322.

Bengtsson, H. (2005, February). *`R` Coding Conventions - a first draft.* Retrieved 3 July 2009, from `http://www1.maths.lth.se/help/R/RCC/` (Version 0.8)

Berman, M. & Turner, T. R. (1992). Approximating point process likelihoods with GLIM. *Applied Statistics*, *41*(1), 31–38.

Breitenacher, M., Obermeier, M. & Schomaker, M. (2004, October). *Zeitlicher Zusammenhang zwischen Influenza- und Meningokokkenausbrüchen.* Available from `http://www.rappenantilope.de/Dokumente/Endberichtpdf.pdf`

Daley, D. J. & Vere-Jones, D. (2003). *An introduction to the theory of point processes* (2nd ed., Vol. I: Elementary Theory and Methods; J. M. Gani, C. C. Heyde & T. G. Kurtz, Eds.). New York: Springer-Verlag.

Daley, D. J. & Vere-Jones, D. (2008). *An introduction to the theory of point processes* (2nd ed., Vol. II: General Theory and Structure; J. M. Gani, C. C. Heyde & T. G. Kurtz, Eds.). New York: Springer-Verlag.

Dargatz, C., Georgescu, V. & Held, L. (2005, September). *Stochastic modelling of the spatial spread of influenza in germany* (Tech. Rep. No. 450). München: Department of Statistics, Collaborative Research Center (SFB) 386, Ludwig-Maximilians-Universität. Available from `http://epub.ub.uni-muenchen.de/1819/`

Davis, P. J. & Rabinowitz, P. (1984). *Methods of numerical integration* (2nd ed.). New York: Academic Press.

DiDonato, A. R. & Hageman, R. K. (1982). A method for computing the integral of the bivariate normal distribution over an arbitrary polygon. *SIAM Journal on Scientific and Statistical Computing*, *3*(4), 434–446. Available from `http://link.aip.org/link/?SCE/3/434/1`

Diggle, P. J. (2006). Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Statistical Methods in Medical Research*, *15*(4), 325–336.

Diggle, P. J. (2007). Spatio-temporal point processes: Methods and applications. In B. Finkenstädt, L. Held & V. Isham (Eds.), *Statistical methods for spatio-temporal systems* (pp. 1–45). Boca Raton: Chapman & Hall/CRC.

Diggle, P. J., Kaimi, I. & Abellana, R. (2009, August). Partial-likelihood analysis of spatio-temporal point-process data. *Biometrics*. (In press, published online)

Elias, J., Harmsen, D., Claus, H., Hellenbrand, W., Frosch, M. & Vogel, U. (2006, November). Spatiotemporal analysis of invasive meningococcal disease, Germany. *Emerging Infectious Diseases*, *12*(11), 1689–1695. Available from `http://www.cdc.gov/ncidod/EID/vol12no11/06-0682.htm`

Elias, J., Schouls, L. M., Pol, I. van de, Keijzers, W. C., Martin, D. R., Glennie, A. et al. (2010). Slow migration of meningococcal clone belonging to ST-41/44 complex facilitates targeting by vaccination. *Emerging Infectious Diseases*. (Accepted for publication)

Environmental Systems Research Institute. (1998, July). *ESRI shapefile technical description.* Available from `http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf`

Fahrmeir, L., Kneib, T. & Lang, S. (2007). *Regression: Modelle, Methoden und Anwendungen*. Berlin Heidelberg: Springer-Verlag.

Falcon, S. (2009). `weaver`: Tools and extensions for processing Sweave documents [Computer software manual]. Available from `http://www.bioconductor.org/packages/2.4/bioc/html/weaver.html` (R package version 1.10.0)

Fox, P. A., Hall, A. P. & Schryer, N. L. (1978, June). The PORT Mathematical Subroutine Library. *ACM Transactions on Mathematical Software*, *4*(2), 104–126. Available from `http://www.bell-labs.com/project/PORT/`

Fuest, A. (2009). *Modelling temporal dependencies of extreme events via point processes*. Unpublished master's thesis, Department of Statistics, Ludwig-Maximilians-Universität, München.

GENESIS-Online. (2009). *12411-0005: Bevölkerung: Deutschland, Stichtag, Altersjahre.* Wiesbaden, Germany: Statistisches Bundesamt (DESTATIS). Retrieved 23 November 2009, from `https://www-genesis.destatis.de/genesis/online`

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F. et al. (2009). `mvtnorm`: Multivariate normal and t distributions [Computer software manual]. Available from `http://CRAN.R-project.org/package=mvtnorm` (R package version 0.9-7)

Genz, A. C. & Malik, A. A. (1980, December). An adaptive algorithm for numerical integration over an N-dimensional rectangular region. *Journal of Computational and Applied Mathematics*, *6*(4), 295–302.

German National Reference Centre for Meningococci. (2009). *Data of the German Reference Centre for Meningococci for the year 2008.* Würzburg. Retrieved 15 December 2009, from `http://www.meningococcus.uni-wuerzburg.de/startseite/berichte/data_2008_eng/` (Version 1.4)

Gómez-Rubio, V., Ferrándiz-Ferragud, J. & Lopez-Quílez, A. (2005). Detecting clusters of disease with `R`. *Journal of Geographical Systems*, *7*(2), 189–206.

Hawkes, A. G. (1971, April). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90.

Hawkes, A. G. & Oakes, D. (1974, September). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, *11*(3), 493–503. Available from `http://www.jstor.org/stable/3212693`

Held, M. (2001, August). FIST: Fast industrial-strength triangulation of polygons. *Algorithmica*, *30*(4), 563–596.

Hellenbrand, W., Vogel, U. & Elias, J. (2008, August). Invasive Meningokokken-Erkrankungen im Jahr 2007. *Epidemiologisches Bulletin*(32), 265–272.

Hengl, T., Loon, E., Sierdsema, H. & Bouten, W. (2008). Advancing spatio-temporal analysis of ecological data: Examples in `R`. In *ICCSA '08: Proceedings of the international conference on computational science and its applications, part I* (Vol. 5072/2008, pp. 692–707). Berlin, Heidelberg: Springer-Verlag.

Höhle, M. (2007). `surveillance`: An `R` package for the surveillance of infectious diseases. *Computational Statistics*, *22*(4), 571–582.

Höhle, M. (2009a, December). Additive-multiplicative regression models for spatio-temporal epidemics. *Biometrical Journal*, *51*(6), 961–978.

Höhle, M. (2009b). *Statistical modeling and monitoring of infectious diseases*. Habilitation thesis, Faculty of Mathematics, Informatics and Statistics, Ludwig-Maximilians-Universität, München.

Höhle, M. & Feldmann, U. (2007, October). `RLadyBug` — an `R` package for stochastic epidemic models. *Computational Statistics & Data Analysis*, *52*(2), 680–686.

Höhle, M., Feldmann, U. & Meyer, S. (2009). `RLadyBug`: Analysis of infectious diseases using stochastic epidemic models [Computer software manual]. Available from `http://cran.r-project.org/package=RLadyBug` (R package version 0.6-0)

Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Chichester: Wiley.

ISO. (2004). *ISO 8601:2004 - data elements and interchange formats — information interchange — representation of dates and times.* Geneva, Switzerland: International Organization for Standardization.

Jefferson, T. (2004, September). How to deal with influenza? *BMJ*, *329*(7467), 633–634.

Jensen, E. S., Lundbye-Christensen, S., Samuelsson, S., Sørensen, H. T. & Schønheyder, H. C. (2004, February). A 20-year ecological study of the temporal association between influenza and meningococcal. *European Journal of Epidemiology*, *19*(2), 181–187.

Keitt, T. H., Bivand, R., Pebesma, E. & Rowlingson, B. (2009). `rgdal`: Bindings for the Geospatial Data Abstraction Library [Computer software manual]. Available from `http://CRAN.R-project.org/package=rgdal` (R package version 0.6-20)

Krommer, A. R. & Ueberhuber, C. W. (1998). *Computational integration*. Philadelphia: SIAM.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods*(6), 1481–1496.

Lawson, A. B. & Leimich, P. (2000). Approaches to the space-time modelling of infectious disease behaviour. *IMA Journal of Mathematics Applied in Medicine and Biology*, *17*(1), 1–13.

Leisch, F. (2002). `Sweave`: Dynamic generation of statistical reports using literate data analysis. In W. Härdle & B. Rönz (Eds.), *Compstat 2002 — Proceedings in computational statistics* (pp. 575–580). Physica Verlag, Heidelberg. Available from `http://www.stat.uni-muenchen.de/~leisch/Sweave`

Martinussen, T. & Scheike, T. H. (2002, June). A flexible additive multiplicative hazard model. *Biometrika*, *89*(2), 283–298.

Meyer, S. & Wimmer, V. (2009, March). *Räumlich-zeitliche Analyse der FSME-Infektionen in Baden-Württemberg.* (Consulting project, Department of Statistics, Ludwig-Maximilians-Universität München)

Møller, J. & Waagepetersen, R. P. (2004). *Statistical inference and simulation for spatial point processes.* Chapman & Hall/CRC.

Murta, A. (2009). *GPC: General Polygon Clipper library.* Advanced Interfaces Group, Department of Computer Science, University of Manchester. Available from `http://www.cs.manchester.ac.uk/~toby/alan/software/`

Møller, J. & Rasmussen, J. G. (2005, September). Perfect simulation of hawkes processes. *Advances in Applied Probability*, *37*(3), 629–646.

Nelder, J. A. & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, *7*(4), 308–313.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, *30*(2), 243–261.

Ogata, Y. (1988, March). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27. Available from `http://www.jstor.org/stable/2288914`

Ogata, Y. (1998, June). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402.

Ogata, Y. (1999, August). Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics*, *155*(2-4), 471–507.

Ogata, Y., Katsura, K. & Tanemura, M. (2003, October). Modelling heterogeneous space-time occurrences of earthquakes and its residual analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *52*(4), 499–509.

Pebesma, E. J. & Bivand, R. S. (2005, November). Classes and methods for spatial data in R. *R News*, *5*(2), 9–13. Available from `http://CRAN.R-project.org/doc/Rnews/`

Peng, R. D. (2009). `gpclib`: General Polygon Clipping library for `R` [Computer software manual]. Available from `http://CRAN.R-project.org/package=gpclib` (R package version 1.4-4)

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2007). *Numerical recipes: The art of scientific computing* (3rd ed.). Cambridge, New York: Cambridge University Press.

R Development Core Team. (2009a). `R`: A language and environment for statistical computing [Computer software manual]. Wien, Austria. Available from `http://www.R-project.org`

R Development Core Team. (2009b). Writing `R` Extensions (2.9.1 (2009-06-26) ed.) [Computer software manual]. Wien, Austria.

Rathbun, S. L. (1996). Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *Journal of Statistical Planning and Inference*, *51*(1), 55–74.

Rathbun, S. L. & Cressie, N. (1994, March). Asymptotic properties of estimators for the parameters of spatial inhomogeneous poisson point processes. *Advances in Applied Probability*, *26*(1), 122–154.

Robert Koch-Institut. (2009). *SurvStat@RKI.* Berlin, Germany. Retrieved 4 November 2009, from `http://www3.rki.de/SurvStat`

Rosenstein, N. E., Perkins, B. A., Stephens, D. S., Popovic, T. & Hughes, J. M. (2001). Meningococcal Disease. *The New England Journal of Medicine*, *344*(18), 1378–1388.

Rowlingson, B., Diggle, P. et al. (2009). `splancs`: Spatial and space-time point pattern analysis [Computer software manual]. Available from `http://CRAN.R-project.org/package=splancs` (R package version 2.01-25)

Rue, H. & Held, L. (2005). *Gaussian Markov random fields: Theory and applications* (Vol. 104). London: Chapman & Hall/CRC.

Schoenberg, F. P., Brillinger, D. R. & Guttorp, P. M. (2002). Point processes, spatial-temporal. In A. H. El-Shaarawi & W. W. Piegorsch (Eds.), *Encyclopedia of Environmetrics* (Vol. 3, pp. 1573–1577). New York: John Wiley & Sons.

Schrauder, A., Claus, H., Elias, J., Vogel, U., Haas, W. & Hellenbrand, W. (2007, May). Capture-recapture analysis to estimate the incidence of invasive meningococcal disease in Germany, 2003. *Epidemiology and Infection*, *135*(4), 657–664.

Smyth, G. (2009). `statmod`: Statistical modeling [Computer software manual]. Available from `http://CRAN.R-project.org/package=statmod` (R package version 1.4.1)

Sommariva, A. & Vianello, M. (2007a, March). *Polygauss: MATLAB code for Gauss-like cubature over polygons.* Retrieved 7 October 2009, from `http://www.math.unipd.it/~alvise/software.html`

Sommariva, A. & Vianello, M. (2007b, June). Product Gauss cubature over polygons based on Green's integration formula. *Bit Numerical Mathematics*, *47*(2), 441-453(13).

Statistisches Bundesamt (DESTATIS). (2009, February). *Gemeindeverzeichnis GV 2000* [CD]. Wiesbaden, Germany: Statistische Ämter des Bundes und der Länder. (Districts as of 31/12/2008. Data as of 31/12/2007.)

Thompson, M. J., Ninis, N., Perera, R., Mayon-White, R., Phillips, C., Bailey, L. et al. (2006). Clinical recognition of meningococcal disease in children and adolescents. *Lancet*, *367*(9508), 397–403.

Veen, A. & Schoenberg, F. P. (2006, June). *Estimation of space-time branching process models in seismology using an EM-type algorithm* (Tech. Rep.). UC Los Angeles (UCLA): Department of Statistics. Available from `http://escholarship.org/uc/item/7zm0v7tg`

Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.

Vere-Jones, D. (2009, June). Some models and procedures for space-time point processes. *Environmental and Ecological Statistics*, *16*(2), 173–195.

Wikipedia. (2009, December). *Nomenclature of territorial units for statistics — Wikipedia, the free encyclopedia.* Retrieved 7 December 2009, from `http://en.wikipedia.org/w/index.php?title=Nomenclature_of_Territorial_Units_for_Statistics&oldid=327378537`

Zelen, M. & Severo, N. C. (1970). Probability functions. In M. Abramowitz & I. A. Stegun (Eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (9th ed., chap. 26). New York: Dover Publications.

Appendices

A Contents of the Enclosed CD

The results of this Master's Thesis can be reproduced with the material on the enclosed CD. It contains the following files:

`CITATION` File on how to cite this Master's Thesis: BibTeX entry, plain text.

`LICENCE` File stating usage permissions and restrictions.

`data` Data folder.

`graphics` Directory containing the presented figures.

`literature` Directory with some of the literature referenced to.

`mydata` Collection of cached data objects and model fits.

`R` Directory with `R` source code.
The following components might be especially interesting for future research:

- The file `polygon-integration.R` contains the functions `polyint.*` of Section 3.2 implementing the various cubature rules over polygons.
- The subdirectory `epidemic` contains the main code for the analysis in Chapter 5 (which was briefly described in Section 4.3), and might soon be integrated into a contributed `R` package of the same name. The file `epidata.R` provides the converter `as.epidata` to generate an object of the `S3` class `"epidata"`. The file `twinstim.R` contains the partial and full log-likelihood (cf. Section 4.2.1), the score function (cf. Section 4.2.2), the estimation of the fisher information matrix (cf. Section 4.2.3), as well as the interfacing function `twinstim`. Some typical `S3` methods for model fits are implemented in the file `twinstim-methods.R`.

`thesis` Directory containing the LaTeX and `Sweave` sources.

With GNU `make`, `R` (and necessary packages) and pdfLaTeX installed, one can compile this Master's Thesis using the `Makefile` in the root directory by `make Thesis.pdf`.

B Model Selection Steps for Finetype B:P1.7-2,4:F1-5

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	-20.401054	0.135303	-150.781	$< 2 \cdot 10^{-16}$
`I(start/365)`	-0.028137	0.032795	-0.858	0.39092
`sin(start*2*pi/365)`	0.275318	0.099357	2.771	0.00559
`cos(start*2*pi/365)`	0.374639	0.095277	3.932	$8.42 \cdot 10^{-05}$
`influenza0`	0.015578	0.011083	1.406	0.15984
`influenza1`	-0.015540	0.019406	-0.801	0.42326
`influenza3`	-0.004647	0.024891	-0.187	0.85189
`(Intercept)`	-15.9547	0.2484	-64.23	$< 2 \cdot 10^{-16}$
AIC:	9984.4			
Log-likelihood:	-4984			

Table B.1: Dropped `influenza2` from the initial model of Table 5.2.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	-20.40267	0.13533	-150.759	$< 2 \cdot 10^{-16}$
`I(start/365)`	-0.02820	0.03253	-0.867	0.38605
`sin(start*2*pi/365)`	0.27208	0.09677	2.811	0.00493
`cos(start*2*pi/365)`	0.37575	0.09515	3.949	$7.84 \cdot 10^{-05}$
`influenza0`	0.01710	0.01094	1.563	0.11812
`influenza1`	-0.01980	0.01777	-1.114	0.26522
`(Intercept)`	-15.9532	0.2468	-64.64	$< 2 \cdot 10^{-16}$
AIC:	9982.5			
Log-likelihood:	-4984			

Table B.2: Dropped `influenza3` from the model of Table B.1.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.50301	0.07703	−266.169	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.28117	0.09612	2.925	0.00344
`cos(start*2*pi/365)`	0.37706	0.09537	3.954	$7.7 \cdot 10^{-05}$
`influenza0`	0.01713	0.01093	1.567	0.11721
`influenza1`	−0.02037	0.01773	−1.149	0.25060
`(Intercept)`	−15.9341	0.2422	−65.8	$< 2 \cdot 10^{-16}$
AIC:	9981.3			
Log-likelihood:	−4985			

Table B.3: Dropped `I(start/365)` from the model of Table B.2.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.507315	0.076450	−268.246	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.271011	0.094458	2.869	0.00412
`cos(start*2*pi/365)`	0.378273	0.095335	3.968	$7.25 \cdot 10^{-05}$
`influenza0`	0.003368	0.007652	0.440	0.65982
`(Intercept)`	−15.9394	0.2426	−65.71	$< 2 \cdot 10^{-16}$
AIC:	9980.9			
Log-likelihood:	−4985			

Table B.4: Dropped `influenza1` from the model of Table B.3.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.50140	0.07503	−273.255	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.28095	0.09159	3.067	0.00216
`cos(start*2*pi/365)`	0.38228	0.09500	4.024	$5.73 \cdot 10^{-05}$
`(Intercept)`	−15.9494	0.2435	−65.49	$< 2 \cdot 10^{-16}$
AIC:	9979.2			
Log-likelihood:	−4986			

Table B.5: Dropped `influenza0` from the model of Table B.4.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.50095	0.07507	−273.085	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.25062	0.09406	2.664	0.00771
`cos(start*2*pi/365)`	0.39376	0.09706	4.057	$4.97 \cdot 10^{-05}$
`sin(2*start*2*pi/365)`	0.11151	0.09207	1.211	0.22581
`cos(2*start*2*pi/365)`	−0.11048	0.09368	−1.179	0.23825
`(Intercept)`	−15.9775	0.2494	−64.07	$< 2 \cdot 10^{-16}$
AIC:	9980.3			
Log-likelihood:	−4984			

Table B.6: Adding the first harmonic to the model of Table B.5 worsens AIC.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	-20.55360	0.07111	-289.057	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.24900	0.09790	2.543	0.01098
`cos(start*2*pi/365)`	0.28510	0.09408	3.030	0.00244
`influenza0`	0.01734	0.01087	1.595	0.11077
`influenza1`	-0.01864	0.01719	-1.084	0.27837
`(Intercept)`	-12.38526	0.18894	-65.55	$< 2 \cdot 10^{-16}$
`siaf.1`	2.91781	0.09656	30.22	$< 2 \cdot 10^{-16}$
AIC:	9856.7			
Log-likelihood:	-4921			

Table B.7: Dropped `tiaf.1` from the model of Table 5.3.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	-20.577543	0.070636	-291.319	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.238214	0.096247	2.475	0.013322
`cos(start*2*pi/365)`	0.317140	0.094123	3.369	0.000753
`influenza0`	0.004566	0.007705	0.593	0.553435
`(Intercept)`	-12.39553	0.18872	-65.68	$< 2 \cdot 10^{-16}$
`siaf.1`	2.91780	0.09624	30.32	$< 2 \cdot 10^{-16}$
AIC:	9855.8			
Log-likelihood:	-4922			

Table B.8: Dropped `influenza1` from the model of Table B.7.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	-20.58306	0.06981	-294.826	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.27712	0.09406	2.946	0.003218
`cos(start*2*pi/365)`	0.32557	0.09443	3.448	0.000565
`(Intercept)`	$-1.205e+01$	$2.593e-01$	-46.475	$< 2 \cdot 10^{-16}$
`sexmale`	$-7.905e-01$	$3.402e-01$	-2.324	0.0201
`age`	$7.149e-04$	$8.048e-03$	0.089	0.9292
`siaf.1`	$2.918e+00$	$9.478e-02$	30.785	$< 2 \cdot 10^{-16}$
AIC:	9746.4			
Log-likelihood:	-4866			

Table B.9: Additional effects for `age` and `sex` added to the model of Table 5.4. Due to missing values of the marks, the log-likelihood (and AIC) can not be compared with the models above.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	-20.58111	0.06981	-294.829	$< 2 \cdot 10^{-16}$
`sin(start*2*pi/365)`	0.27243	0.09402	2.897	0.003762
`cos(start*2*pi/365)`	0.31773	0.09446	3.364	0.000769
`(Intercept)`	-12.02286	0.23244	-51.726	$< 2 \cdot 10^{-16}$
`sexmale`	-0.78091	0.34060	-2.293	0.0219
`age==0TRUE`	-0.22028	0.47743	-0.461	0.6445
`siaf.1`	2.91733	0.09475	30.791	$< 2 \cdot 10^{-16}$
AIC:	9746.3			
Log-likelihood:	-4866			

Table B.10: Additional effects for infants (`age==0`) and `sex` added to the model of Table 5.4. The log-likelihood (and AIC) can only be compared with the model in Table B.9 above.

C Model Selection Steps for Finetype C:P1.5,2:F3-3

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.324569	0.124507	−163.240	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.057444	0.031955	−1.798	0.072234
`sin(start*2*pi/365)`	0.357462	0.098497	3.629	0.000284
`cos(start*2*pi/365)`	0.170430	0.095187	1.790	0.073379
`influenza0`	−0.024424	0.040898	−0.597	0.550372
`influenza2`	0.012009	0.030140	0.398	0.690313
`influenza3`	−0.008471	0.022961	−0.369	0.712174
`(Intercept)`	−16.4533	0.3134	−52.5	$< 2 \cdot 10^{-16}$
AIC:	9196.9			
Log-likelihood:	−4590			

Table C.1: Dropped `influenza1` from the initial model of Table 5.6.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.324472	0.124401	−163.379	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.057884	0.031917	−1.814	0.069745
`sin(start*2*pi/365)`	0.354253	0.098490	3.597	0.000322
`cos(start*2*pi/365)`	0.171482	0.095115	1.803	0.071404
`influenza0`	−0.023049	0.040447	−0.570	0.568778
`influenza2`	0.003978	0.024616	0.162	0.871632
`(Intercept)`	−16.4542	0.3135	−52.49	$< 2 \cdot 10^{-16}$
AIC:	9195			
Log-likelihood:	−4591			

Table C.2: Dropped `influenza3` from the model of Table C.1.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.32399	0.12432	−163.476	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.05781	0.03188	−1.814	0.069743
`sin(start*2*pi/365)`	0.35616	0.09797	3.635	0.000278
`cos(start*2*pi/365)`	0.17002	0.09499	1.790	0.073459
`influenza0`	−0.01965	0.03203	−0.614	0.539538
`(Intercept)`	−16.4543	0.3131	−52.56	$< 2 \cdot 10^{-16}$
AIC:	9193			
Log-likelihood:	−4591			

Table C.3: Dropped `influenza2` from the model of Table C.2.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.32709	0.12440	−163.407	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.06155	0.03125	−1.970	0.048864
`sin(start*2*pi/365)`	0.32735	0.09305	3.518	0.000435
`cos(start*2*pi/365)`	0.15764	0.09291	1.697	0.089754
`(Intercept)`	−16.4361	0.3077	−53.41	$< 2 \cdot 10^{-16}$
AIC:	9193.3			
Log-likelihood:	−4592			

Table C.4: Dropped `influenza0` from the model of Table C.3.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.27251	0.11845	−171.149	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.07061	0.03172	−2.226	0.026008
`sin(start*2*pi/365)`	0.33355	0.09540	3.496	0.000472
`cos(start*2*pi/365)`	0.21354	0.09150	2.334	0.019601
`influenza0`	−0.02040	0.03170	−0.644	0.519828
`(Intercept)`	−13.0923	0.3117	−42.00	$< 2 \cdot 10^{-16}$
`siaf.1`	2.8360	0.1497	18.94	$< 2 \cdot 10^{-16}$
AIC:	9156.4			
Log-likelihood:	−4571			

Table C.5: Dropped `tiaf.1` from the model of Table 5.7.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.29276	0.11874	−170.901	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.06969	0.03104	−2.245	0.02474
`sin(start*2*pi/365)`	0.29416	0.09083	3.239	0.00120
`cos(start*2*pi/365)`	0.19196	0.08993	2.135	0.03280
`(Intercept)`	−12.93890	0.43377	−29.829	$< 2 \cdot 10^{-16}$
`sexmale`	0.21221	0.48015	0.442	0.659
`age`	−0.01377	0.01784	−0.772	0.440
`siaf.1`	2.79113	0.15757	17.713	$< 2 \cdot 10^{-16}$
AIC:	9100.4			
Log-likelihood:	−4542			

Table C.6: Additional effects for `age` and `sex` added to the model of Table 5.8. Due to missing values of the marks, the log-likelihood (and AIC) can not be compared with the models above.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.28979	0.11901	−170.494	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.07187	0.03113	−2.308	0.02097
`sin(start*2*pi/365)`	0.29955	0.09114	3.287	0.00101
`cos(start*2*pi/365)`	0.19430	0.09002	2.158	0.03090
`(Intercept)`	−13.2294	0.4379	−30.209	$< 2 \cdot 10^{-16}$
`sexmale`	0.3267	0.4626	0.706	0.480
`age==0TRUE`	−0.2919	0.7731	−0.378	0.706
`siaf.1`	2.8359	0.1484	19.107	$< 2 \cdot 10^{-16}$
AIC:	9101.1			
Log-likelihood:	−4543			

Table C.7: Additional effects for infants (`age==0`) and `sex` added to the model of Table 5.8. The log-likelihood (and AIC) can only be compared with the model in Table C.6 above.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`(Intercept)`	−20.29142	0.11907	−170.419	$< 2 \cdot 10^{-16}$
`I(start/365)`	−0.07159	0.03113	−2.300	0.021468
`sin(start*2*pi/365)`	0.30381	0.09111	3.334	0.000855
`cos(start*2*pi/365)`	0.19525	0.09007	2.168	0.030175
`(Intercept)`	−13.2472	0.4252	−31.152	$< 2 \cdot 10^{-16}$
`sexmale`	0.3126	0.4584	0.682	0.495
`siaf.1`	2.8360	0.1483	19.129	$< 2 \cdot 10^{-16}$
AIC:	9099.3			
Log-likelihood:	−4543			

Table C.8: Additional effect for `sex` (only) added to the model of Table 5.8. Due to a different set of observations caused by the missing values, the log-likelihood (and AIC) can not be compared with the models above.

Declaration of Authorship

I hereby confirm that I have authored this Master's Thesis independently and without use of others than the indicated resources.

München, 18 December 2009 .....