LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Sebastian Meyer, Johannes Elias, Michael Höhle

A space-time conditional intensity model for infectious disease occurence

Technical Report Number 095, 2010
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

A space-time conditional intensity model for infectious disease occurence

Sebastian Meyer*[1,2] Johannes Elias[3] Michael Höhle[4,2]

22nd November 2010

SUMMARY. A novel point process model continuous in space-time is proposed for infectious disease data. Modelling is based on the conditional intensity function (CIF) and extends an additive-multiplicative CIF model previously proposed for discrete space epidemic modelling. Estimation is performed by means of full maximum likelihood and a simulation algorithm is presented. The particular application of interest is the stochastic modelling of the transmission dynamics of the two most common meningococcal antigenic sequence types observed in Germany 2002–2008. Altogether, the proposed methodology represents a comprehensive and universal regression framework for the modelling, simulation and inference of self-exciting spatio-temporal point processes based on the CIF. Application is promoted by an implementation in the `R` package `RLadyBug`.

KEY WORDS: Conditional intensity function; Invasive meningococcal disease; Spatio-temporal point process; Stochastic epidemic modelling

1. Introduction

Infectious diseases – such as influenza, gastroenteritis, and the "swine flu" among humans, or foot and mouth disease, the "bird flu", and classical swine fever among animals – are a matter of tremendous public concern especially gaining attention in case of outbreaks. Collaboration of public health decision makers, veterinaries, microbiologists, epidemiologists, statisticians and many others is indispensable for understanding and

*To whom correspondence should be addressed: `Sebastian.Meyer@med.uni-muenchen.de`
[1]Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-Universität, München, Germany
[2]Department of Statistics, Ludwig-Maximilians-Universität, München, Germany
[3]German Reference Centre for Meningococci at the Institute for Hygiene and Microbiology, University of Würzburg, Germany
[4]Department for Infectious Disease Epidemiology, Robert Koch Institute, Berlin, Germany

controlling disease dynamics. The statistician's contribution is typically based on stochastic epidemic models inheriting from the stochastic susceptible-infectious-recovered (SIR) model described by, e.g., Daley and Gani (1999) or Andersson and Britton (2000).

The present work concentrates on stochastic modelling and associated inference for spatio-temporal epidemic point referenced 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 modelling of invasive meningococcal disease (IMD). IMD is a life-threatening human bacterial disease mostly manifesting as meningitis or sepsis. Its pathogenic agent, *Neisseria meningitidis* (aka *Meningococcus*), can be transmitted by large droplet secretions from the respiratory tract of colonized or infected humans. The only reservoir of Meningococci is the human (mostly nasopharyngeal) mucosa (Rosenstein et al., 2001). Data on cases of IMD related to the two most common meningococcal finetypes B:P1.7-2,4:F1-5 and C:P1.5,2:F3-3 in Germany 2002–2008 are obtained from the German Reference Centre for Meningococci (Nationales Referenzzentrum für Meningokokken, NRZM) hosted at the Institute for Hygiene and Microbiology at the University of Würzburg, Germany. Here, a 'finetype' represents a unique combination of serogroup, sequence type of variable region 1 and 2 of the outer membrane protein PorA, and sequence type of the variable region of the outer membrane protein FetA. One specific question of interest for the researchers at the NRZM is whether the two finetypes (in what follows abbreviated B and C) exhibit different spatio-temporal behaviour. In purely descriptive spatio-temporal visualisations it appears as if finetype B exhibits a more stationary pattern than finetype C, but quantifying this relationship would be an important step in the finetype characterisation of IMD dynamics. Furthermore, a connection between outbreaks of meningococci and influenza is hypothesised. For example, Jensen et al. (2004) found an association between the influenza detection rate and the number of IMD cases during the same week in temporal analysis of data from Northern Jutland County in Denmark, during 1980–1999.

We want to perform such an investigation in a spatio-temporal manner. Therefore, we use spatio-temporal point processes as framework for modelling the dynamics of the epidemic. Point process modelling has in the context of epidemics been used in a discrete spatial setting in, e.g., Lawson and Leimich (2000), Neal and Roberts (2004), Diggle (2006), Scheel et al. (2007) and Jewell et al. (2009). Spatio-temporal epidemic modelling in an explicit continuous spatial setting, however, is rare with Diggle et al. (2005) being one of the few examples of covariate adjusted modelling. One explanation is the balancing between optimal spatial resolution of the data for statistical analyses and confidentiality of cases.

Recently, there have been suggestions for splitting the dynamics of infectious disease into endemic and epidemic components, see Held et al. (2005), Held et al. (2006) and Paul et al. (2008) for a discrete spatial – discrete time perspective and Höhle (2009) for a discrete spatial – continuous time perspective. For the continuous spatial – continuous time setting, similar modelling approaches have been seen in the point process analysis of earthquake data using a trend, cyclic, and clustering decomposition, see e.g. Ogata (1988) and Ogata (1999). Earthquakes and cases of infectious diseases have in common

that they feature so-called *self-excitement*, i.e. events promote the future evolution of the point process by producing "offspring" events. Strong earthquakes cause subsequent aftershocks, and an infected individual may transmit the infectious agent to susceptible individuals. Point process modelling of such clustering in space and time often inherits in structure from *Hawkes' self-exciting process* (Hawkes, 1971), which is a birth process with immigration formulated as a purely temporal process based on the conditional intensity function. Inspired by the self-exciting Hawkes process, Ogata (1988) proposed the so-called *epidemic-type aftershock-sequences* (ETAS) model for earthquake occurrences. The name of the ETAS model clearly indicates a relationship between the modelling of earthquake occurences and infectious disease cases. Ogata's ETAS model has been extended to a spatio-temporal version also covering the locations of earthquakes, and several alternatives for the earthquakes' triggering function have been considered in the seismological literature over the years (cf. Ogata, 1998). Other areas of application drawing on similar modelling approaches are the modelling of forest fires (Peng et al., 2005; Møller and Díaz-Avalos, 2010), of rainstorms (Rodriguez-Iturbe et al., 1987; Cox and Isham, 1988), residential burglaries (Mohler et al., 2010), and the analysis of bird nesting patterns (Diggle et al., 2009). In these applications modelling is, however, often tailored towards the specific use and thus there appears to be a need for a more unifying regression approach for the modelling of space-time phenomena.

In the present work, the multivariate counting process approach of the two-component epidemic modelling of Höhle (2009) is extended to the context of continuous space. Furthermore, the modelling of the epidemic component is additionally generalised by marks and covariates. In the spatial SIR model of Höhle (2009), possible event locations constituted a finite subset of $\mathbb{R}^2$. This enabled compartmental model formulations in an extended SIR context, and re-infection of the same unit could be modelled. This is not possible in the spatially continuous setting of the present work. Instead, the "population" is now considered as a subset of $\mathbb{R}^2$ having infinite size. This means that the random number of events theoretically has no upper bound, that events may occur at any location in the observational region, and that a re-infection has zero probability. The spatial and temporal distance kernels in Höhle (2009) were modelled as a linear combination of basis functions, whereas in our work this restriction is dropped and any Riemann-integrable functions with any number of parameters may be supplied. Altogether, the proposed modelling class provides a very general regression framework – beyond epidemics – for the modelling, inference and simulation of spatio-temporal point processes. For the IMD application this means

1. quantifying spatio-temporal associations between disease occurrence and explaining factors represented as covariates, e.g. population density and outbreaks of influenza,

2. characterising differences in offspring behaviour concerning the two finetypes, as well as the age and gender of infected individuals.

This paper is organised as follows: Section 2 presents the spatio-temporal two-component epidemic model based on the CIF, whereas Sections 3 and 4 discuss inference and si-

mulation for the proposed model. Section 5 analyses the IMD data, and a discussion in Section 6 finalises the paper.

2. Spatio-Temporal Two-Component CIF Model

In the following, we propose a novel additive-multiplicative model for the conditional intensity function of an infectious disease process continuous in space-time with events occuring in a prespecified observation period $[0, T]$, $T > 0$, and observation region $W \subset \mathbb{R}^2$. The CIF $\lambda^*(t, \boldsymbol{s})$ represents the instantaneous rate or hazard for events at time t and location $\boldsymbol{s}$ given all the observations up to time t (the asterisk notation shall represent the conditioning on the random past history of the process).

The basic framework of the proposed model is to superimpose endemic and epidemic components – an idea similar to the ETAS model (Ogata, 1998) or the two-component spatial SIR model (Höhle, 2009):

$$\lambda^*(t, \boldsymbol{s}) = h(t, \boldsymbol{s}) + e^*(t, \boldsymbol{s}) \qquad (t > 0, \boldsymbol{s} \in W) \,. \tag{1}$$

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

2.1. Specification of the Endemic Component $h(t, s)$

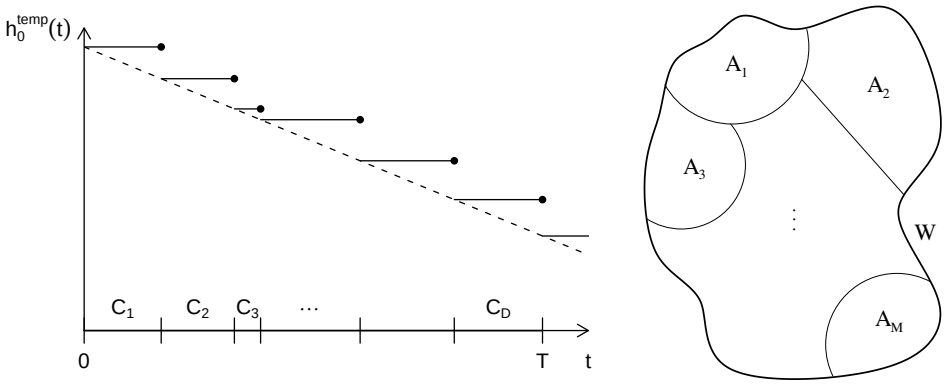
The endemic component is of the multiplicative Cox-type

$$h(t, \boldsymbol{s}) = \exp\left(h_0(t, \boldsymbol{s}) + \tilde{\boldsymbol{\beta}}' \tilde{\boldsymbol{z}}(t, \boldsymbol{s})\right) , \tag{2}$$

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 exogenous covariates actually result from another jointly evolving point process. For example, in the IMD application, an endemic covariate is the number of influenza cases on a week $\times$ district grid (possibly time-lagged). 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 or otherwise standardize to adjust for known confounders. In our case, we model $h_0^{\text{spat}}(\boldsymbol{s})$ as an offset for the logarithmic population density in the district 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: in principle this could be a semiparametric zero-degree B-Spline but we only consider parametric trends (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.



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

(b) Decomposition of W.

Figure 1: 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 $C_1, \ldots, C_D$.

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 (e.g. weeks) are denoted by $C_1, \ldots, C_D \subset (0, T]$, and the spatial tiles (e.g. districts) are denoted by $A_1, \ldots, A_M \subset W$. An illustration is provided in Figure 1. 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

$$h(t, \boldsymbol{s}) = \exp\left(o_{\xi(\boldsymbol{s})} + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right), \tag{3}$$

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 notational convenience, $h_0^{\text{temp}}(t)$ is in (3) now included in the terms of the linear predictor as ordinary covariates.

2.2. 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. This so called triggering function is factorised into separate effects of marks, elapsed time, and relative location:

$$e_j(t, \boldsymbol{s}) = e^{\eta_j}\, g(t - t_j)\, f(\boldsymbol{s} - \boldsymbol{s}_j)\,, \qquad (t > t_j) \tag{4}$$

where $\eta_j = \gamma_0 + \boldsymbol{\gamma}'\boldsymbol{m}_j$ is a linear predictor based on the vector of unpredictable marks $\boldsymbol{m}_j$ attached to the infected individual, and g and f are positive, Riemann-integrable, temporal and spatial interaction functions, respectively. The effects $\boldsymbol{\gamma}$ of marks shall reflect that different individuals might cause more or less secondary cases, depending on individual characteristics; e.g. probability of meningococcal transmission increases with carriage rate, which is highest in adolescents and young adults (Claus et al., 2005).

As in classical regression models, interaction terms of different marks can also be included in the predictor.

The interaction functions describe the decay of infectivity with an increasing spatial or temporal distance from the infection source. In infectious disease applications, f is often taken to be 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

$$f(\boldsymbol{s}) = \exp\left(-\frac{\|\boldsymbol{s}\|^2}{2\sigma^2}\right) \qquad (\boldsymbol{s} \in \mathbb{R}^2,\, \sigma > 0)\,, \tag{5}$$

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

$$g(t) = e^{-\alpha t} \qquad (t > 0,\, \alpha > 0) \tag{6}$$

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.

The resulting epidemic component $e^*(t, \boldsymbol{s})$ then is the sum of the contributions (4) of all infectious individuals at time t and location $\boldsymbol{s}$. Formally,

$$\begin{aligned} e^*(t, \boldsymbol{s}) &= \int_{(0,t)\times W\times\mathcal{M}} \mathbb{1}_{(0,\varepsilon]}(t - \tilde{t})\, \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s} - \tilde{\boldsymbol{s}}\|)\, e^{\eta_j}\, g(t - \tilde{t})\, f(\boldsymbol{s} - \tilde{\boldsymbol{s}})\, N(\mathrm{d}\tilde{t} \times \mathrm{d}\tilde{\boldsymbol{s}} \times \mathrm{d}\tilde{\boldsymbol{m}})\,, \\ &= \sum_{j \in I^*(t,\boldsymbol{s};\varepsilon,\delta)} e^{\eta_j}\, g(t - t_j)\, f(\boldsymbol{s} - \boldsymbol{s}_j)\,, \end{aligned} \tag{7}$$

where $\mathcal{M}$ is the mark space and N is the time-space-mark point process counting the infections. In the above, the hyperparameters $\varepsilon, \delta > 0$ are introduced as known maximum temporal and spatial interaction ranges. A past event only influences the process at time t and location $\boldsymbol{s}$, if both indicator functions are true, i.e. if it occurred 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 separate hyperparameters suggests a convenient decomposition by defining $I^*(t, \boldsymbol{s}; \varepsilon, \delta)$ as the history-dependent set that contains the indices of all past events which are still infective at time t and location $\boldsymbol{s}$. In what follows, the explicit conditioning on ε and δ will be dropped from the function I^*.

2.3. Characteristics of the Model

Altogether, the proposed CIF model for a self-exciting spatio-temporal point process with components (3) and (7) is

$$\lambda^*(t, \boldsymbol{s}) = \exp\left(o_{\xi(\boldsymbol{s})} + \boldsymbol{\beta}'\boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right) + \sum_{j \in I^*(t,\boldsymbol{s})} e^{\eta_j}\, g(t - t_j)\, f(\boldsymbol{s} - \boldsymbol{s}_j)\,, \tag{8}$$

which we shall call `twinstim` to indicate a *two*-component *s*patio-*t*emporal (conditional) *i*ntensity *m*odel.

Similar models have been considered in an epidemic modelling context by Diggle (2007, Section 1.3.2.2). He describes a self-exciting spatio-temporal point process resembling the epidemic component $e^*(t, \boldsymbol{s})$ from equation (7). Specifically, his model is an instance of `twinstim` having neither an endemic component ($h_0(t, \boldsymbol{s}) \equiv -\infty$) nor a weighting of infectives ($\eta_j \equiv 0$), and assuming $\varepsilon = \delta = \infty$. Thus, the model basically is a Hawkes process without immigration, spatially enriched by a Thomas process. The population growth model described in Diggle et al. (2009) extends this by a temporally limited interaction range ε and an endemic component being a space-time constant value, i.e. also this model is an instance of `twinstim`.

For the proposed model an interesting quantity is the individual-specific mean number μ_j of infections caused by individual j inside its spatio-temporal range of interaction:

$$\begin{aligned} \mu_j &= \int_0^\infty \int_{\mathbb{R}^2} e_j(t, \boldsymbol{s})\, \mathbb{1}_{(0,\varepsilon]}(t - t_j)\, \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s} - \boldsymbol{s}_j\|)\, \mathrm{d}t\, \mathrm{d}\boldsymbol{s} \\ &= e^{\eta_j} \cdot \left[\int_0^\varepsilon g(t)\, \mathrm{d}t\right] \cdot \left[\int_{b(\mathbf{0},\delta)} f(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s}\right]\,. \end{aligned} \tag{9}$$

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 occurs in unlimited time and space. In practice this is not observable, but individuals near the border would be attributed a truncated value of μ_j if integrating over W only. 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 $[0, T]$. These edge effects are thus overcome by the formulation (9), which also simplifies interpretation and hence provides a quantity similar to the basic reproduction number R_0 known from classical epidemic modelling (Anderson and May, 1991). Specifically, the number μ_j offers an intuitive way of interpreting the parameters $\boldsymbol{\gamma}$ in the linear predictor η_j. An "intercept" term γ_0 multiplied by the two integral values would represent 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} multiplies the mean number by the corresponding parameter e^{γ_l}.

2.4. Extension: Type-Specific `twinstim`

The IMD data actually represent a spatio-temporal point pattern marked by the fine-type 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. Both finetypes should however have the same relation to the population at risk and, as we assume for simplicity, to the numbers of influenza cases and the time trend. The only finetype-specific element in the endemic component would be the intercept, corresponding to the global background rate. 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 in the specific application and marked point patterns in general.

Denote by $\mathcal{K} = \{1, \ldots, K\} \subset \mathbb{N}$ 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:

$$\begin{aligned}
\lambda^*(t, \boldsymbol{s}, \kappa) &= h(t, \boldsymbol{s}, \kappa) + e^*(t, \boldsymbol{s}, \kappa) \qquad (10)\\
h(t, \boldsymbol{s}, \kappa) &= \exp\left(h_0^{\text{type}}(\kappa) + o_{\xi(\boldsymbol{s})} + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right)\\
e^*(t, \boldsymbol{s}, \kappa) &= \sum_{j \in I^*(t,\boldsymbol{s},\kappa)} e_j(t, \boldsymbol{s})\\
e_j(t, \boldsymbol{s}) &= \exp(\eta_j) \cdot g(t - t_j | \kappa_j) \cdot f(\boldsymbol{s} - \boldsymbol{s}_j | \kappa_j)\\
I^*(t, \boldsymbol{s}, \kappa) &= \left\{ j \in \{1, \ldots, N_g(t-)\} : \; \mathbb{1}_{(0,\varepsilon]}(t - t_j) = 1 \; \wedge \; \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s} - \boldsymbol{s}_j\|) = 1 \; \wedge \; q_{\kappa_j,\kappa} = 1 \right\}
\end{aligned}$$

Here, the transmission indicators from the matrix $\boldsymbol{Q}$ have been integrated into $I^*(t, \boldsymbol{s}, \kappa)$, where $N_g(t-)$ is the overall number of infections just before time t, and $\eta_j = \boldsymbol{\gamma}' \boldsymbol{m}_j$ is a linear predictor based on the event marks from the history of the process. Note that the event type κ_j is part of the vector $\boldsymbol{m}_j$, which enables type-specific epidemic intercepts as well as type-interactions with individual covariates. The unmarked `twinstim` of the previous subsection can be treated as a special case of the above model by assuming $\mathcal{K} = \{1\}$, i.e. a single type of event. 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 $h(t, \boldsymbol{s}, \kappa)$ must no longer contain an intercept. For the remainder of the endemic predictor, the model assumes independence of κ, which means that the effect of endemic covariates is homogeneous over the event types. 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 by using type-specific σ_κ^2 and α_κ in (5) and (6), respectively.

3. Statistical Inference

After formulating the marked `twinstim` model in (10) this section deals with likelihood inference for the parameters of the CIF based on the observed marked spatio-temporal point pattern $\boldsymbol{x} = \{(t_i, \boldsymbol{s}_i, \boldsymbol{m}_i) : i = 1, \ldots, n\}$, where the event type κ_i is part of the vector of marks $\boldsymbol{m}_i$, and n is the number of events, i.e. a realisation of $N_g(T)$. The parameter vector in question is

$$\boldsymbol{\theta} = (\boldsymbol{\beta}_\mathbf{0}', \boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\sigma}', \boldsymbol{\alpha}')' \,, \tag{11}$$

where $\boldsymbol{\beta}_\mathbf{0} = (\beta_{0,1}, \ldots, \beta_{0,K})'$ (type-specific) or $\boldsymbol{\beta}_\mathbf{0} = \beta_0$ (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.

In a frequentist framework, parameter estimates can be obtained by maximisation of the log-likelihood or the partial log-likelihood 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 for their special instance of a `twinstim` model. However, we do not see great benefit in using the partial likelihood approach because the need for spatial integration – which is the computational bottleneck of statistical inference – remains. Furthermore, the parameters e^{β_0} and e^{γ_0} would not both be identifiable, but only their ratio. As a consequence, we will concentrate on full maximum likelihood estimation – although our implementation also provides the partial likelihood alternative.

In the next subsections, the log-likelihood and score functions related to the type-specific `twinstim` are derived. Furthermore, estimation of the expected Fisher information matrix and asymptotic properties of the maximum likelihood estimators are discussed.

3.1. Log-Likelihood Function

In this framework, no attempt is made to model unpredictable marks like gender and age but they are taken as given predictor variables in models of the CIF. In this case, the conditional intensity function of the underlying point process N on $[0; T] \times W \times \mathcal{M}$ may be conveniently written as

$$\lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \boldsymbol{m}) = \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) \cdot f_{\tilde{\mathcal{M}}}(\tilde{\boldsymbol{m}} | t, \boldsymbol{s}, \kappa) \,, \tag{12}$$

where the tilde separates the event type κ from the unpredictable marks $\tilde{\boldsymbol{m}}$, and $f_{\tilde{\mathcal{M}}}$ is the density of the mark distribution at a specific time t, location $\boldsymbol{s}$ and event type κ. Then, the log-likelihood function $l(\boldsymbol{\theta}; \boldsymbol{x}, T, W, \mathcal{K})$ of the point process is given by

$$\left[\sum_{i=1}^{n} \log \lambda^*_{\boldsymbol{\theta}}(t_i, \boldsymbol{s}_i, \kappa_i) - \int_0^T \int_W \sum_{\kappa \in \mathcal{K}} \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s}\right] + \left[\sum_{i=1}^{n} \log f_{\tilde{\mathcal{M}}}(\tilde{\boldsymbol{m}}_i | t_i, \boldsymbol{s}_i, \kappa_i)\right] \tag{13}$$

(Daley and Vere-Jones, 2003), where the second term is an ordinary i.i.d. log-likelihood for the unpredictable marks, and the first term has the same form as the log-likelihood of an ordinary point process uniquely specified through its CIF. However, observe that the history of the process hidden in the asterisk of the CIF also contains past values of the unpredictable marks which influence the future evolution of the process. Furthermore, if the process depends on some exogeneous variables (like e.g. the infection intensity of IMD might depend on waves of influenza), the conditional intensity function is controlled by some larger history, which also incorporates the pasts of all related jointly evolving processes. As we do not attempt to model the unpredictable marks $\tilde{\boldsymbol{m}}$, interest is in maximizing the first term in (13) only as a kind of partial likelihood (Vere-Jones, 2009, p. 180). Therefore, 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 above 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:

$$\int_0^T \int_W \sum_{\kappa \in \mathcal{K}} h_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s} = \left(\sum_{\kappa \in \mathcal{K}} \exp\left(h_0^{\text{type}}(\kappa) \right) \right) \cdot \sum_{\tau=1}^{D} \sum_{\xi=1}^{M} |C_\tau| |A_\xi| \exp\left(o_\xi + \boldsymbol{\beta}' \boldsymbol{z}_{\tau,\xi} \right) \tag{14}$$

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

$$\begin{aligned}
&\int_0^T \int_W \sum_{\kappa \in \mathcal{K}} \sum_{j=1}^{n} \mathbb{1}_{(0,\varepsilon]}(t - t_j) \, \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s} - \boldsymbol{s}_j\|) \, q_{\kappa_j,\kappa} \, e^{\eta_j} \, g_{\boldsymbol{\alpha}}(t - t_j | \kappa_j) \, f_{\boldsymbol{\sigma}}(\boldsymbol{s} - \boldsymbol{s}_j | \kappa_j) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{s} \\
&= \sum_{j=1}^{n} q_{\kappa_j,\bullet} \, e^{\eta_j} \left[\int_0^T \mathbb{1}_{(0,\varepsilon]}(t - t_j) g_{\boldsymbol{\alpha}}(t - t_j | \kappa_j) \, \mathrm{d}t \right] \left[\int_W \mathbb{1}_{[0,\delta]}(\|\boldsymbol{s} - \boldsymbol{s}_j\|) f_{\boldsymbol{\sigma}}(\boldsymbol{s} - \boldsymbol{s}_j | \kappa_j) \, \mathrm{d}\boldsymbol{s} \right] \\
&= \sum_{j=1}^{n} q_{\kappa_j,\bullet} \, e^{\eta_j} \left[\int_0^{\min\{T - t_j; \varepsilon\}} g_{\boldsymbol{\alpha}}(t | \kappa_j) \, \mathrm{d}t \right] \left[\int_{R_j} f_{\boldsymbol{\sigma}}(\boldsymbol{s} | \kappa_j) \, \mathrm{d}\boldsymbol{s} \right].
\end{aligned} \tag{15}$$

Here, $q_{\kappa_j,\bullet} := \sum_{\kappa \in \mathcal{K}} q_{\kappa_j,\kappa}$ is the number of different event types which can be triggered by an event of type κ_j, and $R_j := \left[W \cap b(\boldsymbol{s}_j; \delta) \right] - \boldsymbol{s}_j$ is the spatial interaction region of the jth event centred at its location. 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. Meyer (2009) compared accuracy and speed of different cubature rules for performing the numerical integration. Here, the two-dimensional midpoint rule (see e.g. Stroud, 1971) proved to be best suited for the task. For the special case of the type-specific Gaussian kernel (5), robust accuracy for any value of σ_κ can be achieved by an adaptive choice of the bandwidth $h = \phi\,\sigma_\kappa$, see Meyer (2009, 3.2.7) for details.

In contrast, the evaluation of the definite integral over the temporal interaction function is analytically accessible for the choice of $g_{\boldsymbol{\alpha}}$ used in our work. Provided $G_{\boldsymbol{\alpha}}(t|\kappa)$ denotes an antiderivative of $g_{\boldsymbol{\alpha}}(t|\kappa)$, the first integral thus equals

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

For instance, the type-specific exponential decay function $g_{\boldsymbol{\alpha}}$ from equation (6) has antiderivative

$$G_{\boldsymbol{\alpha}}(t|\kappa) = -\frac{e^{-\alpha_\kappa\, t}}{\alpha_\kappa} \qquad (\alpha_\kappa > 0)\,. \tag{16}$$

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

Altogether, an analytical maximisation of the log-likelihood for the `twinstim` class is not feasible, and a numerical optimisation routine such as BFGS (see e.g. Nocedal and Wright, 1999, Section 8.1) is required. Here, it is advantageous to know the derivative with respect to $\boldsymbol{\theta}$, i.e. the score function which is presented below.

3.2. Score Function

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

$$\boldsymbol{s}_{\boldsymbol{\vartheta}}(\boldsymbol{\theta}) := \frac{\partial}{\partial\boldsymbol{\vartheta}} l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\boldsymbol{\vartheta}}\lambda^*_{\boldsymbol{\theta}}(t_i,\boldsymbol{s}_i,\kappa_i)}{\lambda^*_{\boldsymbol{\theta}}(t_i,\boldsymbol{s}_i,\kappa_i)} - \int_0^T\int_W \sum_{\kappa\in\mathcal{K}} \frac{\partial}{\partial\boldsymbol{\vartheta}}\lambda^*_{\boldsymbol{\theta}}(t,\boldsymbol{s},\kappa)\,\mathrm{d}t\,\mathrm{d}\boldsymbol{s}\,, \tag{17}$$

and the score function is

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

The necessary partial derivatives of the CIF with their respective time-space-mark integrals are given in Appendix A, and can be plugged into the equation (17).

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. For instance, a type-specific spatial Gaussian kernel with $\boldsymbol{\sigma} = (\sigma_1,\ldots,\sigma_K)'$ similar to equation (5) has partial derivatives

$$\frac{\partial}{\partial\sigma_k} f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa) = \mathbb{1}_{k=\kappa}(\kappa)\cdot\exp\left(-\frac{\|\boldsymbol{s}\|^2}{2\sigma_k^2}\right)\frac{\|\boldsymbol{s}\|^2}{\sigma_k^3}, \qquad \text{for any } k\in\mathcal{K}\,. \tag{18}$$

The type-specific temporal exponential decay function with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ from equation (6) has partial derivatives

$$\frac{\partial}{\partial \alpha_k} g_{\boldsymbol{\alpha}}(t|\kappa) = \mathbb{1}_{k=\kappa}(\kappa) \cdot \left(-t\, e^{-\alpha_k\, t}\right), \qquad \text{for any } k \in \mathcal{K}\,. \tag{19}$$

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.

3.3. 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 discussed in Meyer (2009, Section 4.2.3) and are assumed satisfied.

As an alternative, parametric bootstrap using the simulation algorithm from Section 4 may be useful for obtaining approximate standard errors (Schoenberg et al., 2002). However, we will restrict estimation of the uncertainty of $\hat{\boldsymbol{\theta}}_{ML}$ to the asymptotic properties based on the Fisher information matrix as this computation is much faster. The *expected* Fisher information $\mathcal{I}(\boldsymbol{\theta})$ can be estimated by the "optional variation process" – adapted from Martinussen and 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

$$\hat{\mathcal{I}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(\frac{\partial}{\partial \tilde{\boldsymbol{\theta}}} \log \lambda^*_{\tilde{\boldsymbol{\theta}}}(t_i, \boldsymbol{s}_i, \kappa_i)\Big|_{\tilde{\boldsymbol{\theta}}=\boldsymbol{\theta}} \right)^{\otimes 2} = \sum_{i=1}^{n} \left(\frac{\frac{\partial}{\partial \tilde{\boldsymbol{\theta}}} \lambda^*_{\tilde{\boldsymbol{\theta}}}(t_i, \boldsymbol{s}_i, \kappa_i)}{\lambda^*_{\tilde{\boldsymbol{\theta}}}(t_i, \boldsymbol{s}_i, \kappa_i)}\Big|_{\tilde{\boldsymbol{\theta}}=\boldsymbol{\theta}} \right)^{\otimes 2}, \tag{20}$$

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 (17).

Uncertainty of the parameter estimates is thus deduced from the diagonal of $\hat{\mathcal{I}}^{-1/2}(\hat{\boldsymbol{\theta}}_{ML})$, which contains their standard errors. Utilizing the asymptotic normality of $\hat{\boldsymbol{\theta}}_{ML}$, ordinary component-wise Wald confidence intervals 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.

With respect to model selection, likelihood ratio tests and Akaike's information criterion will be used. The former permits e.g. to test whether the null hypothesis of a type-invariant endemic intercept $\beta_0 = \beta_{0,1} = \cdots = \beta_{0,K}$ can be rejected.

4. Simulation Algorithm

In general, the usability of a model class is greatly improved by the ability of simulation from a specific model. For instance, it enables model checking and parametric bootstrap. For evolutionary point processes specified by their conditional intensity function, *Ogata's modified thinning algorithm* (Daley and Vere-Jones, 2003, Algorithm 7.5.V.) provides a convenient way to simulate realisations of the process. The algorithm requires piecewise upper bounds for the intensity $\lambda_g^*(t)$ of the ground process $N_g(t) := N((0,t] \times W \times \mathcal{K})$, which counts the total number of points occuring up to time t anywhere in W and of any type $\kappa \in \mathcal{K}$. This intensity is determined as

$$\begin{aligned}
\lambda_g^*(t) &= \int_W \sum_{\kappa \in \mathcal{K}} \lambda^*(t, \boldsymbol{s}, \kappa) \, \mathrm{d}\boldsymbol{s} && (21)\\
&= \left(\sum_{\kappa \in \mathcal{K}} e^{\beta_{0,\kappa}} \right) \left(\sum_{\xi=1}^{M} |A_\xi| \, e^{o_\xi + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(t),\xi}} \right) && (22)\\
&\quad + \sum_{j=1}^{N_g(t-)} \left[\left(\sum_{\kappa \in \mathcal{K}} q_{\kappa_j,\kappa} \right) e^{\eta_j} \, \mathbb{1}_{(0,\varepsilon]}(t - t_j) \, g(t - t_j | \kappa_j) \int_{R_j} f(\boldsymbol{s} | \kappa_j) \, \mathrm{d}\boldsymbol{s} \right] . && (23)
\end{aligned}$$

This function is a.s. bounded above by the CIF $\overline{\lambda_g^*}(t)$, which is defined by replacing $g(t|\kappa)$ by the *constant* temporal interaction function $\overline{g}(t|\kappa) = \max_{u>0} g(u|\kappa)$. This CIF is piecewise constant in time as it only jumps at time points where any of the endemic covariates in $\boldsymbol{z}_{\tau(t),\xi}$ in any tile ξ changes its value, or when the set of currently infectious individuals changes, i.e. when a new event occurs or a previous event stops triggering.

Given a parameter vector $\boldsymbol{\theta}$, the ranges of interaction ε and δ, as well as a sampling scheme for the marks $\boldsymbol{m}_j$, the time point of the next infection starting from the current time $t = t_0$ can be generated as follows:

Draw an exponentially distributed random variate Δ with rate $\overline{\lambda_g^*}(t_0)$. The simulated value of Δ is a proposal for the waiting time to the next event, i.e. the next time point of infection might be $\tilde{t} = t_0 + \Delta$. However, this proposal is not valid if the rate $\overline{\lambda_g^*}(t)$ had changed between t_0 and $\tilde{t}$. In this case, time is set to the first changepoint after t_0 and a new Δ is simulated. Eventually, a proposed time point $\tilde{t}$ is valid. It is then accepted with probability $\lambda_g^*(\tilde{t}) / \overline{\lambda_g^*}(\tilde{t})$. If it is rejected, time is set to $t = \tilde{t}$ and a new waiting time Δ is simulated as above. If it is accepted, location $\tilde{\boldsymbol{s}}$ and type $\tilde{\kappa}$ of the event have to be simulated. At first, the source of infection is sampled with probabilities proportional to

the respective components of $\lambda_g^*(\tilde{t})$:

$$\begin{aligned}
\mathbb{P}(\text{endemic source}) \cdot \lambda_g^*(\tilde{t}) &= \left(\sum_{\kappa \in \mathcal{K}} e^{\beta_{0,\kappa}}\right) \left(\sum_{\xi=1}^{M} |A_\xi| \, e^{o_\xi + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(\tilde{t}),\xi}}\right) \\
\mathbb{P}(\text{source} = \text{event } j) \cdot \lambda_g^*(\tilde{t}) &= \left(\sum_{\kappa \in \mathcal{K}} q_{\kappa_j,\kappa}\right) e^{\eta_j} \, \mathbb{1}_{(0,\varepsilon]}(\tilde{t} - t_j) \, g(\tilde{t} - t_j | \kappa_j) \int_{R_j} f(\boldsymbol{s}|\kappa_j) \, \mathrm{d}\boldsymbol{s} \,,
\end{aligned}$$

for $j \in \{1, \ldots, N_g(\tilde{t}-)\}$. On the one hand, if the new event has an endemic source, then

$$\begin{aligned}
\mathbb{P}(\tilde{\kappa} = k) &\propto \exp(\beta_{0,k}) \qquad (k \in \mathcal{K}) \\
\mathbb{P}(\tilde{\boldsymbol{s}} \in A_\xi) &\propto |A_\xi| \, e^{o_\xi + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(\tilde{t}),\xi}} \qquad (\xi = 1, \ldots, M)
\end{aligned}$$

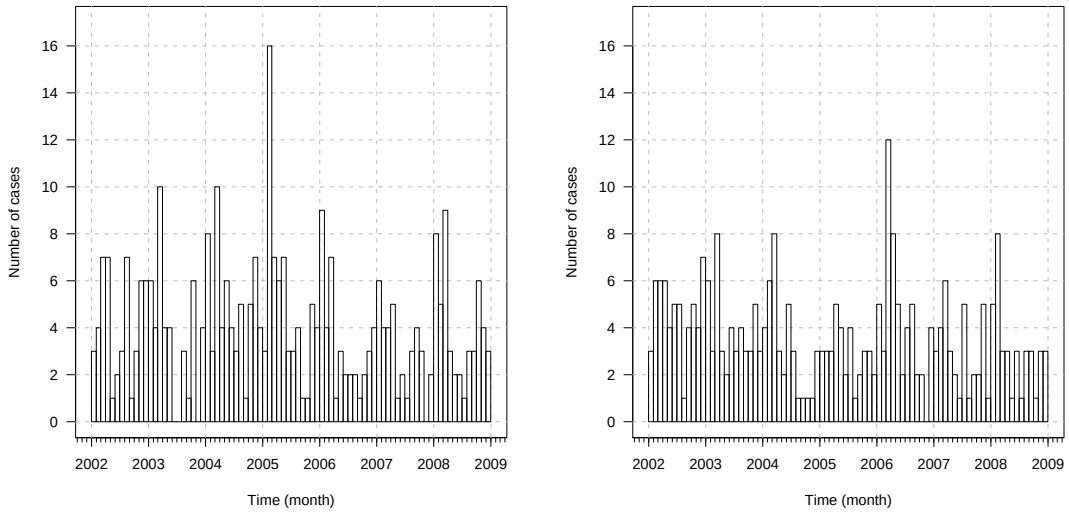
In the sampled tile $A_{\tilde{\xi}}$, the location $\tilde{\boldsymbol{s}}$ is uniformly distributed. On the other hand, if the new event was triggered by the previous event j, then $\tilde{\kappa} \sim U(\{k : q_{\kappa_j,k} = 1\})$ (i.e. one of the types which can be triggered by the type κ_j without any weighting), and $\tilde{\boldsymbol{s}} = \boldsymbol{s}_j + V$, where V is drawn from the density $f(\boldsymbol{s}|\kappa_j) / \int_{R_j} f(\boldsymbol{s}|\kappa_j) \, \mathrm{d}s$ on R_j, e.g. using rejection sampling.

5. Application to the IMD Data

In what follows we analyse the spatio-temporal spread of IMD based on the 2002-2008 NRZM data. The postal code of the patient's home address was the spatial resolution available for our analysis. Despite being spatially discrete we consider centroids of postal code areas as quasi-continuous in space when looking at entire Germany. As usual with infectious diseases, the actual time point of infection is unknown for the IMD cases. Therefore, we define the beginning of illness and infectivity as the date of specimen sampling.

All in all, $n = 636$ infections with finetypes B (336) and C (300) have been registered. Figure 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.

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 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 postal code, 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 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 of the districts, which



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

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

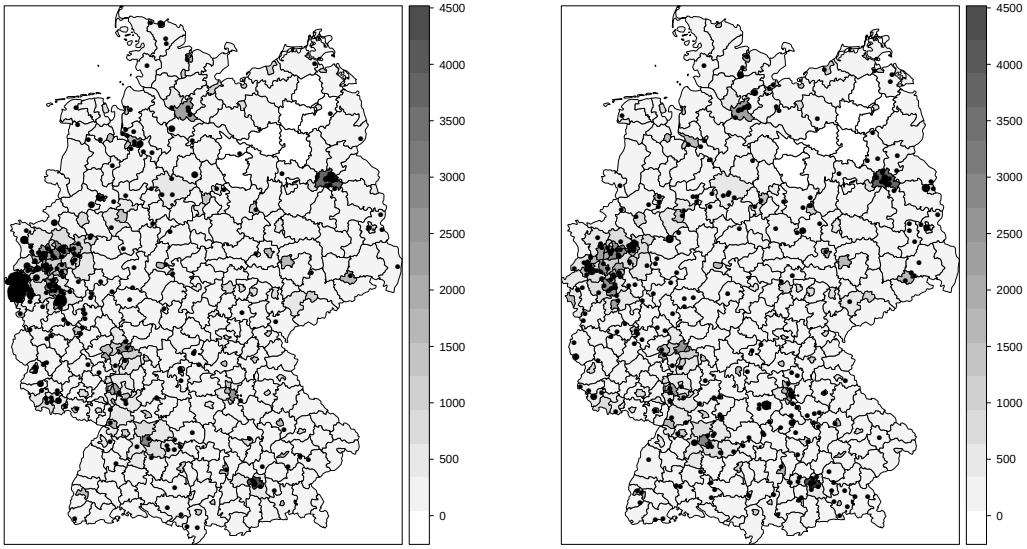
Figure 2: Monthly numbers of IMD cases for both finetypes separately.

can be assumed to be roughly proportional to the population at risk of infection. Spatial heterogeneity of the observed point patterns thus partially arises from spatial variation in the population density. Not surprisingly, the intensity of points in metropolitan areas like Berlin, Munich or the Ruhr is higher.

Animated graphics of the space-time locations of infections give more insight into the infectiousness of the finetypes, i.e. of their epidemic character, and can be found as Web Animation 1. The most common finetype B 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, yet not proven, that this phenomenon is due to differences in the mucosal immune reaction elicited; specifically, finetype B might be more successful than C in evading mucosal clearance.

Concerning the supplied patient data there is not much difference between the finetypes. About 46.3% of all infections affected females, while the finetype-specific proportions are similar with 45.9% (B) and 46.6% (C). Both finetypes have a modal age of < 1 year (13.1% of all cases) followed by children of one (10.1%) and two (7.4%) 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 second large group of 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.

As mentioned in the introduction, a concurrent viral infection of the upper respiratory tract is hypothesized to be an important risk factor for IMD. Since influenza is a notifiable



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

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

Figure 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 districts (source: Federal Statistical Office (DESTATIS) (2009)).

pathogen we investigate the effect of the weekly number of influenza cases in the 413 districts of Germany taken from the SurvStat database (Robert Koch-Institut, 2009) on the occurrence of the two finetypes.

Although visual comparisons between the finetypes and heuristic comparisons of the estimates of separate finetype-specific models are possible, this does not allow to assess potential differences statistically. We thus opt for a joint analysis of the two finetypes by the marked `twinstim` of Section 2.4 in order to test whether the weight of the epidemic component and hence the basic reproduction number is significantly higher for the serogroup B finetype.

We perform model selection for the joint point pattern of 630 cases of IMD with complete age and gender information by using Akaike's information criterion (AIC) to compare all models with the CIF composed by subsets of the following terms:

- Endemic component: Common or finetype-specific intercept, linear time trend, sine-cosine time-of-year effects (zero, one or two harmonics), and linear effect of weekly number of influenza cases registered in the district of a point (lag 0 – lag 3).
- Epidemic component: Age (categorized as 0-2, 3-18 and $\geq$19 years), gender, finetype and age-finetype interaction.

	Estimate	Std. Error	z value	$\mathbb{P}(\lvert Z\rvert > \lvert z\rvert)$
`h.(Intercept)`	-20.3652	0.0872	-233.53	$< 2 \cdot 10^{-16}$
`h.I(start/365)`	-0.0493	0.0223	-2.21	0.027
`h.sin(start*2*1*pi/365)`	0.2618	0.0649	4.03	$5.5 \cdot 10^{-05}$
`h.cos(start*2*1*pi/365)`	0.2668	0.0644	4.14	$3.4 \cdot 10^{-05}$
`e.(Intercept)`	-12.5746	0.3128	-40.21	$< 2 \cdot 10^{-16}$
`e.agegrp[3,19)`	0.6463	0.3195	2.02	0.04310
`e.agegrp[19,Inf)`	-0.1868	0.4321	-0.43	0.66558
`e.typeC`	-0.8496	0.2574	-3.30	0.00097
`e.siaf`	2.8287	0.0819		
AIC:	18968			
Log-likelihood:	-9475			

Table 1: Parameter estimates of the AIC-best model with `h.<NAME>` denoting endemic and `e.<NAME>` denoting epidemic components, e.g. `e.siaf`$=\log\sigma$. The p-values correspond to Wald tests.

Fixed hyperparameters of $\varepsilon = 30$ days and $\delta = 200$ km were used for all analyses. Furthermore, a constant temporal interaction function g was used. To restrict the model search, and hence computing time, we first performed the search for all 600 models with constant spatial interaction function f. Hereafter, the top 10 models of this search were investigated further with two Gaussian spatial interaction functions: one with joint variance parameter and one with finetype-specific variance parameter.

The CIF of the resulting AIC-best model obtained by this search has the following form:

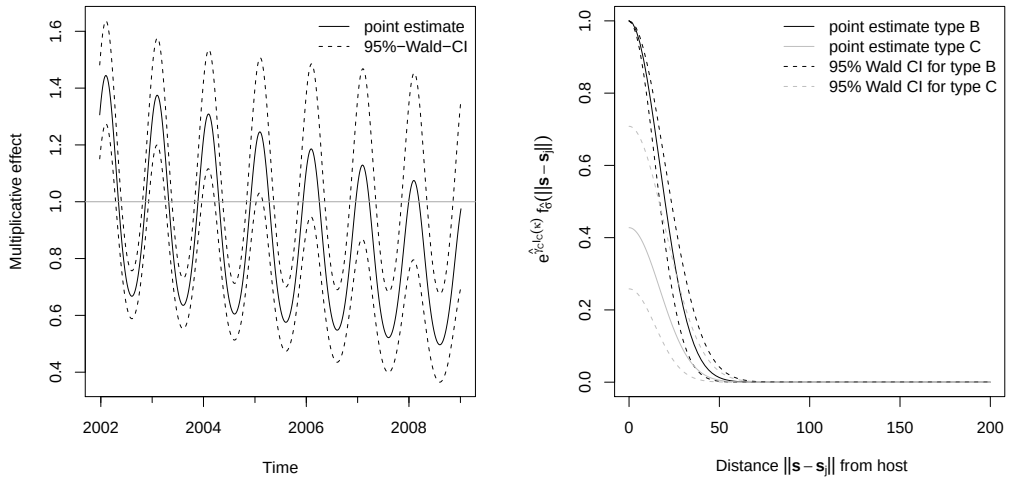
$$\begin{aligned}\lambda^*_{\boldsymbol{\theta}}(t,\boldsymbol{s},\kappa) &= \rho_{\xi(\boldsymbol{s})} \cdot \exp\left(\beta_0 + \beta_{\text{trend}}\tfrac{\lfloor t\rfloor}{365} + \beta_{\text{sin}}\sin\left(\lfloor t\rfloor\tfrac{2\pi}{365}\right) + \beta_{\text{cos}}\cos\left(\lfloor t\rfloor\tfrac{2\pi}{365}\right)\right)\\ &\quad + \sum_{j\in I^*(t,\boldsymbol{s},\kappa;\varepsilon,\delta)} q_{\kappa_j,\kappa}\, e^{\gamma_0+\gamma_{3\text{-}18}\mathbb{1}_{[3,18]}(\text{age}_j)+\gamma_{\geq 19}\mathbb{1}_{[19,\infty)}(\text{age}_j)+\gamma_{\text{C}}\mathbb{1}_{\{C\}}(\kappa_j)}\, f_\sigma(\boldsymbol{s}-\boldsymbol{s}_j).\end{aligned}$$

Here, $(t,\boldsymbol{s},\kappa)$ denotes days since 31 December 2001, coordinate in ETRS89 (kilometre scale) and finetype. With $\lfloor t\rfloor$ we denote monday of week $\tau(t)$, i.e. the lower bound of time intervals $C_1,\ldots,C_D$. Furthermore, $\rho_{\xi(\boldsymbol{s})}$ denotes the district-specific population density (inhabitants per km^2). In the linear predictor of the epidemic component, age group $[0,2]$ and type B serve as reference categories. The corresponding parameter estimates of the best model, now fitted to the 635 cases with available age, are found in Table 1.

Thus, there appears to be no noteworthy difference in the endemic behaviour of the two types: a linear downward time trend superimposed with one harmonic best describes the endemic behaviour of the point pattern (see Figure 4(a)). An additional effect of past numbers of influenza cases does not improve the model. In contrast, there is an effect of past IMD cases, i.e. the process is indeed self-exciting. Comparing the endemic-only model with the model enriched by an epidemic intercept only, greatly improves the fit (ΔAIC=202.84). In the epidemic component, there is a detectable dependence on marks with type C being less aggressive than type B (i.e. $\hat{\gamma}_C = -0.85 < 0$). Figure 4(b)

shows the resulting finetype-specific spatial interaction functions which for type C is $\exp(\hat{\gamma}_C) \cdot 100\% = 43\%$ of type B. Finally, there is a significant age difference in the infectivity of cases: the highest potential is found in the 3-18 year old, which could be interpreted as the school-aged children having a higher contact behaviour than e.g. adults.

Based on the selected model, basic reproduction numbers of $\hat{\mu}_{\mathrm{B}} = 0.25$ (95%-CI: 0.19-0.34) vs. $\hat{\mu}_{\mathrm{C}} = 0.11$ (95%-CI: 0.07-0.18) are obtained by calculating the type-specific expectation of (9) over the empirical distribution function of the additional covariates in the epidemic predictor (here: age group). The confidence intervals are given as the 0.025 and 0.975 quantiles of samples obtained by re-computing $\hat{\mu}_{\mathrm{B}}$ and $\hat{\mu}_{\mathrm{C}}$ for 999 additional coefficient vectors drawn from the asymptotic multivariate normal distribution of the parameter estimates in Table 1. The confidence intervals thus indicate a higher epidemic potential of the serogroup B finetype.



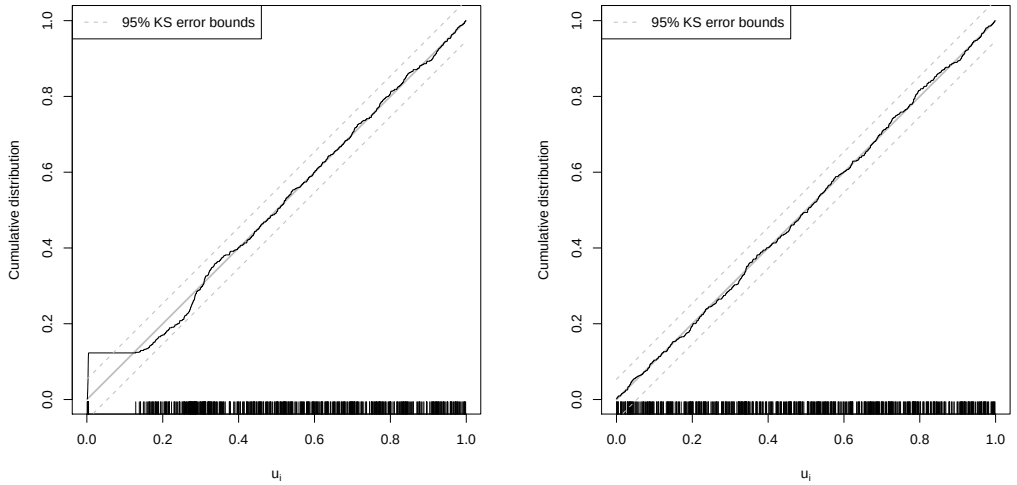
(a) (b)

Figure 4: (a) Trend and seasonal component of the fitted model; one observes the typical IMD peak in late February and minimum in August. Furthermore, (b) shows the spatial interaction function multiplied by the type modifier illustrating the higher epidemic potential of type B.

To inspect the goodness-of-fit of the selected spatio-temporal point process model, we follow the suggestion by Berman (1983) (see also Ogata, 1988; Rathbun, 1996) by computing

$$Y_i = \hat{\Lambda}_g^*(t_i) - \hat{\Lambda}_g^*(t_{i-1}), \quad i = 2, \ldots, n,$$

where $\hat{\Lambda}_g^*(t)$ is the fitted cumulative intensity function of the ground process. If the estimated CIF describes the true CIF well, then $U_i = 1 - \exp(-Y_i) \overset{\text{iid}}{\sim} U(0,1)$. Figure 5(a) contains a plot of the cumulative density function (CDF) of the observed U_i and for



(a) ϵ-scheme. (b) $U(0,1)$-scheme.

Figure 5: CDF of the observed U_i together with 95% Kolmogorov-Smirnov error bounds for data with tie breaking according to the (a) ϵ scheme and (b) $U(0,1)$ scheme.

comparison the CDF of the $U(0,1)$-distribution together with error bounds computed by inverting the one sample Kolmogorov-Smirnov test. The fit appears good, but noticable deviations for $u_i < 0.15$ can be observed, which we suspect to occur due to the tie-breaking strategy of subtracting $\epsilon = 0.01$ days from ties. As observations are on a per-day basis and thus are interval censored we re-estimated the model for a data set where ties were broken by subtracting a $U(0,1)$-distributed random number from each observation time. Figure 5(b) shows the improved fit of this analysis. The relative changes in the parameter estimates are minor and in the order of 0.98 - 1.07 except for the estimate $\hat{\gamma}_{\geq 19}$ where the change is of order 1.43. The "residual plots" in Figure 5 thus provided important insights with respect to the obtained fit and suggest reasonable goodness-of-fit. In addition, a scatterplot of the observed U_i against U_{i+1} (not shown) yields no evidence for serial correlation.

Altogether, we are led to the conclusion that the proposed model provides a useful description of the spread of IMD. It allows a quantification that the serogroup B finetype has a higher epidemic potential than the serogroup C finetype and shows, based on the German data, no spatio-temporal influence of influenza cases.

Another way of assessing the goodness-of-fit is by simulation from the fitted CIF. Figure 6 shows the observed 7-year incidences (per 100,000 inhabitants) of the 413 districts for both finetypes together. In order to identify extreme observations which are not explained by the selected model, we simulated 100 realisations of the process along the lines of Section 4, and determined the 2.5% and 97.5% quantiles of the district-specific 7-year incidences. In the figure, districts with observed incidences outside the simulated 95%-range are marked by triangles. There are 17 districts, which experienced outbreaks of IMD on top of the clustering described by the model. Most of these districts

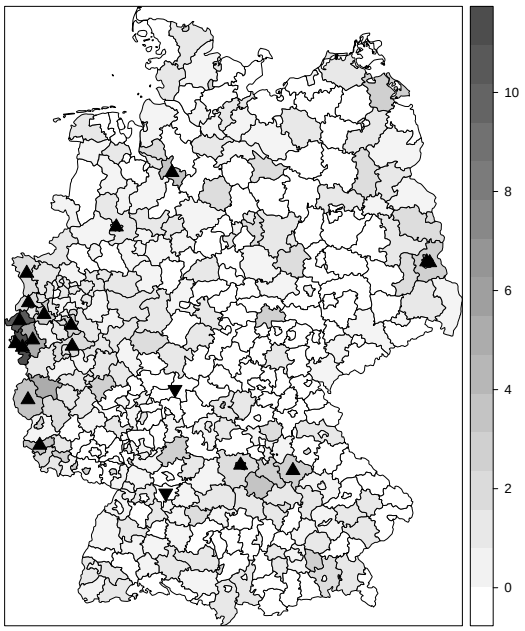


Figure 6: Observed incidence (per 100,000 inhabitants) during 2002-2008 for both finetypes together. Triangles pointing up (down) indicate districts with a higher (lower) incidence than explained by 100 simulations from the model.

are found in and around Aachen at the border to the Netherlands. The deviation from the model could thus be explained by edge effects hiding potential transmissions across the border and resulting in an underestimation of the epidemic weight.

6. Discussion

We presented a comprehensive framework for modelling, inference and simulation for general self-exciting spatio-temporal point patterns. Our motivating example originated from epidemic modelling, but the proposal is flexible enough to suit other applications, e.g. the modelling of earthquakes and forest fires. Using an additive-multiplicative decomposition of the conditional intensity function we divide modelling into endemic and epidemic parts. In the case of IMD the infected individual is effectively removed from the transmission network once the disease becomes manifest. Secondary cases are thought to acquire the infective strain either from the case during incubation or from asymptomatic carriers close to the case. While marks attached to the case can naturally not account for the latter mode of transmission, they represent a valid proxy for the transmission network of the case when analyzing surveillance data, which typically lack information regarding carriage.

It is our experience that parameters in the epidemic component can be hard to identify when looking at data sets with a limited number of direct disease transmissions. In

this case, the epidemic component should be modelled relatively simple and it might be worthwhile to investigate a histogram of the number of potential sources of the events. Identifiability was also the main reason for proposing independent space and time interaction functions – if data are plenty there is no conceptual hindrance to use a joint space-time interaction kernel and to compute its derivatives for inference. Also, we did only consider two artificial procedures for breaking tied event times for the analysis. More accurate analyses must carry out sensitivity analyses on the specific choices made with respect to breaking ties (e.g. use different randomisations), but residual plots offered useful illustrations for identifying such problems.

An issue currently not dealt with in our estimation are edge effects, i.e. data are only available for Germany, but infections occur outside the observation window. For example, 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 in Germany, which potentially causes underestimation of the epidemic weight. Edge correction for inference in spatio-temporal point processes is still an open methodological issue. For the IMD application we have too few cases to create a buffer zone, but the idea in Cronie (2010) of creating a buffer zone by simulation conditioned on the actual observed point pattern might be worth investigating. However, this would require an appropriate adaption of the simulation algorithm for performing conditional simulation. 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.

One way to address missing covariates or other unexplained heterogeneity is the use of random effects, aka frailties, in the endemic component. As in the discrete time – discrete space setting of Paul and Held (2010) or the survival setting in Kneib and Fahrmeir (2007), this could be done by a spatial Gaussian-Markov random field, or possibly by a space-continous Gaussian random field with inference by penalized likelihood. Similarly, smooth effects based on penalized splines would consistute a further extension fitting nicely into such a penalized likelihood framework, e.g. to describe a smooth baseline hazard function $h_0^{\text{temp}}(t)$ in the endemic component as in Höhle (2009) or flexible covariate effects as in Kneib and Fahrmeir (2007). However, from a computational point of view, maximization of the likelihood is already a daunting task. Inference for such extensions would complicate the likelihood maximization even further.

An additional strength of the proposed modelling is that it offers a parametric framework for conducting prospective change-point analysis in spatio-temporal point processes typical in disease surveillance: Within the framework of stochastic process control one could e.g. use likelihood ratio detectors to monitor the time point where inclusion of an epidemic component is necessary to describe the observed data. This would correspond in idea to the time series setting investigated in Höhle and Paul (2008) or the homogeneous spatio-temporal Poisson process setting of Assunçáo and Correa (2009).

The presented methods for inference and simulation of `twinstim` models will be made available as part of the `R`-package for epidemic modelling `RLadyBug` (Höhle and Feld-

mann, 2007; Höhle et al., 2010) available from the Comprehensive R Archive Network (CRAN).

Supplementary Materials

The Web Animation referenced in Section 5 is available from `http://www.statistik.lmu.de/~hoehle/pubs/twinstim/`.

Acknowledgements

We thank Ludwig Fahrmeir (University of Munich) for providing helpful suggestions and comments. Financial support was provided by the Munich Center of Health Sciences.

References

Anderson, R. M. and May, R. M. (1991). *Infectious diseases of humans*. Oxford University Press.

Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis*, volume 151 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Assunçáo, R. and Correa, T. (2009). Surveillance to detect emerging space-time clusters. *Computational Statistics & Data Analysis*, 53(8):2817–2830.

Berman, M. (1983). Comment on "Likelihood analysis of point processes and its applications to seismological data" by Y. Ogata. *Bulletin of the International Statistical Institute*, 44 (Book 3):412–418.

Claus, H., Maiden, M. C. J., Wilson, D. J., McCarthy, N. D., Jolley, K. A., Urwin, R., Hessler, F., Frosch, M., and Vogel, U. (2005). Genetic analysis of meningococci carried by children and young adults. *Journal of Infectious Diseases*, 191(8):1263–1271.

Cox, D. R. and Isham, V. (1988). A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London. Series A*, 415:317–328.

Cronie, O. (2010). Some edge correction methods for marked spatio-temporal point process models. Technical Report PREPRINT 2010:9, Department of Mathematical Sciences, Chalmers University of Technology, Sweden.

Daley, D. J. and Gani, J. (1999). *Epidemic Modelling: An introduction*. Cambridge University Press.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*, volume I: Elementary Theory and Methods of *Probability and its Applications*. Springer-Verlag, New York, 2nd edition.

Diggle, P., Rowlingson, B., and li Su, T. (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics*, 16(5):423–434.

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 Finkenstädt, B., Held, L., and Isham, V., editors, *Statistical Methods for Spatio-Temporal Systems*, pages 1–45. Chapman & Hall/CRC, Boca Raton.

Diggle, P. J., Kaimi, I., and Abellana, R. (2009). Partial-likelihood analysis of spatio-temporal point-process data. *Biometrics*, 66(2):347–354.

Elias, J., Schouls, L. M., van de Pol, I., Keijzers, W. C., Martin, D. R., Glennie, A., Oster, P., Frosch, M., Vogel, U., and van der Ende, A. (2010). Vaccine preventability of meningococcal clone, Greater Aachen Region, Germany. *Emerging Infectious Diseases*, 16(3):465–472.

Federal Statistical Office (DESTATIS) (2009). Gemeindeverzeichnis GV 2000. Districts as of 31/12/2008. Data as of 31/12/2007.

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

Held, L., Hofmann, M., Höhle, M., and Schmid, V. (2006). A two component model for counts of infectious diseases. *Biostatistics*, 7:422–437.

Held, L., Höhle, M., and Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance data. *Statistical Modelling*, 5:187–199.

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

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

Höhle, M., Feldmann, U., and Meyer, S. (2010). *`RLadyBug`: Analysis of Infectious Diseases using Stochastic Epidemic Models*. R package version 0.7-0.

Höhle, M. and Paul, M. (2008). Count data regression charts for the monitoring of surveillance time series. *Computational Statistics & Data Analysis*, 52(9):4357–4368.

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

Jewell, C. P., Kypraios, T., Neal, P., and Roberts, G. O. (2009). Bayesian analysis for emerging infectious diseases. *Bayesian Analysis*, 4(4):465–496.

Kneib, T. and Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, 34:207–228.

Lawson, A. B. and 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.

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

Meyer, S. (2009). Spatio-temporal infectious disease epidemiology based on point processes. Master's thesis, Department of Statistics, Ludwig-Maximilians-Universität, München. Available as `http://epub.ub.uni-muenchen.de/11703/`.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2010). Self-exciting point process modeling of crime. Technical report, Department of Mathematics and Computer Science, Santa Clara University.

Møller, J. and Díaz-Avalos, C. (2010). Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. *Scandinavian Journal of Statistics*, 37(1):2–25.

Neal, P. and Roberts, G. O. (2004). Statistical inference and model selection for the 1861 Hagelloch measles epidemic. *Biostatistics*, 5(2):249–261.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27.

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

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

Paul, M. and Held, L. (2010). Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. Preprint, Biostatistics Unit, Institute for Social and Preventive Medicine, University of Zurich, Switzerland.

Paul, M., Held, L., and Toschke, A. M. (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine*, 27:6250–6267.

Peng, R. D., Schoenberg, F. P., and Woods, J. A. (2005). A space-time conditional intensity model for evaluating a wildfire hazard index. *Journal of the American Statistical Association*, 100(469):26–35.

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.

Robert Koch-Institut (2009). SurvStat@RKI.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987). Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London. Series A*, 410(1839):269–288.

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

Scheel, I., Aldrin, M., Frigessi, A., and Jansen, P. A. (2007). A stochastic model for infectious salmon anemia (ISA) in Atlantic salmon farming. *Journal of the Royal Society, Interface*, 4(15):699–706.

Schoenberg, F. P., Brillinger, D. R., and Guttorp, P. M. (2002). Point processes, spatial-temporal. In El-Shaarawi, A. H. and Piegorsch, W. W., editors, *Encyclopedia of Environmetrics*, volume 3, pages 1573–1577. John Wiley & Sons, Ltd.

Stroud, A. H. (1971). *Approximate Calculation of Multiple Integrals*. Prentice Hall, Englewood Cliffs.

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

A. Calculus of the Score Function

Endemic intercept(s) $\boldsymbol{\beta_0}$:
Let $\beta_{0,k}$, $k \in \{1, \dots, K\}$ be one of the type-specific intercepts in $\boldsymbol{\beta_0}$. Then,

$$\frac{\partial}{\partial \beta_{0,k}} \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) = \mathbb{1}_{k=\kappa}(\kappa) \cdot \exp\left(\beta_{0,k} + o_{\xi(\boldsymbol{s})} + \boldsymbol{\beta}' \boldsymbol{z}_{\tau(t),\xi(\boldsymbol{s})}\right)$$

since the parameter $\beta_{0,k}$ appears in the endemic component $h_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa)$ if 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 (14). 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 $\boldsymbol{\gamma}$:

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

and the corresponding integral can be deduced similar to equation (15) as

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

Parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}$ of the interaction functions:
For a general spatial kernel $f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa)$,

$$\frac{\partial}{\partial \boldsymbol{\sigma}} \lambda^*_{\boldsymbol{\theta}}(t, \boldsymbol{s}, \kappa) = \sum_{j \in I^*(t,\boldsymbol{s},\kappa)} e^{\eta_j} \, g_{\boldsymbol{\alpha}}(t - t_j|\kappa_j) \, \left[\frac{\partial}{\partial \boldsymbol{\sigma}} f_{\boldsymbol{\sigma}}(\boldsymbol{s} - \boldsymbol{s}_j|\kappa_j)\right]$$

with corresponding integral

$$\sum_{j=1}^{n} q_{\kappa_j,\bullet}\, e^{\eta_j} \left[\int_0^{\min\{T-t_j;\varepsilon\}} g_{\boldsymbol{\alpha}}(t|\kappa_j)\,\mathrm{d}t \right] \left[\int_{R_j} \frac{\partial}{\partial \boldsymbol{\sigma}} f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa_j)\,\mathrm{d}\boldsymbol{s} \right].$$

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

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \lambda^*_{\boldsymbol{\theta}}(t,\boldsymbol{s},\kappa) = \sum_{j \in I^*(t,\boldsymbol{s},\kappa)} e^{\eta_j} \left[\frac{\partial}{\partial \boldsymbol{\alpha}} g_{\boldsymbol{\alpha}}(t-t_j|\kappa_j) \right] f_{\boldsymbol{\sigma}}(\boldsymbol{s}-\boldsymbol{s}_j|\kappa_j)$$

with corresponding integral

$$\sum_{j=1}^{n} q_{\kappa_j,\bullet}\, e^{\eta_j} \left[\int_0^{\min\{T-t_j;\varepsilon\}} \frac{\partial}{\partial \boldsymbol{\alpha}} g_{\boldsymbol{\alpha}}(t|\kappa_j)\,\mathrm{d}t \right] \left[\int_{R_j} f_{\boldsymbol{\sigma}}(\boldsymbol{s}|\kappa_j)\,\mathrm{d}\boldsymbol{s} \right].$$