

Contents lists available at [ScienceDirect](#)

# Spatial Statistics

journal homepage: [www.elsevier.com/locate/spasta](http://www.elsevier.com/locate/spasta)

## A multi-scale area-interaction model for spatio-temporal point patterns

Adina Iftimi<sup>a,\*</sup>, Marie-Colette van Lieshout<sup>b,c</sup>,  
Francisco Montes<sup>d</sup>

<sup>a</sup> Department of Bioscience and Nutrition, Karolinska Institutet, Stockholm, Sweden

<sup>b</sup> CWI, P.O. Box 94079, NL-1090 GB Amsterdam, The Netherlands

<sup>c</sup> Department of Applied Mathematics, University of Twente, P.O. Box 217, NL-7500 AE Enschede, The Netherlands

<sup>d</sup> Department of Statistics and Operations Research, University of Valencia, Doctor Moliner, 50, 46100 Burjassot-Valencia, Spain

### ARTICLE INFO

#### Article history:

Received 26 July 2017

Accepted 5 June 2018

Available online 19 June 2018

#### Keywords:

Gibbs point processes

Multi-scale area-interaction model

Spatio-temporal point processes

Varicella

### ABSTRACT

Models for fitting spatio-temporal point processes should incorporate spatio-temporal inhomogeneity and allow for different types of interaction between points (clustering or regularity). This paper proposes an extension of the spatial multi-scale area-interaction model to a spatio-temporal framework. This model allows for interaction between points at different spatio-temporal scales and for the inclusion of covariates. We present a simulation study and fit the new model to varicella cases registered during 2013 in Valencia, Spain.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Spatio-temporal patterns are increasingly observed in many different fields, including ecology, epidemiology, seismology, astronomy and forestry. The common feature is that all observed events have two basic characteristics: the location and the time of the event. In this paper we are mainly concerned with epidemiology (Stallybrass, 1931), which studies the distribution, causes and control of diseases in a defined human population. The locations of the occurrence of cases give information on the spatial behavior of the disease, whereas the times, measured on different scales (days, weeks, years, period of times), give insights on the temporal response of the overall process. An essential point

\* Corresponding author.

E-mail address: [iftimi@uv.es](mailto:iftimi@uv.es) (A. Iftimi).

to take into consideration is that people are not uniformly distributed in space, hence information on the spatial distribution of the population at risk is crucial when analyzing spatio-temporal patterns of diseases.

Realistic models to fit epidemiological data should incorporate spatio-temporal inhomogeneity and allow for different types of dependence between points. One important class of such models is the family of Gibbs point processes, defined in terms of their probability density function (van Lieshout, 2000; Ripley, 1988, 1990), and, in particular, the sub-class of pairwise interaction processes. Well-known examples of pairwise interaction processes are the Strauss model (Kelly and Ripley, 1976; Strauss, 1975) or the hard core process, a particular case of the Strauss model where no points ever come closer to each other than a given threshold. However, pairwise interaction models are not always a suitable choice for fitting clustered patterns. A family of Markov point processes that can fit both clustered and inhibitory patterns is that of the area- or quermass-interaction models (Baddeley and van Lieshout, 1995; Kendall et al., 1999). These models are defined in terms of stochastic geometric functionals and display interactions of all orders. Methods for inference and perfect simulation are available in Dereudre et al. (2014), Häggström et al. (1999), Kendall (2000) and Møller and Helisová (2010).

Most natural processes exhibit interaction at multiple scales. The classical Gibbs processes model spatial interaction at a single scale, nevertheless *multi-scale* generalizations have been proposed in the literature (Ambler and Silverman, 2010; Gregori et al., 2003; Picard et al., 2009). In this paper we propose an extension of the spatial multi-scale area-interaction model to a spatio-temporal framework.

The outline of the paper is as follows. Section 2 provides some preliminaries in relation to notation and terminology. Section 3 gives the definition and Markov properties of our spatio-temporal multi-scale area-interaction model. Section 4 adapts simulation algorithms, such as the Metropolis–Hastings algorithm, to our context. Section 5 treats the logistic regression approach and presents a simulation study. The model is applied to a varicella data set in Section 6. Section 7 presents final remarks and a discussion of future work.

## 2. Preliminaries

A realization of a spatio-temporal point process  $X$  consists of a finite number  $n \geq 0$  of distinct points  $(x_i, t_i)$ ,  $i = 1, \dots, n$ , that are observed within a compact spatial domain  $W_S \subset \mathbb{R}^2$  and time interval  $W_T \subset \mathbb{R}$ . The pattern formed by the points will be denoted by  $\mathbf{x} = \{(x_i, t_i)\}_{i=1}^n$ . For a mathematically rigorous account, the reader is referred to Daley and Vere-Jones (2003, 2008).

We define the Euclidean norm  $\|x\| = (x_1^2 + x_2^2)^{1/2}$  and the Euclidean metric  $d_{\mathbb{R}^2}(x, y) = \|x - y\|$  for  $x = (x_1, x_2) \in \mathbb{R}^2$  and  $y = (y_1, y_2) \in \mathbb{R}^2$ . We need to treat space and time differently, thus on  $\mathbb{R}^2 \times \mathbb{R}$  we consider the supremum norm  $\|(x, t)\|_{\infty} = \max\{\|x\|, |t|\}$  and the supremum metric  $d((x, t), (y, s)) = \|(x, t) - (y, s)\|_{\infty} = \max\{\|x - y\|, |t - s|\}$ , where  $(x, t), (y, s) \in \mathbb{R}^2 \times \mathbb{R}$ . Note that  $(\mathbb{R}^2 \times \mathbb{R}, d(\cdot, \cdot))$  as well as its restriction to  $W_S \times W_T$  is a complete, separable metric space. We write  $\mathcal{B}(\mathbb{R}^2 \times \mathbb{R}) = \mathcal{B}(\mathbb{R}^2) \otimes \mathcal{B}(\mathbb{R})$  for the Borel  $\sigma$ -algebra and  $\ell$  for Lebesgue measure. We denote by  $\oplus$  the Minkowski addition of two sets  $A, B \subset \mathbb{R}^2$ , defined as the set  $A \oplus B = \{a + b : a \in A, b \in B\}$ .

As stated in Section 1, Gibbs models form an important class of models able to fit epidemiological data exhibiting spatio-temporal inhomogeneity and interaction between points. In space, the Widom–Rowlinson *penetrable sphere model* (Widom and Rowlinson, 1970) produces clustered point patterns; the more general area-interaction model (Baddeley and van Lieshout, 1995) fits both clustered and inhibitory point patterns. In its most simple form, the area-interaction model is defined by its probability density

$$p(\mathbf{x}) = \alpha \lambda^{n(\mathbf{x})} \gamma^{-A(\mathbf{x})} \quad (1)$$

with respect to a unit rate Poisson process on  $W_S$ . Here  $\alpha$  is the normalizing constant,  $\mathbf{x}$  is a spatial point configuration in  $W_S \subset \mathbb{R}^2$ ,  $n(\mathbf{x})$  is the cardinality of  $\mathbf{x}$  and  $A(\mathbf{x})$  is the area of the union of discs of radius  $r$  centered at  $x_i \in \mathbf{x}$  restricted to  $W_S$ . The positive scalars  $\lambda$ ,  $\gamma$  and  $r > 0$  are the parameters of the model. Note that, as emphasized in van Lieshout (2000), Gibbsian interaction terms can be combined to yield more complex models. Doing so, Ambler and Silverman (2010), Gregori et al. (2003) and Picard

et al. (2009) develop an extension of the area-interaction process which incorporates both inhibition and attraction. We propose a further generalization of the area-interaction model to allow multi-scale interaction in a spatio-temporal framework.

### 3. Space–time area-interaction processes

Let  $\mathbf{x}$  be a finite spatio-temporal point configuration on  $W_S \times W_T \subset \mathbb{R}^2 \times \mathbb{R}$ , that is, a finite set of points, including the empty set.

**Definition 1.** The *spatio-temporal multi-scale area-interaction process* is the point process with density

$$p(\mathbf{x}) = \alpha \prod_{(x,t) \in \mathbf{x}} \lambda(x,t) \prod_{j=1}^m \gamma_j^{-\ell(\mathbf{x} \oplus G_j)} \tag{2}$$

with respect to a unit rate Poisson process on  $W_S \times W_T$ , where  $\alpha > 0$  is a normalizing constant,  $\lambda \geq 0$  is a measurable and bounded function,  $\ell$  is Lebesgue measure restricted to  $W_S \times W_T$ ,  $\gamma_j > 0$  are the interaction parameters,  $G_j$  are some compact subsets of  $\mathbb{R}^2 \times \mathbb{R}$  with size depending on  $j, j = 1, \dots, m, m \in \mathbb{N}$ , where  $G_1 \subset G_2 \subset \dots \subset G_m$ , and  $\oplus$  denotes Minkowski addition.

Note that when  $\mathbf{x}$  is the empty set,  $p(\mathbf{x}) = \alpha$ . The interaction parameters have the same interpretation as for the spatial area-interaction model (1). For fixed  $j \in \{1, \dots, m\}$ , when  $0 < \gamma_j < 1$  we would expect to see *inhibition* between points at spatio-temporal scales determined by the definition of the compact set  $G_j$ . On the other hand, when  $\gamma_j > 1$  we expect *clustering* between the points. We observe that (2) reduces to an inhomogeneous Poisson process when  $\gamma_j = 1$  for all  $j \in \{1, \dots, m\}$ .

Covariates can be introduced in the model by letting the function  $\lambda$  be a measurable and bounded function  $\lambda(x,t) = \rho(Z(x,t))$  of the covariate vector  $Z(x,t)$ .

The new model proposed in (2) successfully extends the area-interaction model to multi-scale interaction for spatio-temporal point patterns.

**Lemma 1.** The density (2) is measurable and integrable for all  $\gamma_j, j = 1, \dots, m, m \in \mathbb{N}$ .

**Proof.** Consider a point configuration,  $\mathbf{x}$ . Since  $\ell$  is  $\sigma$ -finite and  $G_j$  is compact, the map  $\mathbf{x} \mapsto \ell(\mathbf{x} \oplus G_j)$  is measurable for any  $j = 1, \dots, m$ . It follows that the map  $\mathbf{x} \mapsto \exp[-\ell(\mathbf{x} \oplus G_j) \log \gamma_j]$  is measurable for any  $j = 1, \dots, m$ . The map  $\mathbf{x} \mapsto \prod_{(x_i,t_i) \in \mathbf{x}} \lambda(x_i,t_i)$  is also measurable by assumption, hence the density (2) is measurable.

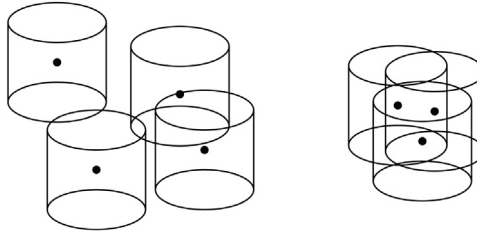
To determine if (2) is integrable, we observe that  $0 \leq \ell(\mathbf{x} \oplus G_j) \leq \ell(W_S \times W_T) < \infty$ . The function  $\lambda$  is integrable by assumption, hence (2) is dominated by an integrable function, and therefore integrable.

As a further simplification, for fixed  $j \in \{1, \dots, m\}$ , consider the case where  $\mathbf{x} \oplus G_j = \bigcup_{(x,t) \in \mathbf{x}} C_{r_j}^{t_j}(x,t)$  is the union of all cylinders with radius  $(r_j, t_j)$  centered in  $(x,t)$  taken over all  $(x,t) \in \mathbf{x}$ . We define the cylinder with radius  $(r_j, t_j)$  by  $C_{r_j}^{t_j}(x,t) = \{(y,s) \in W_S \times W_T : \|x-y\| \leq r_j, |t-s| \leq t_j\}$ . Note that we could consider only the upper half of the cylinder if we would not like to look backwards in time, thus distinguish future from past. Alternatively, a cone could be used rather than a cylinder. This would be appropriate in situations where the spatial range of interaction decreases with time. In this paper, we consider a symmetric cylinder, thus treating space and time identically, taking into account both past and future.

Expression (2) reads

$$p(\mathbf{x}) = \alpha \prod_{(x,t) \in \mathbf{x}} \lambda(x,t) \prod_{j=1}^m \gamma_j^{-\ell(\bigcup_{(x,t) \in \mathbf{x}} C_{r_j}^{t_j}(x,t))}, \tag{3}$$

where  $(r_j, t_j)$  are pairs of *irregular parameters* (Baddeley et al., 2016) of the model and  $\gamma_j$  are interaction parameters,  $j = 1, \dots, m$ . The function  $\lambda$  is here assumed known for simplicity, but could also depend on further parameters.



**Fig. 1.** An illustration of possible  $\mathbf{x} \oplus G$  (cylinders around the points), where the black dots represent points of the process.

Fig. 1 shows an illustration of  $\mathbf{x} \oplus G_j$ . When  $0 < \gamma_j < 1$ , point configurations such as the one on the left are likely to be observed (inhibition between points), whereas for large  $\gamma_j > 1$ , point configurations such as the one on the right are more likely to be observed (attraction between points).

### 3.1. Markov properties

Let  $\sim$  on  $\mathbb{R}^2 \times \mathbb{R}$  be a symmetric and reflexive relation on  $\mathbb{R}^2 \times \mathbb{R}$ , i.e. for any  $(x, t), (y, s) \in \mathbb{R}^2 \times \mathbb{R}$ ,  $(x, t) \sim (y, s) \Leftrightarrow (y, s) \sim (x, t)$  and  $(x, t) \sim (x, t)$ . Two points  $(x, t)$  and  $(y, s)$  are said to be *neighbors* if  $(x, t) \sim (y, s)$ .

The fixed range relation at range  $r$  is defined as

$$(x, t) \sim (y, s) \Leftrightarrow \|(x, t) - (y, s)\|_\infty \leq r \Leftrightarrow C_{r/2}^{r/2}(x, t) \cap C_{r/2}^{r/2}(y, s) \neq \emptyset. \tag{4}$$

**Definition 2.** A point process has the *Markov property* (van Lieshout, 2000; Ripley and Kelly, 1977) with respect to the symmetric, reflexive relation  $\sim$ , if, for all point configurations  $\mathbf{x}$  with  $p(\mathbf{x}) > 0$ , the following conditions are fulfilled:

1.  $p(\mathbf{y}) > 0$  for all  $\mathbf{y} \subset \mathbf{x}$ ;
2. the likelihood ratio  $\frac{p(\mathbf{x} \cup \{(y, s)\})}{p(\mathbf{x})}$  for adding a new point  $(y, s) \notin \mathbf{x}$  to a point configuration  $\mathbf{x}$  depends only on points  $(x, t) \in \mathbf{x}$  such that  $(y, s) \sim (x, t)$ , i.e. depends only on the neighbors of  $(y, s)$ .

**Lemma 2.** The spatio-temporal multi-scale area-interaction process (2) is a Markov point process with respect to the relation (4) in the sense of Ripley and Kelly (1977).

**Proof.** Note that if  $p(\mathbf{x}) > 0$ , since  $\lambda(x, t) > 0$  for all  $(x, t) \in \mathbf{x}$ , then whenever  $\mathbf{y} \subset \mathbf{x}$ , also  $p(\mathbf{y}) > 0$ . The likelihood ratio

$$\begin{aligned} \frac{p(\mathbf{x} \cup \{(y, s)\})}{p(\mathbf{x})} &= \frac{\alpha \left( \prod_{(x,t) \in \mathbf{x}} \lambda(x, t) \right) \lambda(y, s) \prod_{j=1}^m \gamma_j^{-\ell((\mathbf{x} \cup \{(y, s)\}) \oplus G_j)}}{\alpha \prod_{(x,t) \in \mathbf{x}} \lambda(x, t) \prod_{j=1}^m \gamma_j^{-\ell(\mathbf{x} \oplus G_j)}} \\ &= \lambda(y, s) \prod_{j=1}^m \gamma_j^{-\ell((y, s) \oplus G_j) \setminus (\mathbf{x} \oplus G_j)}. \end{aligned} \tag{5}$$

Note that

$$\begin{aligned} ((y, s) \oplus G_j) \setminus (\mathbf{x} \oplus G_j) &= ((y, s) \oplus G_j) \cap \left[ \bigcup_{(x,t) \in \mathbf{x}} (x, t) \oplus G_j \right]^c \\ &= ((y, s) \oplus G_j) \cap \left[ \bigcup_{(x,t) \sim (y,s) \in \mathbf{x}} (x, t) \oplus G_j \right]^c, \quad \forall j = 1, \dots, m. \end{aligned}$$

Thus (5) depends only on the newly added point  $(y, s)$  and its neighbors. Hence (2) defines a Markov point process with respect to  $\sim$ .

It follows that the density  $p(\cdot)$  in (3) is Markov at range  $2 \max\{r_m, t_m\}$ , see (4).

Define the Papangelou conditional intensity of a point process with density  $p$  by

$$\lambda((y, s); \mathbf{x}) = \frac{p(\mathbf{x} \cup \{(y, s)\})}{p(\mathbf{x})},$$

whenever  $p(\mathbf{x}) > 0$  and  $(y, s) \notin \mathbf{x}$ . Then, for the spatio-temporal multi-scale area-interaction process, by the proof of Lemma 2 we obtain that

$$\lambda((y, s); \mathbf{x}) = \lambda(y, s) \prod_{j=1}^m \gamma_j^{-\ell(C_{r_j}^{t_j}(y, s) \setminus \bigcup_{(x, t) \in \mathbf{x}} C_{r_j}^{t_j}(x, t))}, \tag{6}$$

or, upon transformation to a logarithmic scale,

$$\log \lambda((y, s); \mathbf{x}) = \log \lambda(y, s) - \sum_{j=1}^m (\log \gamma_j) \ell \left( C_{r_j}^{t_j}(y, s) \setminus \bigcup_{(x, t) \in \mathbf{x}} C_{r_j}^{t_j}(x, t) \right).$$

Note that  $\lambda(y, s)$  may be 0, thus making  $\log \lambda(y, s)$  ill-defined.

Write  $\eta_j = \log \gamma_j$ . Then, whenever well-defined,

$$\begin{aligned} \log \lambda((y, s); \mathbf{x}) &= \log \lambda(y, s) - \sum_{j=1}^m \eta_j \int_{W_S \times W_T} \mathbf{1}\{(z, u) \in C_{r_j}^{t_j}(y, s) \setminus \bigcup_{(x, t) \in \mathbf{x}} C_{r_j}^{t_j}(x, t)\} dz du \\ &= \log \lambda(y, s) - \sum_{j=1}^m \int_{F_{r_j}^{t_j}(y, s)} \sum_{i=j}^m \eta_i \mathbf{1}\{(z, u) \notin \bigcup_{(x, t) \in \mathbf{x}} C_{r_i}^{t_i}(x, t)\} dz du, \end{aligned} \tag{7}$$

where  $F_{r_j}^{t_j}(y, s)$  is the difference between two concentric cylinders  $C_{r_j}^{t_j}(y, s)$  and  $C_{r_{j-1}}^{t_{j-1}}(y, s)$ .

Indeed,

$$\begin{aligned} F_{r_j}^{t_j}(y, s) &= C_{r_j}^{t_j}(y, s) \setminus C_{r_{j-1}}^{t_{j-1}}(y, s) \\ &= \left\{ (y, s) \in W_S \times W_T : \begin{array}{l} r_{j-1} < \|x - y\| \leq r_j, |t - s| \leq t_{j-1} \text{ or} \\ r_{j-1} < \|x - y\| \leq r_j, t_{j-1} < |t - s| \leq t_j \text{ or} \\ \|x - y\| \leq r_{j-1}, t_{j-1} < |t - s| \leq t_j \end{array} \right\}. \end{aligned}$$

We assume the cylinders are nested, thus  $0 = r_0 < r_1 < \dots < r_m$  and  $0 = t_0 < t_1 < \dots < t_m$ . The left-most panel of Fig. 2 shows an illustration of  $F_{r_j}^{t_j}$  for fixed  $r_j, t_j$ . The blue annulus corresponds to  $\{(y, s) \in W_S \times W_T : r_{j-1} < \|x - y\| \leq r_j, |t - s| \leq t_{j-1}\}$ , the two green annuli represent  $\{(y, s) \in W_S \times W_T : r_{j-1} < \|x - y\| \leq r_j, t_{j-1} < |t - s| \leq t_j\}$  and the two red cylinders form  $\{(y, s) \in W_S \times W_T : \|x - y\| \leq r_{j-1}, t_{j-1} < |t - s| \leq t_j\}$ .

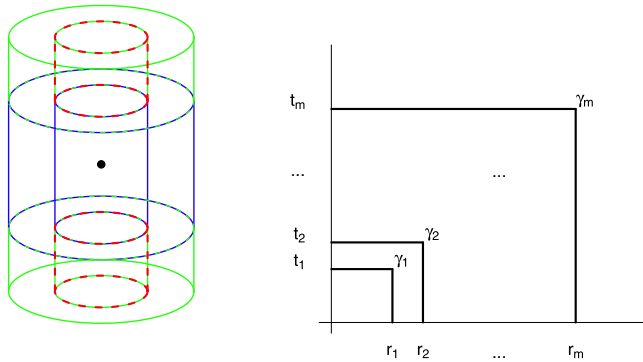
If, for  $(y, s), \|y - x\| > 2r_m, |s - t| > 2t_m, \forall (x, t) \in \mathbf{x}$ , then

$$\begin{aligned} \log \lambda((y, s); \mathbf{x}) &= \log \lambda(y, s) - \sum_{j=1}^m \left( \sum_{i=j}^m \eta_i \right) \ell(F_{r_j}^{t_j}(y, s)) \\ &= \log \lambda(y, s) - \sum_{j=1}^m \eta_j \ell(C_{r_j}^{t_j}(y, s)). \end{aligned}$$

To conclude this discussion, note that, in accordance with (Gregori et al., 2003),

$$p(\mathbf{x}) = \alpha \prod_{(x, t) \in \mathbf{x}} \lambda(x, t) \exp\left[- \sum_{j=1}^m \alpha_j \ell(F_{r_j}^{t_j}(\mathbf{x}))\right], \tag{8}$$

where  $\alpha_j = \sum_{i \geq j} \eta_i$  and  $F_{r_j}^{t_j}(\mathbf{x}) = (\mathbf{x} \oplus G_j) \setminus (\mathbf{x} \oplus G_{j-1})$ . As before,  $G_j = C_{r_j}^{t_j}(0, 0)$ .



**Fig. 2.** (Left) An illustration of  $F_{r_j}^{t_j}$  where the blue annulus corresponds to  $\{(y, s) \in W_S \times W_T : r_{j-1} < \|x - y\| \leq r_j, |t - s| \leq t_{j-1}\}$ , the two green annuli represent  $\{(y, s) \in W_S \times W_T : r_{j-1} < \|x - y\| \leq r_j, t_{j-1} < |t - s| \leq t_j\}$  and the two red cylinders are  $\{(y, s) \in W_S \times W_T : \|x - y\| \leq r_{j-1}, t_{j-1} < |t - s| \leq t_j\}$ . (Right) Multi-scale behavior. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In particular, when  $r_j = t_j$ , we get that

$$p(\mathbf{x}) = \alpha \prod_{(x,t) \in \mathbf{x}} \lambda(x, t) \exp \left[ - \sum_{j=1}^m \alpha_j \int_{W_S} \int_{W_T} 1\{d_\infty((z, u), \mathbf{x}) \in (r_{j-1}, r_j]\} dz du \right],$$

so a point  $(z, u)$  at  $d_\infty$ -distance to  $\mathbf{x}$  in the interval  $(r_{j-1}, r_j]$  contributes  $-\alpha_j$  to the exponent (Gregori et al., 2003).

The model in (3) with Papangelou conditional intensity defined by (7) allows for interaction behavior that varies across spatio-temporal scales, for example, inhibition at small scales, attraction at larger scales and randomness beyond. The different spatio-temporal scales,  $(r_j, t_j)$ , are defined according to  $F_{r_j}^{t_j}$ . Indeed, a point  $(z, u)$  in  $F_{r_j}^{t_j}(\mathbf{x})$  contributes a term  $\alpha_j$  to the energy (the negative of the exponential term) in  $p(\mathbf{x})$ . The right-most panel of Fig. 2 shows a visual representation of this multi-scale behavior.

An important property of Markov densities is the fact that the Papangelou conditional intensity,  $\lambda((y, s); \mathbf{x})$ , depends only on  $(y, s)$  and its neighbors in  $\mathbf{x}$ , and is computationally convenient. This property will be exploited in the next section to design simulation algorithms for generating realizations of the model.

#### 4. Simulation. The Metropolis–Hastings algorithm

Consider a Markov point process on  $W_S \times W_T \subset \mathbb{R}^2 \times \mathbb{R}$  defined by its density  $p(\cdot)$ . The *Metropolis–Hastings algorithm*, first introduced in statistical physics (Barker, 1965; Metropolis et al., 1953), is a tool for constructing a Markov process with limit distribution defined by  $p(\cdot)$ .

Metropolis–Hastings algorithms are *discrete time Markov* processes where transitions are defined as the proposal of a new state that is accepted or rejected based on the likelihood of the proposed state compared with the old state. We consider two types of proposals: addition (*birth*) and deletion (*death*) of a point. The likelihood ratio of the new state in comparison with the old state, for these type of transitions, is the (reciprocal) conditional intensity.

More precisely, consider the point configuration  $\mathbf{x}$ . We can propose either a birth or a death with respective probabilities  $q(\mathbf{x})$  and  $1 - q(\mathbf{x})$  that depend on  $\mathbf{x}$ . For a birth, a new point  $u \in W_S \times W_T$  is sampled from a probability density  $b(\mathbf{x}, \cdot)$  and the new point configuration  $\mathbf{x} \cup \{u\}$  is accepted with probability  $A(\mathbf{x}, \mathbf{x} \cup \{u\})$ , otherwise the state remains unchanged,  $\mathbf{x}$ . For a death, the point  $x \in \mathbf{x}$  chosen to be eliminated is selected according to a discrete probability distribution  $d(\mathbf{x}, \cdot)$  on  $\mathbf{x}$ , and the proposal  $\mathbf{x} \setminus \{x\}$  is accepted with probability  $A(\mathbf{x}, \mathbf{x} \setminus \{x\})$ , otherwise the state remains unchanged.

In general, we can choose  $b(\cdot, \cdot)$ ,  $d(\cdot, \cdot)$  and  $q(\cdot)$  as we prefer. However, an important condition to consider is that of *detailed balance*, and therefore time-reversibility of the Markov process,

$$q(\mathbf{x}) b(\mathbf{x}, u) A(\mathbf{x}, \mathbf{x} \cup \{u\}) p(\mathbf{x}) = (1 - q(\mathbf{x} \cup \{u\})) d(\mathbf{x} \cup \{u\}, u) A(\mathbf{x} \cup \{u\}, \mathbf{x}) p(\mathbf{x} \cup \{u\}). \tag{9}$$

For simplicity, consider the case that births and deaths are equally likely and sampled uniformly, that is,  $q \equiv 1/2$ ,  $b \equiv 1/\ell(W_S \times W_T)$  and  $d(\mathbf{x}, \cdot) = 1/n(\mathbf{x})$ , where  $n(\mathbf{x})$  is the number of points in the point configuration  $\mathbf{x}$ . Then, provided  $p(\mathbf{x}) > 0$ , (9) reduces to

$$\begin{aligned} \frac{1}{2} \frac{1}{\ell(W_S \times W_T)} A(\mathbf{x}, \mathbf{x} \cup \{u\}) p(\mathbf{x}) &= \left(1 - \frac{1}{2}\right) \frac{1}{n(\mathbf{x}) + 1} A(\mathbf{x} \cup \{u\}, \mathbf{x}) p(\mathbf{x} \cup \{u\}) \\ \frac{1}{\ell(W_S \times W_T)} A(\mathbf{x}, \mathbf{x} \cup \{u\}) p(\mathbf{x}) &= \frac{1}{n(\mathbf{x}) + 1} A(\mathbf{x} \cup \{u\}, \mathbf{x}) p(\mathbf{x} \cup \{u\}) \\ \frac{A(\mathbf{x}, \mathbf{x} \cup \{u\})}{A(\mathbf{x} \cup \{u\}, \mathbf{x})} &= \underbrace{\frac{\ell(W_S \times W_T)}{n(\mathbf{x}) + 1}}_{=r(\mathbf{x}, u)} \times \frac{p(\mathbf{x} \cup \{u\})}{p(\mathbf{x})}. \end{aligned}$$

Thus, more likely configurations can be favored by setting  $A(\mathbf{x}, \mathbf{x} \cup \{u\}) = \min\{1, r(\mathbf{x}, u)\}$ , and  $A(\mathbf{x} \cup \{u\}, \mathbf{x}) = \min\{1, 1/r(\mathbf{x}, u)\}$ . Therefore, using Eq. (6), for the spatio-temporal multi-scale area-interaction process (3), the ratio  $r(\mathbf{x}, u)$  for  $u = (y, s)$  reduces to

$$r(\mathbf{x}, u) = \frac{\ell(W_S \times W_T)}{n(\mathbf{x}) + 1} \lambda(y, s) \prod_{j=1}^m \gamma_j^{-\ell(c_{ij}^t(y,s) \setminus \cup_{(x,t) \in \mathbf{x}} c_{ij}^t(x,t))}. \tag{10}$$

In practice, we will use the logarithmic form of the conditional intensity as given in Eq. (7). When the region  $W_S \times W_T$  is irregular we use rejection sampling to generate a point uniformly at random from  $W_S \times W_T$ .

**5. Inference**

*5.1. A logistic regression estimating function approach*

Suppose  $X$  is a spatio-temporal point process on  $W_S \times W_T$  whose distribution is given by a density  $p_\theta$  with respect to the law of a unit rate Poisson process on  $W_S \times W_T$ . As before,  $W_S \subset \mathbb{R}^2$  is compact,  $W_T \subset \mathbb{R}$  is compact and  $\theta$  is an unknown parameter to be estimated.

A general approach to estimate  $\theta$  is the Takacs–Fiksel method (Takacs, 1986; Takacs and Fiksel, 1986). It relies on the Georgii–Nguyen–Zessin equation (Georgii, 1976; Nguyen and Zessin, 1979)

$$\mathbb{E} \left[ \sum_{x \in X \cap W_S \times W_T} h(x, X \setminus \{x\}) \right] = \mathbb{E} \left[ \int_{W_S} \int_{W_T} h(x, X) \lambda_\theta(x; X) dx \right] \tag{11}$$

holding for all (measurable) non-negative functions  $h$  (in the sense that the left-hand side is  $\infty$  whenever the right-hand side is and vice versa). The Papangelou conditional intensity  $\lambda_\theta(x; X)$  depends on the parameter  $\theta$ .

The Takacs–Fiksel idea is to choose a function  $h$ , estimate the left-hand side as well as the right-hand side based on an observation  $\mathbf{x}$  of  $X$  in  $W_S \times W_T$  and equate the two. The equation will involve  $\theta$ , so solving for  $\theta$  yields an estimator  $\hat{\theta}$ . Clearly, if  $\theta$  is a vector, a system of equations is needed, that is, several choices of  $h$ .

Since the function  $h$  may be chosen freely, (11) is quite flexible and includes the maximum pseudo-likelihood (Besag, 1977; Jensen and Møller, 1991) and the logistic regression (Baddeley et al., 2014) methods. Indeed, maximum pseudo-likelihood corresponds to (11) with

$$h(x, X) = \frac{\partial}{\partial \theta} \log \lambda_\theta(x; X),$$

the gradient of the log conditional intensity.

Logistic regression is based on (11) with

$$h(x, X) = \frac{\partial}{\partial \theta} \log \left[ \frac{\lambda_\theta(x; X)}{\lambda_\theta(x; X) + \rho(x)} \right] \tag{12}$$

where  $\rho : W_S \times W_T \rightarrow \mathbb{R}^+$  is an arbitrary (measurable) function.

We have experimented (Iftimi et al., 2017) with the maximum pseudo-likelihood approach, just as (Ambler and Silverman, 2010) did, but found the logistic regression approach to perform better. Naturally, in both instances, we assume  $\lambda_\theta$  is differentiable. The left-hand side of (11) with  $h$  given by (12) may be straightforwardly estimated by

$$\sum_{x \in \mathbf{x}} h(x, \mathbf{x} \setminus \{x\}) = \sum_{x \in \mathbf{x}} \frac{\rho(x)}{\lambda_\theta(x; \mathbf{x} \setminus \{x\}) + \rho(x)} \frac{\frac{\partial}{\partial \theta} \lambda_\theta(x; \mathbf{x} \setminus \{x\})}{\lambda_\theta(x; \mathbf{x})}.$$

For the right-hand side of (11), note that

$$\begin{aligned} \mathbb{E} \left[ \int_{W_S} \int_{W_T} h(x; X) \lambda_\theta(x; X) dx \right] &= \mathbb{E} \left[ \int_{W_S} \int_{W_T} \frac{\rho(x)}{\lambda_\theta(x; X) + \rho(x)} \frac{\partial}{\partial \theta} \lambda_\theta(x; X) dx \right] = \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \sum_{u \in D} \frac{\frac{\partial}{\partial \theta} \lambda_\theta(u; X)}{\lambda_\theta(u; X) + \rho(u)} \middle| X \right] \right] \end{aligned}$$

for any point process  $D$  independent of  $X$  on  $W_S \times W_T$  having intensity function  $\rho(\cdot)$  by the Campbell-Mecke formula (Chiu et al., 2013). We conclude that the right-hand side may be estimated by

$$\sum_{u \in d} \frac{\frac{\partial}{\partial \theta} \lambda_\theta(u; \mathbf{x})}{\lambda_\theta(u; \mathbf{x}) + \rho(u)}.$$

For further details, see Baddeley et al. (2014).

Thus we arrive at the estimating equation

$$\sum_{x \in \mathbf{x}} \frac{\rho(x)}{\lambda_\theta(x; \mathbf{x} \setminus \{x\}) + \rho(x)} \frac{\nabla \lambda_\theta(x; \mathbf{x} \setminus \{x\})}{\lambda_\theta(x; \mathbf{x} \setminus \{x\})} - \sum_{x \in d} \frac{\nabla \lambda_\theta(x; \mathbf{x})}{\lambda_\theta(x; \mathbf{x}) + \rho(x)}$$

where  $\nabla$  means the gradient, which is the score function of

$$\sum_{x \in \mathbf{x}} \log \left[ \frac{\lambda_\theta(x; \mathbf{x} \setminus \{x\})}{\lambda_\theta(x; \mathbf{x} \setminus \{x\}) + \rho(x)} \right] + \sum_{x \in d} \log \left[ \frac{\rho(x)}{\lambda_\theta(x; \mathbf{x} \setminus \{x\}) + \rho(x)} \right]$$

using that the dummy points do not lie in  $\mathbf{x}$ . Therefore the score function coincides with the loglikelihood for independent Bernoulli trials  $Y(x)$  taking value one for data points, zero for dummy points. The logit is

$$\log \left[ \frac{\lambda_\theta(x; \mathbf{x} \setminus \{x\})}{\rho(x)} \right].$$

Specializing to model (3), we denote by  $\theta = (\theta_1, \dots, \theta_m)$ ,  $\theta_j = \log(\gamma_j) \ell(C_{r_j}^{t_j})$ , the parameters in model (3), where  $\ell(C_{r_j}^{t_j})$  is the volume of the cylinder with radius  $(r_j, t_j)$  and  $(\gamma_1, \dots, \gamma_m)$  are the interaction parameters. We rescale the parameters in order to increase numerical stability and have the sufficient statistics bounded between 0 and 1. Recall the Papangelou conditional intensity for the spatio-temporal multi-scale area-interaction process is given by (6)

$$\lambda_\theta((y, s); \mathbf{x}) = \lambda(y, s) \prod_{j=1}^m \exp \left[ -\theta_j \ell(C_{r_j}^{t_j}(y, s) \setminus \bigcup_{(x,t) \in \mathbf{x}} C_{r_j}^{t_j}(x, t)) / \ell(C_{r_j}^{t_j}) \right].$$

Following Baddeley and Turner (2000) we denote by  $S_j((y, s), \mathbf{x}) = \ell(C_{r_j}^{t_j}(y, s) \setminus \bigcup_{(x,t) \in \mathbf{x}} C_{r_j}^{t_j}(x, t)) / \ell(C_{r_j}^{t_j})$  the sufficient statistics, hence  $\lambda_\theta((y, s); \mathbf{x}) = \lambda(y, s) \prod_{j=1}^m \exp[-\theta_j S_j((y, s), \mathbf{x})]$ . For each parameter



$\theta_j, j = 1, \dots, m$ , the partial derivative of the conditional intensity  $\lambda_\theta$  with respect to  $\theta_j$  reads

$$\frac{\partial}{\partial \theta_j} \lambda_\theta((y, s); \mathbf{x}) = -S_j((y, s), \mathbf{x}) \lambda_\theta((y, s); \mathbf{x}). \tag{13}$$

Plugging in (13), we get that the logit takes the form

$$\log \left[ \frac{\lambda(x, t)}{\rho(x, t)} \right] = \sum_{j=1}^m \theta_j S_j((x, t), \mathbf{x} \setminus (x, t)).$$

This is a linear model in  $\theta$  with offset  $\log [\lambda(x, t)/\rho(x, t)]$ . The term  $\lambda(x, t)$  may also depend on a parameter, say  $\beta$ . We shall consider  $\lambda(x, t) \equiv \beta$  in the simulations and  $\lambda(x, t) = \beta \lambda(x) Z(t)$ . In that case  $\beta$  becomes the intercept.

In summary we have the following algorithm (for  $\lambda(x, t) = \lambda_\beta(x, t) = \beta \mu(x, t)$  with  $\mu(x, t)$  free of any parameters).

- Generate a set of dummy points according to a Poisson or binomial process with intensity function  $\rho(\cdot)$  and merge them with all the data points in  $\mathbf{x}$  to construct the set of ‘quadrature’ points  $(u_j, v_j) \in W_S \times W_T$ ;
- obtain the response variables  $y_j$  (1 for data points, 0 for dummy points);
- compute the values  $S((u_j, v_j); \mathbf{x} \setminus \{(u_j, v_j)\})$  of the vector of sufficient statistics at each quadrature point;
- fit a logistic regression model with explanatory variables  $S((u_j, v_j); \mathbf{x} \setminus \{(u_j, v_j)\})$ , intercept and offset  $\log [\mu(x, t)/\rho(x, t)]$  to obtain estimates  $\tilde{\theta}$  for the  $S$ -vector and intercept  $\tilde{\theta}_0$ ;
- return the parameter estimator  $\hat{\theta} = -\tilde{\theta}$  and  $\hat{\beta} = \exp[\tilde{\theta}_0]$ .

Note that when  $\mu(x, t)/\rho(x, t)$  is constant, the offset parameter may be omitted, but then

$$\hat{\beta} = \frac{\rho(x, t)}{\mu(x, t)} \exp [\tilde{\theta}_0].$$

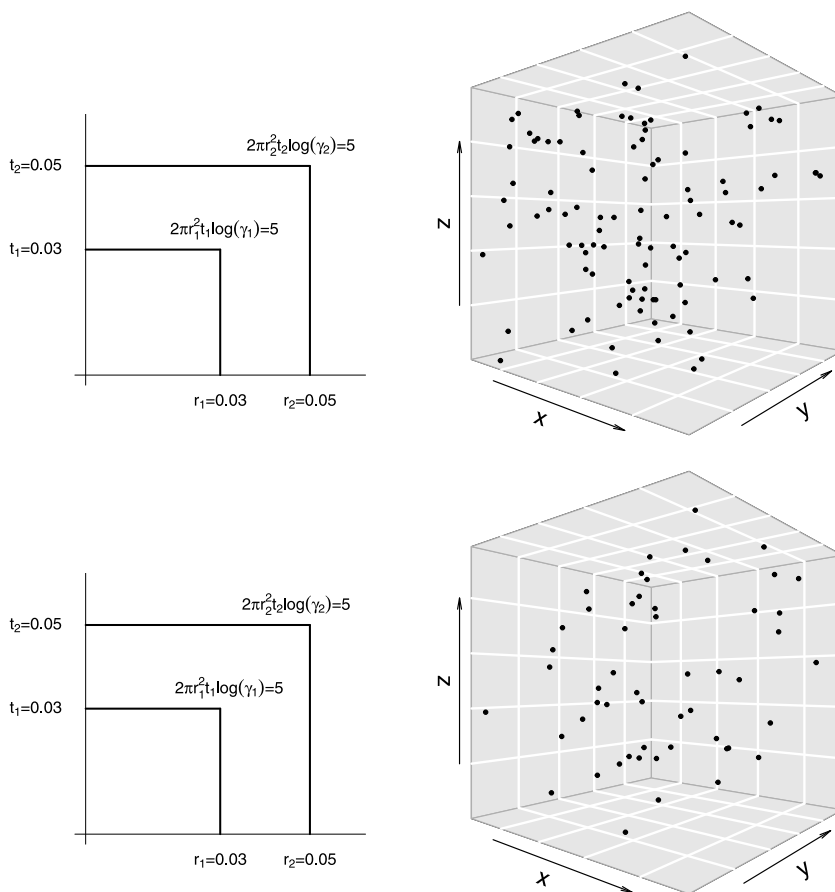
For the values of the *irregular* parameters  $r_j$  and  $t_j, j = 1, \dots, m$ , [Baddeley and Turner \(2000\)](#) suggest fitting the model for a range of values of these parameters and choose the values which maximize the log-likelihood. Additionally, we recommend to first compute some summary statistics, such as the pair correlation or auto-correlation function, to narrow down the search.

### 5.2. Simulation and parameter estimation of a spatio-temporal area interaction process

We generate simulations of two multi-scale spatio-temporal area interaction processes as defined in (3), one which exhibits small scale inhibition and large scale clustering (*sim 1*) and a second one which exhibits small scale clustering and large scale inhibition (*sim 2*).

We consider the spatio-temporal domain  $W = W_S \times W_T = ([0, 1] \times [0, 1]) \times [0, 1]$  and in both cases take constant  $\lambda \equiv 50$ . The parameter values used for the simulations are  $\theta_1 = 2\pi r_1^2 t_1 \log(\gamma_1) = -5, \theta_2 = 2\pi r_2^2 t_2 \log(\gamma_2) = 5$  for *sim 1* and  $\theta_1 = 2\pi r_1^2 t_1 \log(\gamma_1) = 5, \theta_2 = 2\pi r_2^2 t_2 \log(\gamma_2) = -5$  for *sim 2*. For the irregular parameters we choose the same spatio-temporal scales  $r_1 = 0.03, r_2 = 0.05, t_1 = 0.03$  and  $t_2 = 0.05$  for both simulations. We generate 10,000 simulations of the specified processes and estimate  $\theta = (\theta_1, \theta_2)$  using a Poisson dummy point process with  $\rho = 4n(X \cap W)/\ell(W)$  depending on the number of points in each simulation. We use the Metropolis–Hastings algorithm described in Section 4 with a burn in of 20,000, subsampled thereafter every 5,000 steps, implemented in the MPPLIB C++ library ([Steenbeek et al., 2016](#)). We use discretization size 127 in both space and time. The computational time was approximately 18 h for each model. We then compute the sufficient statistics for each data and dummy point using the MPPLIB C++ library and apply logistic regression to obtain the estimates for the parameters. For the implementation of the logistic regression method we use the statistical software R ([R Core Team, 2015](#)).

[Fig. 3](#) (top row) shows the multi-scale spatio-temporal behavior and one simulation from *sim 1*. On the left-hand side of the top row we see the interaction parameters and their corresponding spatial



**Fig. 3.** (Left) Model parameters. (Right) A realization of the specified model. (Top row) *Sim 1*. (Bottom row) *Sim 2*.

and temporal ranges. The expected number of points for this model is 87.29 points. On the right we show a realization of such a model with observed number of points equal to 97. The bottom row in Fig. 3 shows the multi-scale spatio-temporal behavior for *sim 2*. The expected number of points in this model is 45.96. On the right-hand side we show a realization of the model with 53 spatio-temporal points.

Table 1 shows the true parameter values, the estimated parameter values using the logistic regression approach described in Section 5.1, together with the quantiles of the parameter estimates under simulation of the true model. For the first part of the simulation study (*sim 1*) we obtain estimated values of  $-5.07$  ( $[-7.73, -2.68]$ ) for  $\theta_1$  and  $5.01$  ( $[2.95, 7.21]$ ) for  $\theta_2$ . For *sim 2*, we obtain an estimated value of  $5.06$  ( $[1.40, 9.02]$ ) for  $\theta_1$  and  $-5.29$  ( $[-9.06, -1.85]$ ) for  $\theta_2$ . Overall, the values for the estimated parameters are close to the real values for both models. Note that for *sim 2* the intervals are wider.

### 6. Data. Varicella in Valencia

The *Varicella-zoster virus* (VZV) is a highly contagious virus, spread worldwide, which causes two clinical syndromes: *varicella*, also known as chickenpox, and *herpes zoster*, otherwise known as shingles. In this paper we will focus on the spatial, temporal and spatio-temporal behavior of varicella.

**Table 1**Parameter estimates for *sim 1* and *sim 2*.

	<i>sim 1</i>				<i>sim 2</i>		
	Estimate	2.5%	97.5%		Estimate	2.5%	97.5%
$\log \lambda = 3.91$	3.81	2.09	5.32	$\log \lambda = 3.91$	3.72	0.74	5.92
$\theta_1 = -5$	-5.07	-7.73	-2.68	$\theta_1 = 5$	5.06	1.40	9.02
$\theta_2 = 5$	5.01	2.95	7.21	$\theta_2 = -5$	-5.29	-9.06	-1.85

Varicella is transmitted from person to person by direct contact with the rash or inhalation of aerosolized droplets from respiratory tract secretions of patients with varicella. In temperate countries more than 90% of the infections occur before adolescence and less than 5% of adults remain susceptible. Varicella is mostly a mild disorder in childhood, but tends to be more severe in adults. The first symptoms of varicella generally appear after a 10–21 days incubation period. It is characterized by an itchy, vesicular rash, fever and malaise. Varicella is generally self-limited and vesicles gradually develop crusts. It usually takes about 7–10 days for all the vesicles to dry out and for the crusts to disappear. This gives us a time period, from infection to completely dried vesicles, between 17 and 31 days.

Reported infection after household exposure ranges from 61% to 100% (Gershon et al., 2008; WHO, 2008) which indicates small range interaction. The disease may be fatal, especially in neonates and immuno-compromised individuals. The epidemiology of the disease is different in temperate and tropical climates. The reasons behind this behavior may be related to climate, population density and risk of exposure (Health Department, 2014; WHO, 2014).

In this paper we analyze varicella cases registered in Valencia, Spain, during 2013. Valencia is the third largest city in Spain with a population of around 800,000 inhabitants in the administrative center (19 districts) and an area of approximately 134 km<sup>2</sup> (Statistics Office, 2013). The study area is represented by districts 1 to 16. The remaining districts are very sparsely populated and are located far from the urban core. During the year 2013, 921 cases of varicella were registered in the study area in the course of 52 weeks (Health Department, 2014).

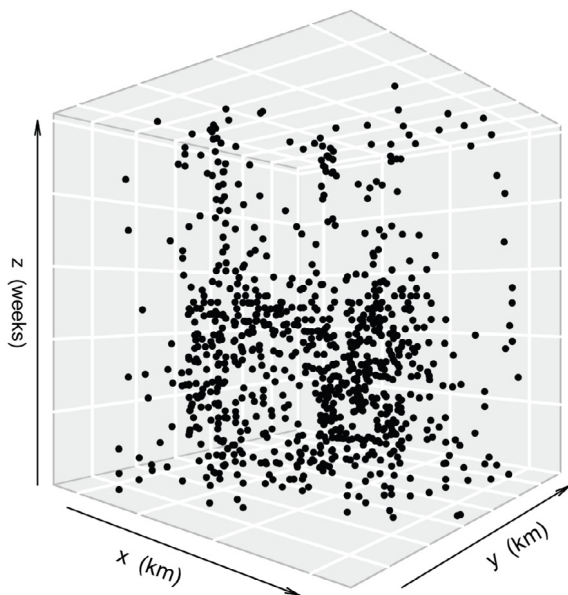
The spatial coordinates of the varicella cases are expressed in latitude and longitude. First we transform them from longitude/latitude to UTM scale expressed in meters (Snyder, 1987). Fig. 5 (left) shows the borders of Valencia together with a spatial projection of the spatio-temporal point pattern. For the analysis we rescale the spatial coordinates to kilometers such that the spatial study area reduces to a square,  $[0, 9] \times [0, 9]$ . This choice is much more convenient to work with and will not matter in practice since we set the first order term to zero outside the Valencia border.

The temporal component of the process takes integer values from 0 to 51. For computational purposes to be explained later, we take the interval  $[0, 52]$  as the time window. Therefore, we set the spatio-temporal study area to  $W_S \times W_T = ([0, 9] \times [0, 9]) \times [0, 52]$  (km<sup>2</sup> × weeks). The spatio-temporal pattern of all varicella cases thus obtained is shown in Fig. 4. The x- and y-axis represent the spatial coordinates in kilometers and the z-axis represents the time component in weeks.

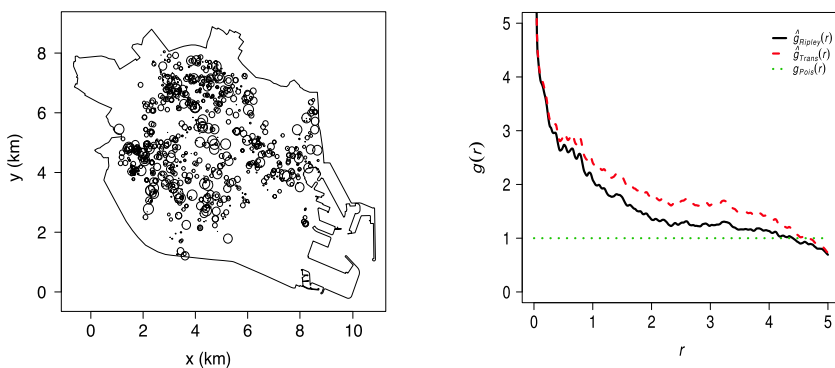
The main focus of our varicella data analysis is to quantify the interactions across a range of spatio-temporal scales. We do so by using the spatio-temporal multi-scale area-interaction model introduced in Section 3.

First we need to get some idea about a plausible upper bound to the values of the irregular parameters  $(r_j, t_j)$ ,  $j = 1, \dots, m$ , in model (3). To this end, we use summary statistics for the spatial and temporal projections of the space-time point pattern

The left panel in Fig. 5 shows the projection of all points onto the spatial region. The sizes of the circles are proportional to time, the bigger the circle, the more recent the event. Due to the projection, duplicate locations are observed, so we jitter the coordinates uniformly on the spatial region around the duplicated points using a maximum jittering distance of 20 m. To get a rough indication of the spatial interaction range, we pretend that the pattern is stationary and isotropic, and estimate the pair correlation function. The result is shown in the right panel of Fig. 5. Recall that for a Poisson process the pair correlation function is equal to 1. Values of the pair correlation function lower than 1 indicate inhibition and values larger than 1 suggest clustering. Fig. 5 suggests that the pair correlation



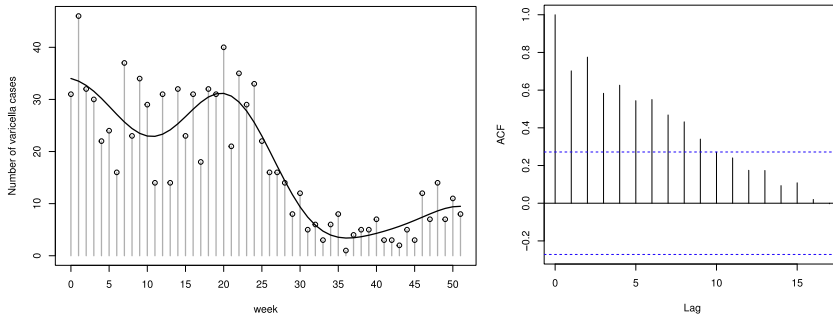
**Fig. 4.** Spatio-temporal pattern of weekly varicella cases in Valencia during 2013, where the spatio-temporal study area is  $W_S \times W_T = ([0, 9] \times [0, 9]) \times [0, 52]$  ( $\text{km}^2 \times \text{weeks}$ ).



**Fig. 5.** (Left) Spatial projection of the spatio-temporal point pattern for the varicella data. After projection, locations were jittered using a maximum jitter distance of 20 m. (Right) Estimated pair correlation function for the jittered spatial point pattern shown in the left panel.

function flattens off at 2 km, and has value 1 beyond 4 km, which indicates that the maximum spatial range should be between 1 and 2 km. On a cautionary note, we need to keep in mind that the estimator only takes into account the spatial pattern of points and assumes isotropy.

The left panel in Fig. 6 shows the temporal evolution of varicella over the 52 weeks, where the small circles  $\circ$  represents the number of registered cases. The right panel displays the estimated auto-correlation function which measures the correlation between the values of the series at different times as a function of the time lag between them. Fig. 6 suggests possible correlation for time lags as big as 15 weeks. This gives us an estimate for the maximum value for the  $t_i$  of about 7.5 weeks. Note that caveats similar to the spatial case apply.



**Fig. 6.** (Left) Weekly reports of varicella cases ( $\circ$ ) and fitted regression curve ( $-$ ). (Right) Estimated auto-correlation function for the data shown in the left panel.

Now that we have estimated the maximum spatial and temporal range for the model, the following step in our analysis is to consider covariate information. The most important factor in the transmission of any kind of disease, and especially a highly contagious one such as varicella, is the population. In areas with very low population we will probably not register as many varicella cases as in highly populated areas. Thus, the pattern of varicella cases can drastically change from one area to another, depending on the spatial distribution of the population, and from one week to the next one.

After a visual inspection of the spatio-temporal point pattern, we consider separability and, as suggested in Diggle et al. (2005), we express the spatio-temporal inhomogeneity term in Eq. (3) as a product  $\lambda(x, t) = \beta\lambda(x)Z(t)$ ,  $x \in [0, 9]^2$ ,  $t \in \{0, \dots, 51\}$ , between a non-parametric estimate of the population density  $\lambda(x)$  and a re-scaled parametric estimate of the temporal component  $Z(t)$ .

First consider the spatial component  $\lambda(x)$ . The population data available to us consist of the number of people living in each census section of the city of Valencia, a total number of 559 sections (districts 1 to 16). We randomly generate within each section  $p$  points, where  $p$  is equal to the number of people living in that particular section. This way, we obtain a sample of the population for the city of Valencia. We estimate its intensity by a kernel estimator, keeping in mind that the bandwidth has to be chosen carefully, to get  $\lambda(x)$ ,  $x \in W_S$ .

Following Diggle et al. (2005) we fit a harmonic regression to the pattern of the weekly varicella counts

$$Z(t) = c_0 + \sum_{j=1}^3 (c_j \cos(2\pi jt/52) + d_j \sin(2\pi jt/52)) + c(a + bt), \quad (14)$$

where  $Z(t)$  denotes the number of varicella cases at time  $t$ ,  $t = 0, \dots, 51$ , and  $c_0, a, b, c, c_j, d_j$ ,  $j = 1, 2, 3$ , are the parameters of the model.

The left panel in Fig. 6 shows the fitted regression curve. We observe a period at the beginning of the year, from Winter until Spring, with large numbers of varicella cases, and a second period starting around week 26, in which the number of cases decreases. These periods correspond roughly with the school term and the summer break. Also, in 2013, in Spain, there were several holidays besides the summer and winter holidays. On March 19, San Jose is celebrated and the period from the 24th to the 31st of March corresponds to the Easter holidays. As a consequence we can observe in Fig. 6 a decrease during the 11th and 12th week. Towards the end of the year, the number of cases picks up again as the Michaelmas term begins.

Finally, we re-scale the parametric estimate of the temporal component  $Z(t)$  by 100, in order to avoid obtaining extreme values for the spatio-temporal inhomogeneity term  $\lambda(x, t)$ .

Since realizations of (3) do not contain points with equal time stamps, we jitter in time as well as space. More precisely, the week index is replaced by a time stamp that is uniformly distributed in the indicated week so that the temporal component falls in  $W_T = [0, 52]$ .

To estimate the parameters we use the logistic regression approach described in Section 5.1, following the steps of the given algorithm. For the dummy process we simulate an inhomogeneous

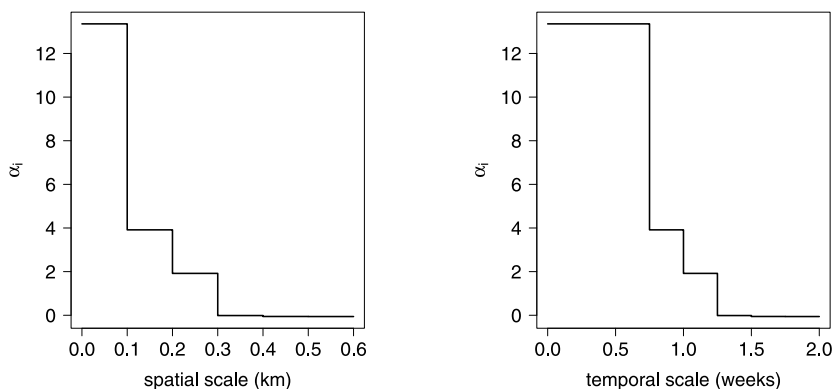


Fig. 7. Step function of the  $\alpha_i$ ,  $i = 1, \dots, 6$  parameters for different spatial (left) and temporal scales (right).

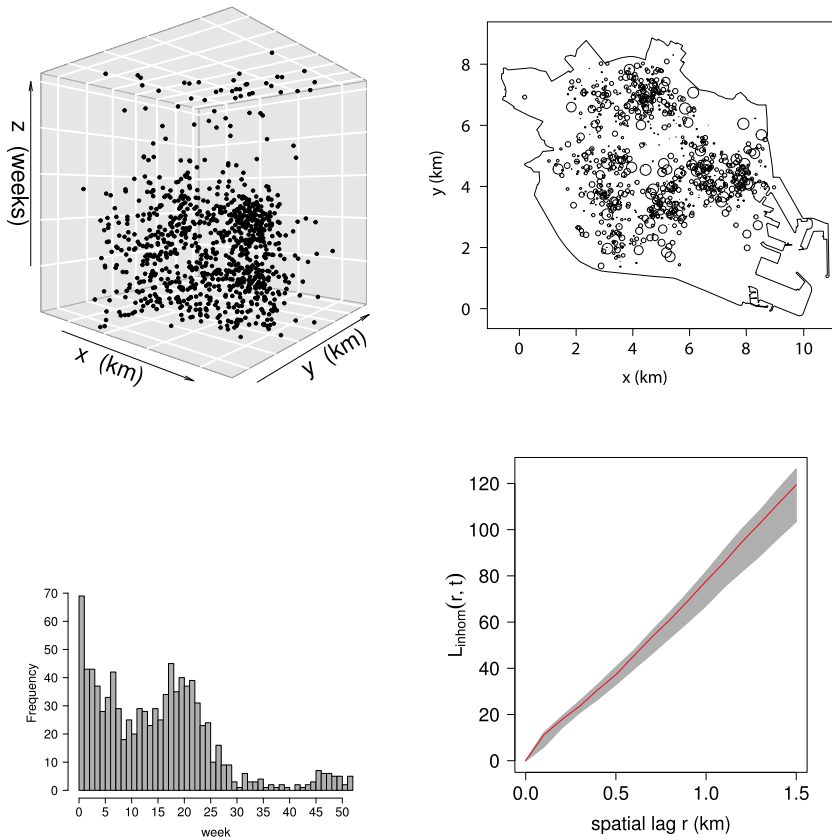
Poisson process with intensity  $\rho(x, t)$  proportional to  $\lambda(x, t) = \lambda(x)Z(t)/100$  with a proportionality constant of four. We obtain a dummy point process with 3888 points. We merge the data points with the dummy points to construct the ‘quadrature’ points. We obtain the response variable and compute the sufficient statistics at each quadrature point using the MPPLIB C++ library of Steenbeek et al. (2016). We then fit a logistic regression model and return the parameter estimators. The computational time was approximately 30 min. The analysis and visual representations have been carried out using the statistical software R (R Core Team, 2015) together with the spatstat (Baddeley et al., 2016) (version v1.55-0), plot3D (Soetaert, 2016) (version 1.1.1) and rgdal (Bivand et al., 2016) (version 1.2-16) packages. For fitting the logistic regression we have used the glm function in R.

Recall that we found indications for the maximum spatial range to be between 1 and 2 km and the maximum temporal range 7.5 weeks. Our objective is to find interaction between varicella cases, and describe how this behavior changes between different spatial and temporal scales. We consider discretization of 128 steps for both space and time. To avoid discretization effects one should not consider distances smaller than about 100 m (9 km/128 is approximately 70 m) and time intervals smaller than about 4 days (52 weeks/128 is approximately 0.4 weeks). Having decided that, we proceed to testing plausible combinations of parameters and spatial and temporal scales. We first start with spatial scales spanning up to 1.5 km and temporal scales up to 7.5 weeks. We obtain the values for the parameters  $\alpha_i$ ,  $i = 1, \dots, m$  in Eq. (8), and gradually zoom into the regions of interest by discarding the ones where  $\alpha_i$ ,  $i = 1, \dots, m$  start to oscillate around zero. We obtain evidence of strong clustering for small distances, decreasing for larger distances. This gives us a hint that our model should consider smaller spatial lags, up to about 1.2 km and smaller temporal lags up to about 4 weeks.

After a detailed and thorough search we found the best fit for our data, a model with number of parameters  $m = 6$ , spatial scales  $r_1 = 0.1, r_2 = 0.2, r_3 = 0.3, r_4 = 0.4, r_5 = 0.5, r_6 = 0.6$  and temporal scales  $t_1 = 0.75, t_2 = 1.0, t_3 = 1.25, t_4 = 1.5, t_5 = 1.75, t_6 = 2.0$ . Fig. 7 shows a step plot of the estimated  $\alpha_i$ ,  $i = 1, \dots, 6$  parameters,  $\alpha_1 = 13.36, \alpha_2 = 3.92, \alpha_3 = 1.92, \alpha_4 = -0.013, \alpha_5 = -0.056$  and  $\alpha_6 = -0.058$ . We can clearly see strong clustering at small distances, decreasing to about zero at  $2r_3 = 600$  m and  $2t_3 = 2.5$  weeks.

As stated before, the time period from infection to completely dried vesicles is between approximately 17 and 31 days. Thus the incubation time corresponds with the findings in the fitted model where we observe interaction among cases up to about 3 weeks.

We use additional information on spatial location of daycares, kindergartens and primary schools in Valencia to calculate the average distance from a varicella case to the nearest school and found a distance of approximately 611 m. This is in concordance with our findings, we observe strong clustering for distances up to 600 m, and no interaction beyond that. Fig. 7 describes the dynamic



**Fig. 8.** (Top row, left) One realization from the model fitted to the varicella data. (Top row, right) Spatial projection of the spatio-temporal point pattern for the simulation. (Bottom row, left) Histogram of the temporal component of the simulated point pattern. (Bottom row, right) Estimated  $L_{\text{inhom}}$ -function for the data (red line) together with the min–max envelopes (gray) based on 99 simulations from the fitted model.

behavior of varicella and, as expected, shows the main feature of the diseases, that is, being a highly contagious disease.

To validate our model, we ran the Metropolis–Hastings algorithm described in Section 4 with the fitted parameters for a burn-in period of 200,000 steps, which seemed enough for convergence based on diagnostic plots. We then subsampled every 20,000 steps to obtain 99 realizations of the fitted model; this subsampling is enough to ensure that some 90 percent of the points in consecutive realizations have been replaced. The computational time is approximately 10 h. The expected number of points was 931.23. Fig. 8 (top row, left) shows 3 dimensional plot of one simulation. Fig. 8 (top row, right) shows the spatial projection corresponding to the simulated pattern. Comparing this figure with Fig. 5 we identify the shape of the pattern is similar to the varicella point pattern. Also, comparing Fig. 8 (bottom row), a histogram of the temporal component of the simulation, we note the similarities between this figure and Fig. 6. This represents a visual confirmation that the proposed model represents a good fit.

A quantitative approach for the validation of the model is to use the space–time inhomogeneous  $L$ -function (Gabriel and Diggle, 2009; Cressie and Wikle, 2011)

$$L_{\text{inhom}}(r, t) = (K_{\text{inhom}}(r, t)/(2\pi))^{1/3}, \quad (15)$$

where  $r$  and  $t$  are the spatial and temporal lags for which we calculate the values of the function and  $K_{\text{inhom}}(r, t)$  is the space–time inhomogeneous  $K$ -function  $K_{ST}^*$  proposed by Gabriel and Diggle (2009).

More specifically, we apply the Monte Carlo approach (Baddeley et al., 2016). Thus, we calculate the estimator (Gabriel, 2014) for the inhomogeneous  $K$ -function and plug it into (15). This way, we can obtain estimates  $\hat{L}_{\text{inhom}}(r, t)$  for the varicella data and for 99 simulations from the fitted model.

Fig. 8 (bottom row, right) shows the minimum–maximum envelopes (gray shading) and the estimated  $L_{\text{inhom}}(r, 3r)$  function for the data (red line), calculated for  $r \in [0, 1.5]$ . We can see that the fit is good. The figure represents a diagonal view of the  $L$  function. Examining the results for the complete  $L_{\text{inhom}}(r, t)$ , we get few inconsistencies for spatial range  $r = 0.1$  and large  $t$ , which correspond to spatial lags in the model of 0.05, that are below the discretization threshold. We can conclude that the model fits very well, except for a bit more clustering at very small distances.

## 7. Discussion and final remarks

In this paper we developed an extension of the area-interaction model that is able to incorporate different types of interaction at different spatio-temporal scales and proposed methods to simulate this process. We discussed inference and demonstrated the logistic regression approach on simulated data. Additionally, we analyzed a spatio-temporal point pattern of varicella in the city of Valencia, Spain. Section 6 represents a practical exercise where we show a step-by-step analysis of a spatio-temporal point pattern with an application of the multi-scale area-interaction model. The results describe the dynamic behavior of varicella, a widely spread disease, that can be interesting from an epidemiological point of view. We carry out the analysis using a ‘manual’ selection for the maximal spatial and temporal scales in combination with discretization. A complete profile likelihood would have been preferred but not computationally feasible in this case.

For future work, it would be interesting to apply our model to other diseases that may exhibit interaction at several scales in space and time. It would also be very interesting to apply this model to data that are not necessarily related to epidemiology. Earthquake patterns, for instance, tend to show aggregation but also inhibition at different scales. Indeed, we believe that the proposed model may find applications in a wide range of research fields, such as forestry, geology and sociology.

As stated in Section 6, varicella is a highly infectious disease. We are certain that, in addition to the effect of the population, there are other covariates that may influence the spatio-temporal behavior of the disease. Therefore, an important goal for future work is to consider adding covariates that can improve the model. For example, WHO (2014) suggests that there are some climatic factors that can influence the epidemiology of varicella. Thus, covariates such as the monthly average temperatures, weekly average levels of rainfall, average hours of sunshine, or other climate related covariates, may provide useful information to the analysis of varicella. Also, additional information on the composition of households, income per capita or other socio-economical covariates might improve the model. Another important covariate that could be taken into account in future work is related to the locations of kindergartens, schools and high-schools. The distance from a case to the nearest school may provide important information for the analysis of varicella.

## Acknowledgments

The authors thank Francisco González of the Surveillance Service and Epidemiological Control, General Division of Epidemiology and Health Surveillance – Department of Public Health, Generalitat Valenciana for providing the varicella data. They are grateful to Professor R. Turner for helpful discussions. The first author gratefully acknowledges financial support from the Ministry of Education, Culture, and Sports (Grants FPU12/04531 and EST15/00174) for a research visit to The Netherlands. She thanks the Stochastics group at CWI and the SOR-chair at Twente University for their hospitality. The work was done while the first author was working at the University of Valencia. The authors also want to thank the two anonymous referees for their constructive comments that helped to improve the paper.



## References

- Ambler, G.K., Silverman, B.W., 2010. Perfect simulation using dominated coupling from the past with application to area-interaction point processes and wavelet thresholding. In: Bingham, N.H., Goldie, C.M. (Eds.), *Probability and Mathematical Genetics*. Cambridge University Press, Cambridge.
- Baddeley, A., Coeurjolly, J.F., Rubak, E., Waagepetersen, R., 2014. Logistic regression for spatial Gibbs point processes. *Biometrika* 101 (2), 377–392.
- Baddeley, A., Rubak, E., Turner, R., 2016. *Spatial Point Patterns: Methodology and Applications with R*. CRC Press, Boca Raton.
- Baddeley, A., Turner, R., 2000. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Aust. N.Z. J. Stat.* 42 (3), 283–322.
- Baddeley, A.J., van Lieshout, M.N.M., 1995. Area-interaction point processes. *Ann. Inst. Statist. Math.* 47 (4), 601–619.
- Barker, A.A., 1965. Monte Carlo calculations of radial distribution functions for a proton-electron plasma. *Aust. J. Phys.* 18 (2), 119–134.
- Besag, J., 1977. Some methods of statistical analysis of spatial data (with discussion). *Bull. Int. Stat. Inst.* 47 (2), 77–91 138–147.
- Bivand, R., Keitt, T., Rowlingson, B., 2016. Rgdal: Bindings for the Geospatial Data Abstraction Library. URL <https://CRAN.R-project.org/package=rgdal>.
- Chiu, S.N., Stoyan, D., Kendall, W.S., Mecke, J., 2013. *Stochastic Geometry and its Applications*, third ed. Wiley.
- Cressie, N., Wikle, C.K., 2011. *Statistics for Spatio-Temporal Data*. Wiley, New Jersey.
- Daley, D.J., Vere-Jones, D., 2003. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Springer, New York.
- Daley, D.J., Vere-Jones, D., 2008. *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Springer, New York.
- Dereudre, D., Lavancier, F., Helisová, K.S., 2014. Estimation of the intensity parameter of the germ-grain quermass-interaction model when the number of germs is not observed. *Scand. J. Stat.* 41 (3), 809–829.
- Diggle, P., Rowlingson, B., Su, T., 2005. Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics* 16 (5), 423–434.
- Gabriel, E., 2014. Estimating second-order characteristics of inhomogeneous spatio-temporal point processes: Influence of edge correction methods and intensity estimates. *Methodol. Comput. Appl. Probab.* 16 (2), 411–431.
- Gabriel, E., Diggle, P.J., 2009. Second-order analysis of inhomogeneous spatio-temporal point process data. *Stat. Neerl.* 63 (1), 43–51.
- Georgii, H.O., 1976. Canonical and grand canonical Gibbs states for continuum systems. *Comm. Math. Phys.* 48 (1), 31–51.
- Gershon, A., Takahashi, M., Seward, J., 2008. Varicella vaccine. In: *Vaccines*, fifth ed. WB Saunders, Philadelphia, pp. 916–958.
- Gregori, P., van Lieshout, M.N.M., Mateu, J., 2003. Modelización de procesos area-interacción generalizados. In: 27 Congreso Nacional de Estadística e Investigación Operativa. Edicions de la Universitat de Lleida, Lleida, (in Spanish).
- Häggström, O., van Lieshout, M.N.M., Möller, J., 1999. Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli* 5 (4), 641–658.
- Health Department. Varicella Report. Epidemiological Surveillance 2013. Public Health. Epidemiology and Health Surveillance. Epidemiological Surveillance and Control, May 2014.
- Iftimi, A., van Lieshout, M.N.M., Montes, F., 2017. A multi-scale area-interaction model for spatio-temporal point patterns. ArXiv e-prints arXiv:1701.02887.
- Jensen, J.L., Möller, J., 1991. Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.* 1 (3), 445–461.
- Kelly, F.P., Ripley, B.D., 1976. A note on Strauss's model for clustering. *Biometrika* 63 (2), 357–360.
- Kendall, W.S., 2000. Perfect simulation for the area-interaction point process. In: Accardi, L., Heyde, C. (Eds.), *Probability Towards 2000*. Springer, New York.
- Kendall, W.S., van Lieshout, M.N.M., Baddeley, A.J., 1999. Quermass-interaction processes: Conditions for stability. *Adv. Appl. Probab.* 31 (2), 315–342.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21 (6), 1087–1092.
- Möller, J., Helisová, K., 2010. Likelihood inference for unions of interacting discs. *Scand. J. Stat.* 37 (3), 365–381.
- Nguyen, X.X., Zessin, H., 1979. Ergodic theorems for spatial processes. *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* 48, 133–158.
- Picard, N., Bar-Hen, A., Mortier, F., Chadoeuf, J., 2009. The multi-scale marked area-interaction point process: A model for the spatial pattern of trees. *Scand. J. Stat.* 36 (1), 23–41.
- R Core Team, 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ripley, B.D., 1988. *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge.
- Ripley, B.D., 1990. Gibbsian interaction models. In: Griffith, D.A. (Ed.), *Spatial Statistics: Past, Present, and Future*. Institute of Mathematical Geography, Ann Arbor, pp. 1–19.
- Ripley, B.D., Kelly, F.P., 1977. Markov point processes. *J. Lond. Math. Soc.* 15 (1), 188–192.
- Snyder, J.P., 1987. *Map Projections – A Working Manual*. United States Government Printing Office, Washington, D.C.
- Soetaert, K., 2016. Plot3D: Plotting Multi-Dimensional Data. URL <https://CRAN.R-project.org/package=plot3D>.
- Stallybrass, C.O., 1931. *The Principles of Epidemiology and the Process of Infection*. Routledge, London.
- Statistics Office, 2013. Municipal register of inhabitants at 1st of January. Valencia Local Council.

- Steenbeek, A.G., van Lieshout, M.N.M., Stoica, R.S., with contributions from Gregori, P., Berthelsen, K.K., Iftimi, A., 2016. MPPLIB, a C++ library for marked point processes. CWI.
- Strauss, D.J., 1975. A model for clustering. *Biometrika* 62 (2), 467–475.
- Takacs, R., 1986. Estimator for the pairpotential of a Gibbsian point process. *Statistics* 17 (3), 429–433.
- Takacs, R., Fiksel, T., 1986. Interaction pair-potentials for a system of ants' nests. *Biom. J.* 28, 1007–1013.
- van Lieshout, M.N.M., 2000. *Markov Point Processes and their Applications*. Imperial College Press, London.
- WHO, 2008. *The Immunological Basis for Immunization Series. Module 10: Varicella-zoster Virus*. WHO Press.
- WHO, 2014. Weekly epidemiological record. Varicella and herpes zoster vaccines: WHO position paper. URL <http://www.who.int/wer>.
- Widom, B., Rowlinson, J.S., 1970. New model for the study of liquid-vapor phase transitions. *J. Chem. Phys.* 52 (4), 1670–1684.