University of California
Los Angeles

Survey of some recent advances in spatial-temporal point processes

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Benjamin M. Greenspan

2013

The thesis of Benjamin M. Greenspan is approved.

Hongquan Xu

Ying Nian Wu

Frederic Paik Schoenberg, Committee Chair

University of California, Los Angeles

2013

To my mom and dad . . .

may they live many good days.

Table of Contents

Acknowledgments

Before all others I would like to praise G-d. Additionally I am grateful to Rick Schoenberg, whose kindness, patience, and understanding has made this thesis possible.

CHAPTER 1

Introduction

Many interesting developments have been made recently in the field of spatial-temporal point processes, and their estimation and applications. Spatial-temporal point processes have been useful for applications in many fields, including the study of earthquakes, wildfires, and other natural disasters, as well as forests and other ecological data, neurological data, invasive species, epidemics, spatial debris, and many others. Schoenberg et. al (2002) write a brief encyclopedic description of the subject. Following that article a scholar is encouraged to examine Vere-Jones (2009). Additionally Daley and Vere-Jones together answer many good questions in two volumes entitled An Introduction to the Theory of Point Processes. Other recent works in these areas create extensions of the general model, applications to earthquakes, higher-order statistics, or analyze residuals.

The thesis is organized in several respects. All above mentioned categories are chapters. Various sorts of spatial-temporal point processes, models, and solution are reviewed under the general model heading. The section on earthquakes traces longitudinal developments in one area of application. Expectations for behavior in spatial-temporal distributions are expanded upon in the chapter on higher-order statistics. Residuals in the spatial-temporal domain conclude the analysis. Within each chapter the principles from citations are summarized. In my estimation the chapter headings and summaries query meaningful contributions from each citation. Here the preference is to transmit consistent notation. The body in each chapter follows chronological order. Data sets uses are noted when applicable.

CHAPTER 2

Theoretical results and descriptions of space-time point processes

A spatial-temporal marked point process N is a random set of points in a $(1+n+d)$-dimensional space formed of one temporal coordinate, n spatial coordinates, and a d-dimensional covariate (mark) associated with each point. The conditional intensity $\lambda(t, x, m)$ is the expected rate that points with the mark m will occur around the location (t, x) of space-time, conditional on the history of the process prior to time t. If this rate does not depend on the history of the process at all, that is, on what points have occurred previously, then N is a (possibly inhomogeneous) Poisson process, and the numbers of points in disjoint spatial-temporal regions are independent Poisson random variables. The conditional intensity is typically estimated by smoothing the observations or fitting a parametric model, for instance by maximum likelihood.

2.1 Choi and Hall (1999)

Data of earthquakes in the Kanto region of Japan are explored and analyzed. The authors achieve this using nonparametric methods based on kernel smoothing. The first estimate is the nonparametric estimation of intensity obtained by

$$\hat{\lambda}(x, t) = \frac{1}{h_1 h_2} \sum_i K_1\left(\frac{X_i - x}{h_1}\right) K_1\left(\frac{T_i - t}{h_2}\right),$$

where K_1 is a univariate kernel and h_1, h_2 are bandwidths. Next the authors examine the strongest activity by location according to chronological order by examining

$$\hat{\mu}(y|t) = \frac{1}{h_2 h_3^2} \sum_{i:T_i \in \mathcal{T}(t)} K_1\left(\frac{T_i - t}{h_2}\right) K_2\left(\frac{Y_i - y}{h_3}\right),$$

where K_2 is a bivariate kernel, $h_3 I_2$ is a 2×2 bandwidth matrix for K_2, I_2 is the 2×2 identity matrix and y is a 2×1 vector. Afterwards a series of graphs show the relationship between location, time and either peak magnitude, the expected strength according to a nonparametric estimator or energy. Whereas peak magnitude and strength are estimated nonparametrically, energy is measured by $E = 10^{1.5(M-M_0)}$. Mixed signals are seen concerning the presence of clustering.

2.2 Schoenberg et al. (2002)

This is a review article describing spatial-temporal point processes, some of which is summarized below. A spatial-temporal point process N is defined in functional form by $\lambda(t, x, y, z)$, the infinitesimal expected rate of events at time t and location (x, y, z) given all the observations up to time t. When the point process N is a Poisson process, λ is deterministic. Put more simply, $\lambda(t, x, y, z)$ depends only on t, x, y, and z. If λ is a constant, the Poisson process is stationary. The stationary case is the most simple of all. Processes with spatial heterogeneity are occasionally modeled as stationary in time but not space.

Stationary spatial-temporal point processes are sometimes described by the second-order parameter measure $\rho(t', x', y', z')$,. This quantity measures the covariance between the number of points in spatial-temporal regions A and B, where region B is A shifted by (t', x', y', z'). For a self-exciting (equivalently, clustered) point process, the function ρ is positive for small values of t', x', y', and z'; N is self-correcting (equivalently, inhibitory) if instead the covariance is negative.

Self-exciting point process models are common in epidemiology and seismology for clusters of events in time and space. One such general model is the Hawkes model, where $\lambda(t, x, y, z)$ is written as

$$\mu(t, x, y, z) + \sum_i \nu(t - t_i, x - x_i, y - y_i, z - z_i)$$

with the sum over the history of points (t_i, x_i, y_i, z_i) with $t_i < t$. The respective functions μ and ν represent the deterministic background rate and clustering density. Often μ is modeled merely as a function of the spatial coordinates (x, y, z), and may be estimated nonparametrically. When observed marks m associated with each point are posited to affect the rate at which future points accumulate, this information is typically incorporated into the function ν. The inclusion of marks yields a conditional rate λ estimated as a background rate plus

$$\sum_i \nu(t - t_i, x - x_i, y - y_i, z - z_i, m - m_i).$$

Sometimes λ is modeled as a product of marginal conditional intensities

$$\lambda(t, x, y, z) = \lambda_1(t)\lambda_2(x, y, z)$$

or even

$$\lambda(t, x, y, z) = \lambda_1(t)\lambda_2(x)\lambda_2(y)\lambda_4(z)$$

These forms convey that the temporal behavior of the process is independent of the spatial behavior, and, in the stronger case, the behavior of each spatial coordinate is independent.

2.3 Schoenberg (2004a)

The author studies the assumptions required for consistently estimated and asymptotically normal parameters. Three relatively lenient assumptions must be satisfied for the Poisson maximum likelihood estimator to be consistent and asymptotically normal. Poisson maximum likelihood estimation refers to estimation of parameters in a Poisson distribution. Recall that the likelihood function of a spatial point process is the sum over all observed points of the log of the conditional intensity at the points minus the integral of the conditional intensity over the whole space. The conditions for consistency of the Poisson MLE are: the parameter space must be piecewise continuous; the variance of the sum over the log of the conditional intensity for all possible solutions must be on the order of $o(\phi(T)^2)$ for some function $\phi(\cdot)$, and both (1) the conditional intensity function using the PMLE and (2) the absolute difference of the logarithm must be bounded away from zero. Example 5.2 (their numbering) shows how one could easily verify these conditions in a spatial-temporal Poisson process.

2.4 Peng et al. (2005)

The authors assess the performance of the burning index used to predict wildfires. The burning index is a composite of meteorological and fuel variables. The data, which have missing values, are clustered by location, year and season. One calculates the background rate and the burning index readings between weather stations by kernel smoothers. Subsequently several constraints on the likelihood are added. Overall fit is measured by AIC as well as by thinned residuals. In the residual analysis, characteristics of the space-time process are clearly described.

2.5 Guan (2008b)

A non-parametric smoothing kernel that relates covariates is described. Although the bias in the local smoother estimate typically is negligible, the variance does not diminish. In other words, one does not claim that as the sample size gets bigger, it becomes less and less likely that such an estimate differs from the population parameter by at least some positive delta. As a result, the local smoother estimator is not consistent. The authors substitute a consistent covariate smoother in place of a local smoother in order to estimate the conditional intensity function of an inhomogeneous spatial-point process. The covariate smoother uses principal component analysis, a particular kind of multivariate data analysis. The technique is demonstrated on locations of *Ocotea whitei* trees on the tropical Barro Colorado island.

2.6 Vere-Jones (2009)

The author writes technical summaries of Poisson processes, first and second order moments, variations of conditional intensity and estimation procedures. A process is Poisson when all of the following are true: points are independent for disjoint regions; the distribution of points is determined by

$$\Pr[N(A) = k] = \left[\Lambda(A)^k / k!\right] e^{-\Lambda(A)};$$

and $E[N(A)] = \mathrm{Var}[N(A)] = \Lambda(A)$. The moments of a point process express the expectations of the counting variables $N(A)$. The first moment measure (equivalently, the expectation measure) is $M(A) = E[N(A)]$. The second moment measure, namely

$$M_2(dz_1 \times dz_2) = E[N(dz_1) \times N(dz_2)]$$

where z is a generic variable, can be interpreted as the probability of finding points of the process in neighborhoods of both z_1 and z_2. The author recognizes the widespread use of partial likelihood estimation and notes that the log-likelihood ratio given by

$$\log[L_1/L_0] = \sum_{i=1}^{N(T)} \log[\lambda^*_{\mathcal{F}}(t_i)_{\mathcal{F}}/\lambda] - \int_0^T [\lambda^*_{\mathcal{F}}(t) - \lambda] dt$$

differentiates between a conditional intensity $\lambda^*_{\mathcal{F}}(t)$ and a simple Poisson process. The author predicts growth for the number of studies in the field of spatial-temporal marked point processes due to increasing performance of computers and increasing availability of high-quality data. The author mentions the `SSLib` package developed for earthquake modeling in the statistical software program R.

2.7 Diggle et al. (2010a)

This study of birds' nests on the delta of the Ebro river in Spain estimates parameters by optimizing the partial likelihood. The authors reference the definition of the log-likelihood for spatial point processes, namely

$$L(\theta) = \sum_{i=1}^{n} \log \lambda(x_i, t_i|\mathcal{H}_{t_i}) - \int_0^T \int_A \lambda(x, t|\mathcal{H}_t) dx dt.$$

Sometimes this quantity is intractable, though the authors are optimistic a good estimate may be obtained by maximizing the partial log-likelihood

$$L_p(\theta) = \sum_{i=1}^{n} \log p_i = \sum_{i=1}^{n} \log \lambda(x_i, t_i|\mathcal{H}_{t_i}) - \sum_{i=1}^{n} \log \left\{ \int_{\mathcal{R}_i} \lambda(x, t_i|\mathcal{H}_{t_i}) dx \right\}.$$

The authors describe Monte Carlo Markov Chain simulation and a population growth model. The authors model bird nests in a wetlands near the Mediterranean Sea as the spatial-temporal point process

$$\lambda(x, t|\mathcal{H}_t) = \lambda_0(t) \exp\{\beta z(x)\} g(x, t|\mathcal{H}_t),$$

where $\lambda_0(t)$ is arbitrary, $z(x)$ denotes the elevation, and $g(x, t|\mathcal{H}_t)$ models dependence. Among the fitted models, the conditional intensity of the model with the greatest maximized likelihood value is

$$\lambda(x, t|\mathcal{H}_t) = \lambda_0(t) \exp\{\beta_1 z(x) + \beta_2 z(x)^2\} \times \left[1 + \theta \exp\left\{ -\frac{u^*(t) - d_0}{\phi} \right\} \right] I[u^*(t) > d_0],$$

where $u^*(t)$ is the distance to the nearest neighbor.

2.8 Diggle et al. (2010b)

The authors extend the spatial-temporal model to data at the individual level for cases and at the ecological level for the the population at risk. A model is fit for meningococcal cases using discrete and continuous predictors, where the population at risk are treated as a spatially inhomogenous Poisson process.

In preparatory remarks the authors report N observations in the data. The first M observations are cases, and M is much smaller than the size of the population N. Let s_i be the spatial locations of the observations. One writes the covariates as $X_i = \{X_{i1}, \cdots, X_{ip}\}$ with $X_{i1} = 1$. The covariates are observed for the first M cases. The spatially aggregated covariates $\tilde{\mu}_{jk} : j = 1, \cdots, p$ are substituted for the at-risk observations in each of the $k = 1, \cdots, K$ subregions of D. The authors define

$$\hat{\mu}_{jk}(\beta) = \sum_{i=1}^{M} \frac{X_j(s)}{f[X(s)'\beta]} I(s_i \in D_k),$$

and note that $\tilde{\mu}_{1k}$ is the number at risk in D_k. They remark that $\tilde{\mu}_{jk}$ and $\hat{\mu}_{jk}(\beta)$ are unbiased estimators for the true attributable number. The result is a set of unbiased estimating equations for β

$$U_j(\beta; W) = \sum_{k=1}^{K} w_k[\tilde{\mu}_{jk} - \hat{\mu}_{jk}(\beta)] = 0, \qquad j = 1, \cdots, p,$$

where $W = \{w_k : k = 1, \cdots, K\}$ is any set of predefined weights. The authors then instruct how to choose the optimal weights under Poisson or non-Poisson conditions. Non-Poisson behavior may be detected by defining

$$G_{kl}(\beta) = \sum_{i_1, i_2 = 1}^{M} \frac{I(s_{i_1} \in D_k, s_{i_2} \in D_l, i_1 \neq I_2)}{\exp[X(s_{i_1})'\beta] \exp[X(s_{i_2})'\beta]}$$

and the estimated pair-correlation function

$$\tilde{g}(u; \beta) = \frac{\sum_{k,l=1}^{K} \kappa[(d_{kl} - u)/h] G_{kl}(\beta)}{\sum_{k,l=1} \kappa[(d_{kl} - u)/h] \mu_{1k}\mu_{1l}},$$

where κ is a kernel function and h is the bandwidth. The pair-correlation function indicates if the process is Poisson, clustered or inhibited by $\tilde{g}(u;\beta) \approx 1$, $\tilde{g}(u;\beta) > 1$ or $\tilde{g}(u;\beta) < 1$, respectively. The authors demonstrate a formal test.

The authors focus on childhood meningococcal disease in the metropolitan county of Merseyside, U.K., arriving at the model

$$\lambda(s,t) = \lambda_0(s,t) \exp[X(s,t)'\beta + h(t)],$$

where $h(t)$ is a cubic spline term for the temporal trend at year t.

CHAPTER 3

Epidemic-type aftershock sequence model and clustering

The ETAS model is a central application of a spatial-temporal model to earthquakes. One can think about such a process as an example of a self-exciting (equivalently, clustered) process where points can trigger additional points and those points can trigger additional points. A self-exciting process is contrastable with a self-correcting (equivalently, inhibitory) process, in which the conditional intensity function is reduced in response to close points.

3.1 Ogata (1998)

The paper models the locations, times and magnitudes of earthquakes using a marked space-time point process. Since the earthquake events are clustered in time and space, the author forms several quantitative models that can tolerate non-homogeneity due to time, location, and magnitude. Ogata notes the predicted rate for the process is conditioned from the history of aftershocks and that it can be described by a Hawkes process. The candidate functions that satisfy the spatial clustering constraint equation are:

$$\nu(t,x,y;M) = \frac{K_0}{(t+c)^p} \exp\left\{-\frac{1}{2}\frac{x^2+y^2}{de^{\alpha(M-M_0)}}\right\},$$
$$\nu(t,x,y;M) = \frac{K_0}{(t+c)^p}\frac{e^{\alpha(M-M_0)}}{(x^2+y^2+d)^q},$$

and

$$\nu(t,x,y;M) = \frac{K_0}{(t+c)^p}\left(\frac{x^2+y^2}{e^{\alpha(M-M_0)}}+d\right)^{-q}.$$

The author forms a theory that the common standard form

$$\nu(t,x,y;M) = \kappa(M) \times \frac{(p-1)c^{p-1}}{(t+c)^p} \times \left[\frac{1}{\pi\sigma(M)} \cdot f\left\{\frac{r^2(\theta)}{\sigma(M)}\right\}\right]$$

represents the respective contributions of magnitude, time, and location. The author formulates

$$r^2(\theta) = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{\sigma_2}{\sigma_1}x^2 - 2\rho xy + \frac{\sigma_1}{\sigma_2}y^2\right)$$

with $\theta = \tan^{-1}(y/x)$ as a model for elliptical aftershock zones. Various applications and extensions are compared using AIC to measure the goodness-of-fit. A simulation is provided for corroboration. The discussion centers on earthquake measurements taken during the twentieth century near Honshu Island, Japan, which is the main island, as well as near the main island's Tohuku region in the Northeast.

3.2 Ogata and Zhuang (2006)

In this installment of modeling earthquake data near Japan, the authors evaluate accumulating knowledge with a view to reducing bias. Assuming that every aftershock can trigger successive incidents, the authors suggest predictions from the epidemic-type aftershock sequence model

$$\lambda(t) = \mu + \sum_{j:t_j<t} e^{\alpha\{M_j-M_c\}} \upsilon(t-t_j)$$

or the space-time ETAS model

$$\lambda(t,x,y) = \mu(x,y) + \sum_{j:t_j<t} \upsilon(t-t_j) \times g(x-x_j, y-y_j, M_j-M_c)$$

with the parameters also obtained by maximum likelihood estimation. In the latter, the space effects are

$$g(x - x_j, y - y_j; M_j - M_c) = \exp\left[-\frac{1}{2}\frac{(x - x_j, y -_y j)S_j(x - x_j, y - y_j)^t}{de^{\alpha(M_j - M_c)}}\right],$$

or

$$g(x - x_j, y - y_j; M_j - M_c) = \frac{e^{\alpha(M_j - M_c)}}{[(x - x_j, y -_y j)S_j(x - x_j, y - y_j)^t + d]^q},$$

or

$$g(x - x_j, y - y_j; M_j - M_c) = \left[\frac{(x - x_j, y -_y j)S_j(x - x_j, y - y_j)^t}{e^{\alpha(M_j - M_c)}} + d\right]^{-q},$$

where S_j is 2×2 dimensional given by the identity matrix or a bivariate covariance matrix of an ellipsoidal cluster for isotropic or anisotropic distributions, respectively. The presence of $e^{\alpha(M_i - M_c)}$ in each of the above formulas traces to the Utsu-Seki law, which describes the inverse power decay of aftershocks with regard to spatial distance. Furthermore the Modified Omori formula specified by

$$v(t) = K(t + c)^{-p}, \qquad (K, c, p; \text{parameters}),$$

with parameters obtained by maximum likelihood estimation commonly holds true for decades for aftershock rate decay notwithstanding limitations. After comparing the AIC of each model, the authors selected the third out of the three spatial effects models. The authors' desire, however, is to continue the development of the model. The authors recall the proposed common standard form

$$\nu(t, x, y; M) = \kappa(M) \times \frac{(p-1)c^{p-1}}{(t+c)^p} \times \left[\frac{1}{\pi\sigma(M)} \cdot g\left\{\frac{(x, y)S(x, y)^t}{\sigma(M)}\right\}\right]$$

and offer an improved model

$$g(x - x_j, y - y_j; M_j - M_c) = e^{(\alpha - \gamma)(M_j - M_c)}\left[\frac{(x - x_j, y -_y j)S_j(x - x_j, y - y_j)^t}{e^{\alpha(M_j - M_c)}} + d\right]^{-q}.$$

Using the data from four seismically active regions in Japan compiled by the Japan Meterological Agency, they confirm the improved model is indeed superior to its antecedent. The authors apply a declustering method – specifically, a thinning method – to generate a random sample of dependent aftershocks.

3.3 Marsan and Lengline (2008)

The authors introduce an entirely nonparametric way of estimating the triggering function in a branching point process model. They then apply it to California earthquakes and show how the triggering decays as a function of time and distance from the main shock.

3.4 Adelfio and Ogata (2010)

This paper follows others in the exploration of earthquakes affecting Japan. The authors fit an ETAS model accompanied by the integral transformation

$$\Lambda(t|\mathcal{H}_t) = \int_0^t \lambda(s|\mathcal{H}_s)dt$$

yielding a cumulative timeline from the theoretical number of events. The rescaled time units are signified by τ. The authors formulate a model with the separable density

$$\lambda(\tau, x, y) = \sum_{i=1}^{n} f(\tau - \tau_i)g(x - x_i, y - y_i).$$

The value of the spatial component $g(x, y)$ is $2\pi r \cdot g(r)$ of distance r from the origin in polar coordinates. The authors consider the following symmetric proposals for the temporal and spatial components:

$$\begin{aligned}
f(\tau) &= \left(\sqrt{2\pi\sigma}\right)^{-1} \exp\left\{-\tau^2/(2\sigma^2)\right\} \\
f(\tau) &= (2\sigma)^{-1} \exp\{-|\tau|/\sigma\} \\
f(\tau) &= (\beta - 1)\sigma^{\beta-1}/\{2(|\tau| + \sigma)^{\beta}\} \\
g(r) &= 1/(2\sigma\rho) \exp\left\{-r^2/(2\rho^2)\right\} \\
g(r) &= (\gamma - 1)\rho^{\gamma-1}/\left\{\pi(r^2 + \rho)^{\gamma}\right\},
\end{aligned}$$

preferring the final choices for f and g. The likelihood is maximized by cross-validation followed by returning the time coordinate to its original units. The results are compared to the parametric ETAS model. Graphical analysis is performed on the Gaussian kernel (equivalently, the pair of initial choices for f and g , the inverse power-law model (equivalently, the latter final choices for f and g) and the full parametrized ETAS model. The authors conclude that the ETAS model is as flexible as the kernel estimate, although the ETAS model has a drawback since $\mu(x, y)$ is not time dependent.

3.5 Marsan and Lengline (2010)

The authors expand previous research on the nonparametric estimation of the triggering function in branching point processes. Let ω_{ij} denote the probability that event i directly causes event j and ω'_{ik} be the probability that event i indirectly causes event k, namely that i triggers j and j triggers k. Assume earthquake separated by time $\triangle t$ interact according to $\lambda_{\triangle t}$. Arbitrary starting values are chosen for $\lambda_{\triangle t}$, since the following algorithm is known to converge. In the first step, calculate ω_{ij} as the percent share for $\lambda_{||j-i||}$, where $||j - i||$ represents the time between i and j. In the second step, update $\lambda_{\triangle t}$, which is the mean number of directly triggered aftershocks during the time interval $\triangle t$. The steps alternate until $\lambda_{\triangle t}$ is stable.

Triggering events condition the intensity. The distributions for the magnitude, time and temporality follow the laws as per Gutenberg-Richter, $r^{-\gamma}$ (power law for decay) and Omori-Utsu, respectively. The authors also consider distance from a quake's fault to the epicenter of its trigger, the number of faults, the number of earthquakes, and earthquakes in dislocated regions. California is shown as an example.

CHAPTER 4

Characterizing the second-moment statistics

Higher-order statistics aid the perception of a point process's properties. These functions assist an investigator in development of hypotheses. They might give answers on interaction or suggest deficiencies in the fitted model.

4.1 Baddeley et al. (2000)

The paper explains the theory to detect interaction using a nonparametric second-order (equivalently, second-moment) statistic named the K-function functions in inhomogeneous (equivalently, non-stationary) processes. Suppose Y is a second-order intensity-reweighted stationary point process. Stated otherwise the second-moment measure of Y is invariant after translation by the vector x. The inhomogenous K-function of Y is

$$K_{\text{inhom}}(t) = \frac{1}{|B|}\mathbb{E}\sum_{y_i \in Y \cap B}\sum_{y_j \in Y \setminus \{y_i\}} \frac{1(||y_i - y_j|| \leqslant t)}{\lambda(y_i)\lambda(y_j)}, \qquad t \geqslant 0.$$

The notation means that the points in Y are restricted to B. When the pair-correlation function exists and is isotropic (equivalently, the second-order measure depends only on the distance between two points), a point-wise, unbiased estimator of K_{inhom} is given by

$$\hat{K}_{\text{inhom}}(t) = \frac{1}{|W|}\sum_{y_i \in Y \cap W}\sum_{y_j \in Y \cup W \setminus \{y_i\}} \frac{w_{y_i,y_j} 1(||y_i - y_j||) \leqslant t)}{\lambda(y_i)\lambda(y_j)}, \qquad 0 \leqslant t \leqslant t^*,$$

In the former equation, w_{y_i,y_j} is Ripley's edge correction factor

$$t^* = \sup\{r \leqslant 0 : |\{s \in W : \partial B(s,r) \cap W \neq \emptyset\}| > 0\},$$

where $\partial B(s,r)$ denotes the boundary of $B(s,r)$.

4.2 Guan (2006)

The authors propose a new theory to investigate independence between marks and points and apply the results in data from an old-growth pine forest native to Arizona.. The authors supply test procedures – both conventional and graphical – based on their definition of the mark K and the mark G functions. The mark K and the mark G functions are analogous to the K and the nearest-neighbor function (equivalently, the G function) in spatial-point processes. By definition,

$$K_m(r) = \lambda^{-1} E\{m(x) \times N_r(x)\},$$
$$G_m(r) = E[m(x) \times I\{\delta(x) \leqslant r\}],$$

where $m(x)$ is a mark, $N_r(x)$ is the number of points in the process N that are within a distance or r from x and $\delta(x)$ is the distance from x to its nearest neighbor. The estimator $\hat{K}_m(r)$ is similar to $\hat{K}_p(r)$ from the spatial-point process with an adjustment to include marks and is defined by

$$\hat{K}_m(r) = \frac{1}{\hat{\lambda}^2} \sum \sum \frac{I(||x_i - x_j|| \leqslant r) \times m(x_i)}{A(W_{x_i} \cup W_{x_j})},$$

where the sums are over all distinct pairs x_i and x_j, $\hat{\lambda}$ is an estimator of λ, $A(W)$ is the volume of W and W_x is W translated by x. The estimator $\hat{G}_m(r)$ is defined by

$$\hat{G}_m(r) = \frac{1}{\hat{\lambda}} \sum_{i=1}^{n} \frac{I\{\delta(x_i) \leqslant r\} \times I\{x_i \in W_{\ominus \delta(x_i)}\} \times m(x_i)}{A\{W_{\ominus \delta(x_i)}\}}$$

A testable null hypothesis of independence, namely $K_m(r) = E\{m(x)\} \times K_p(r)$, implies that $\hat{K}_m(r)$ and $\hat{K}_p(r)$ are linearly related. For the conventional testing procedure, let $\hat{\mu}$ be an estimator of the slope observed from plotting $\hat{K}_m(r)$ against $\hat{K}_p(r)$. Let $\bar{M}$ be the average of the observed marks. The authors state $(\hat{\mu} - \bar{M})^2 / \widehat{\mathrm{Var}}(\hat{\mu} - \bar{M})$ converges to a χ_1^2 random variable.

The authors posit a permutation test for independence. Assume the marked, spatial-point process under consideration is homogeneous. Consider the average of the marks of a thinned process

$$T_1(r) = \frac{\sum m(x_i) \times I\{\delta(x_i \backslash x_1, \cdots, x_{i-1}) \leqslant r\}}{\sum I\{\delta(x_i \backslash x_1, \cdots, x_{i-1} < r\}}.$$

Subsequently, the p value from the two-sided test of the statistic $T_1(R)$ is

$$\frac{2 \times \min\{\#\text{ of } T(R) \leqslant T_1(R),\ \#\text{ of } T(R) \geqslant T_1(R)\}}{\#\text{ of random samples}}.$$

4.3 Adelfio and Schoenberg (2009)

Second-order statistics include covariance density, K-function (a measure of clustering), spectral density (a measure of periodic behavior), R/S statistic and correlation integral. A typical analysis of these types of statistics would reference a stationary Poisson distribution. This paper extends the utility of second-order statistics to an arbitrary conditional intensity function by weighting each point by the inverse of the conditional intensity function at the point's location. The value of the weighted second-order statistic is that the researcher is able to suspend any comparisons to the stationary distribution as well calculation of residual processes. The authors derive the asymptotic properties of the above statistics using martingale theory.

One may gradually build a statistic for the study of long-range dependence properties of temporal processes. Let the function $Z(t)$ be the sum of values from a simple process on the interval $[0, t]$. Let the difference of Z in the interval $[t, t + u]$ and the predicted value corresponding to the same interval estimating during $[t, t + \delta]$ be

$$D(u, t, \delta) = [Z(u + t) - Z(t)] - \frac{u}{\delta}[Z(t + \delta) - Z(t)].$$

The range over bins of lengths δ is given by

$$R(t; \delta) = \max_{0 \leq u \leq \delta} D(u, t, \delta) - \min_{0 \leq u \leq \delta} D(u, t, \delta).$$

Likewise the sample variance is specified by

$$S^2(t;\delta) = \frac{Z(t+\delta)+Z(t)}{\delta} + \left\{\frac{Z(t+\delta)-Z(t)}{\delta}\right\}^2.$$

Then the rescaled range statistic (R/S statistic) is defined as:

$$R/S = \frac{R(t;\delta)}{S(t;\delta)}$$

The slope of the graph of $\log R/S$ versus $\log\delta$ should be the H-constant as δ becomes large. The R/S statistic converges to the range of a Brownian bridge.

The authors present several measures of self-similarity. The correlation dimension is defined as

$$D_{\text{corr}} = \lim_{\delta\to 0}\frac{\log C(\delta)}{\log(\delta)},$$

where $C(\delta)$ is the number of points which have a smaller Euclidean distance that a given distance δ. The authors bring down the definition of a fractal defined by Mandelbrot (1977). An estimate for the correlation dimension can be obtained by the slope of the log-log plot for δ versus $\hat{C}_2$, where

$$\hat{C}_2(\delta) = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} I(|x_i - x_j| \leqslant \delta)$$

The authors pursue the idea further. In the case of a point process, a formula of a simple estimator of $K(\delta)$ is

$$\hat{K}(\delta) = \hat{\lambda}^{-1}\sum_i\sum_{i\neq j} I(|x_i - x_j| \leqslant \delta)/n,$$

where $\hat{\lambda}$ is observed number of events per unit of the area $\ell(A)$. The distribution of the $K-$function follows

$$\hat{K}(\delta) \xrightarrow{d} N\left(\pi\delta^2, \frac{2\pi\delta^2}{\lambda^2\ell(A)}\right)$$

The $K-$function is the expected mass in the circle centered at at a point with radius δ divided by the rate λ. Since

$$\hat{C}(\delta) \approx \frac{\hat{K}(\delta)}{\ell(A)},$$

the distribution of $\hat{C}(\delta)$ is

$$\hat{C}(\delta) \sim \left(\frac{\pi\delta^2}{\ell(A)}, \frac{2\pi\delta^2}{\ell(A)^2\lambda^2} \right)$$

for a Poisson point process.

The authors introduce a weighted process

$$N_w = \int_S \frac{1}{\lambda_*(s)} dN = \int_S \frac{\lambda_{\min}(s)}{\lambda(s)} dN,$$

where N is a process defined on a set $S \in \mathbb{R}^d$. The authors provide the properties for the expectation of the weighted process. Using this definition through the derivation of the R/S statistic yields R/S_w (equivalently, the weighted R/S statistic). The weighted correlation integral is the correlation integral weighted by the inverse intensity, namely

$$\hat{C}_W(\delta) = \frac{2}{n(n-1)} \sum_i^n \frac{1}{\lambda(s_i)} \sum_{j>i}^n \frac{1}{\lambda(s_j)} I(|s_i - s_j| \leqslant \delta).$$

This is estimated by

$$\hat{C}_W(\delta) = \frac{1}{\lambda_{\inf}^2 \ell(A)} \sum_i^n \frac{\lambda_{\min}}{\lambda(s_i)} \sum_{j \neq i}^n \frac{\lambda_{\min}}{\lambda(s_j)} I(|s_i - s_j| \leqslant \delta).$$

The weighted K-function becomes

$$\hat{K}_W(\delta) = \frac{1}{\lambda_{\inf}^2 \ell(A)} \sum_i^n \omega_j \sum_{j \neq i}^n \omega_j I(|(s_i - s_j)| \leq \delta),$$

where in this formula exclusively each λ term follows the null hypothesis and $\omega_k = \lambda_{\inf}/\lambda(s_k)$. Each of the weighted second order statistics are distributed as

$$\hat{K}_W^{(m)}(\delta) \xrightarrow{d} N \left(\pi\delta^2, \frac{2\pi\delta^2}{\ell(A)^{(m)} H((\lambda^{(m)})^2)} \right)$$

and

$$\hat{C}_W^{(m)}(\delta) \xrightarrow{d} N \left(\frac{\pi\delta^2}{\ell(A)^{(m)}}, \frac{2\pi\delta^2}{(\ell(A)^{(m)})^2 H((\lambda^{(m)})^2)} \right),$$

where $H((\lambda^{(m)})^2)$ is the harmonic mean of the squared intensity in the observer region $A^{(m)}$.

4.4 Gabriel and Diggle (2009)

This paper continues to develop the inhomogenous K-function. Whereas the K-function presented by Baddeley et al. (2000) depends upon the distance between two spatial locations, this method also considers temporal distance. The authors analyze incidence of Campylobacter infections of humans in England using a hypothesis test, a permutation test and a simulation to determine bias.

4.5 Guan (2009a)

Let an inhomogenous spatial point processes be called second-order intensity reweighed stationary processes if $\lambda(s_1, s_2) = \lambda(s_1)\lambda(s_2)g(s_1 - s_2)$, where $g(\cdot)$ is the pair-correlation function. Let $k(\cdot)$ denote a kernel function with bandwidth h. The empirical estimate of the PCF given from the author is

$$\hat{g}(t;h) = \frac{1}{2\pi h} \sum_{s_1,s_2 \in (N \cap D)}^{\neq}\sum e(s_1, s_2) \frac{k[(t - ||s_1 - s_2||)/h]}{\lambda(s_1;\hat{\beta})\lambda(s_2;\hat{\beta})||s_1 - s_2||}$$

where N is a spatial-point process and D is a spatial domain of interest. The author states

$$E[\hat{g}(t;h_n)] \approx \int_{\mathbb{R}} k(u)g(t - h_n u)du$$

and

$$\mathrm{Var}[\hat{g}(t;h_n)] \propto g(t;\theta)/t.$$

Ignoring the effect of $\hat{\beta}_n$ on $\hat{g}(t;h_n)$, this means that the estimated PCF is unbiased as long as $g(t)$ is continuous and h_n goes to 0.

The author desires a bandwidth selection procedure for an inhomogenous spatial-point process that rejects the trivial solution where $h = 0$. One may choose h to maximize

$$\sum_{s_1,s_2 \in (N \cap D)}^{\neq}\sum \left\{ \log[\tilde{g}(||s_1 - s_2||;h) - \log\left[\int_D \int_D \lambda(u;\hat{\beta})\lambda(v;\hat{\beta})\hat{g}(||u - v||;h)dudv\right]\right\},$$

where $\tilde{g}(||s_1 - s_2||; h)$ is the cross-validated version of $\hat{g}(||s_1 - s_2||; h)$ obtained by deleting the pair $(s1, s_2)$. The authors suggest the composite likelihood cross-validation criterion as a computationally faster alternative, namely

$$C(h) = \sum_{s_1, s_2 \in (N \cap D)}^{\neq} \sum W(s_1, s_2) \left\{ \log[\tilde{g}(||s_1 - s_2||; h)] - \log \left[\int_0^{r_h} t \hat{g}(t; h) dt \right] \right\},$$

with the weights defined as

$$W(s_1, s_2) = \frac{I(||s_1 - s_2|| \leqslant r_h)}{\lambda(s_1; \hat{\beta}) \lambda(s_2; \hat{\beta}) |D \cup D - s_1 + s_2|}.$$

The tuning parameter r_h is set to be around the dependence range, which can be observed in an empirical PCF plot based on a pilot bandwidth.

The authors report positive results for minimum contrast estimation in the simulation study.

The authors analyze the spatial distribution of three tree species established by the Center for Tropical Forest Science in Barro Colorado Island.

4.6 Guan (2009b)

The author estimates the variance of second-order statistics taken from intensity reweighed stationary processes. In preparatory remarks, the author recalls that a spatial point process is kth-order intensity reweighed stationary if its weight depends only on the interpoint lags. Such a process is isotropic if its weight is the distance between the two points. The author addresses nonparametric variance estimators for second-order statistics of the generic form $S_0(B, \hat{\beta}) = \sum_{u,v \in (N \cup B)}^{\neq} \sum \frac{f(u,v)}{\lambda(u; \hat{\beta}) \lambda(v; \hat{\beta})}$. An example is the K-function, which represents clustering or inhibition by values less than or greater than πt^2, respectively, at a distance of t. The functional form of the variance is stated, then a theoretical justification is produced. The knowledge is applied to tree locations on the Barro Colorado Island.

CHAPTER 5

Residual analysis

Residuals must be extended for spatial reasoning. Various choices of residuals are plausible. In any event checking the residuals is essential so that models can be evaluated, compared, and improved. Also several tests for interaction or separability have been designed.

5.1 Schoenberg (2003)

The author distinguishes epidemic-type aftershock sequence model by rescaled residuals and thinned residuals. Similar to Ogata (1998), the conditional intensity is modeled as

$$\lambda(t, \mathbf{x}, m) = f(m)[\mu(x, y) + \sum_{i; t_i < t} \nu(t - t_i, ||\mathbf{x} - \mathbf{x_i}||, m_i)],$$

$$f(m) \propto \exp\{-\beta(m - m_0)\},$$

and

$$\nu(t, x, y, m) = \frac{K_0 \exp\{\alpha(m - m_0)\}}{(t + c)^p (x^2 + y^2 + d)^q},$$

on account of the ETAS model, Gutenberg-Richter relation and the modified Omori law, respectively. The μ term is estimated from larger earthquakes with kernel smoothing, either by the simple case or by the nonisotropic distance orthogonal to the linear best fit. Under the hypothesis that the model matches the observations, points retained with probability $\min(\lambda(t_i, x_i, y_i, m_i))/\lambda(t_i, x_i, y_i, m_i)$ or alternatively

$$k\hat{\lambda}(t_i, x_i, y_i, m_i)^{-1} / \left(\sum_{i=1}^{N(S)} \hat{\lambda}(t_i, x_i, y_i, m_i)^{-1} \right),$$

will be distributed as homogenous Poisson process. Additionally a plot of the integral transform of points on one dimension versus a covariate will yield uniform dispersed residuals, provided the model is correct. The model is fitted to a catalog of earthquakes in Bear Valley, California.

5.2 Schoenberg (2004b)

The author thoroughly explores the topic of separability including its nature and scope and evaluates a multitude of nonparametric statistical tests with relevance. Let N be a random set of marks in a metric space $\mathcal{X} \in \mathbb{R}^{1+n+d}$ formed as a join of one temporal coordinate, n spatial coordinates, and a d-dimensional mark. The conditional intensity $\lambda(t, x, m)$ is the expected rate that the mark m will be recorded on the infinitesimally small point (t, x) of space-time, conditional on the history of the process prior to time t. A process is separable if $\lambda(t, x, m) = f(m)\lambda_1(t, x)$ or $\lambda(t, x, m) = f(m) + \lambda_1(t, x)$, where f is a nonnegative function and λ_1 is a nonnegative and predictable process. A process is completely separable if $\lambda(t, x, m) = \lambda_1(t) f_1(x) f_2(m)$. One may verify that clustering in spatial-temporal marked point processes is different from separability.

Estimate

$$\bar{\lambda}_1(t, x_1, \cdots, x_n) = \int_{\mathcal{X}} k_{n+1}(t - u, x_1 - y_1, \cdots, x_n - y_n) dN(u, y_1, \cdots, y_n, m),$$

and

$$\bar{f}(m) = \int_{\mathcal{X}} k_d(m - m') dN(t, x_1, \cdots, x_n, m'),$$

where k_{n+1} and k_d are $(n+1)$-dimensional and d-dimensional kernels, respectively. Also estimate

$$\hat{\lambda}(t, x_1, \cdots, x_n, m) = \int_{\mathcal{X}} k_{n+d+1}(t-u, x_1-y_1, \cdots, x_n-y_n, m-m')dN(u, y_1, \cdots, y_n, m')$$

and compare the conditional intensity estimates $\hat{\lambda}(t, x_1, \cdots, x_n, m)$ and $\tilde{\lambda}(t, x_1, \cdots, x_n, m) := \bar{\lambda}_1(t, x_1, \cdots, x_n)\bar{f}(m)/N(\mathcal{X})$. While a variety of tests are considered, the author substantiates the power of the Cramer-von Mises-type statistic given by

$$\int_0^T \int_{\mathbb{R}^n} \int_{\mathbb{R}^d} [\hat{\lambda}(t, x, m) - \tilde{\lambda}(t, x, m)]^2 dmdxdt.$$

The Cramer von-Mises type statistic is powerful amid ordered processes. What's more the L-function on the rescaled residuals is preferable in clustering or inhibition situations. The author draws an example from Los Angeles County wildfire data.

5.3 Baddeley et al. (2005)

The authors define point processes residuals like residuals for generalized linear models. The paper uses innovative graphics and references models from other papers. In addition, the authors created the `spatstat` package. The authors received an enthusiastic response.

The authors define conditional intensity, the quantity of interest, as $\lambda(t) = \mathbb{E}[dN_t|N_s, s < t]/dt$. The innovation or error process is $I(t) = N_t - \int_0^t \lambda(s)ds$. The innovation places a mass of 1 at each point x_i of the spatial-point process and a negative density $-\lambda(u, \mathbf{X})$ at all other spatial locations u. By setting $h(u, \mathbf{x}) = \mathbf{1}\{u \in B\}$, $\mathbb{E}_\theta[I_\theta(B)] = 0$ is satisfied. The raw residual process is $R(t) = N_t - \int_0^t \hat{\lambda}(s)ds$. The likelihood of the point process on the interval [0,t] is $L_\theta(t) = \left\{\prod_{t_i \leqslant t} \lambda_\theta(t_i)\right\} \exp\left\{-\int_0^t \lambda_\theta(s)ds\right\}$. A homogeneous Poisson process, inhomogeneous Poisson process, and pairwise interaction point process have densities

$f(\mathbf{x}) = \alpha\beta^{n(x)}$, $f(\mathbf{x}) = \alpha \prod_{i=1}^{n} b(x_i)$, and $f(\mathbf{x}) = \alpha \left\{ \prod_{i=1}^{n(\mathbf{x})} b(x_i) \right\} \prod_{i<j} c(x_i, x_j)$ respectively, where α represents a normalizing constant, $b(u) \geqslant 0, u \in W$, is the 'activity' and $c(u, v) = c(v, u) \geqslant 0, u, v \in W$, is the 'interaction'. A stationary pairwise interaction process has the function b constant and $c(u, v) = c(u - v)$. The 'hard core' process is obtained by setting $c(u, v) = \mathbf{1}\{\|u - v\| > \delta\}$, where $\delta > 0$, has $\lambda(u, x) = b(u)$ if $\|\mu - \eta\| > \delta$ for all points x_i in $\mathbf{x}$ and $\lambda(u, \mathbf{x}) = 0$ otherwise. To scale the raw residuals, one chooses an alternative of the function h. Then, the variance of the innovations is $\mathrm{var}\{I(B, h, \lambda)\} = \int_B \mathbb{E}[h(u, \mathbf{X})^2 \lambda(u, \mathbf{X})]du + \int_B \int_B \mathbb{E}[S(u, v, \mathbf{X})]dudv$, where $S(u, v, \mathbf{x} = \lambda(u, \mathbf{x})\lambda(v, \mathbf{x})h(u, \mathbf{x})h(v, \mathbf{x}) + \lambda(u, v, \mathbf{x})h(v, \mathbf{x} \cup \{u\})[h(u, \mathbf{x} \cup \{v\}) - 2h(u, \mathbf{x})]$.

5.4 Baddeley et al. (2008)

The article gives a definition of innovations and residuals for point processes. In addition these diagnostics' properties are covered, including first and second moments, variance deflation, conditional independence, a set-indexed martingale property, lack of correlation and marginal distributions. The h-weighted innovation is the signed random measure defined by

$$I_h(B) = \sum_{u \in X_B} h(u, X_W \backslash \{u\}) - \int_B h(u, X_W)\lambda(u, X_W)du.$$

The raw, inverse-λ, and Pearson innovations are special cases where $h = 1, 1/\lambda$ and $1/\sqrt{\lambda}$ respectively. The expected value of the innovations is zero.

Given a realization of a point process, the h-weighted residual becomes

$$R_{\hat{h}}(B) = \sum_{u \in X_B} h(u, X_W \backslash \{u\}) - \int_B h(u, X_W)\lambda(u, X_W)du.$$

The authors state that the the residuals are commonly biased for an inhomogenous Poisson process.

5.5 Guan (2008a)

The author calls $D_c(t;\hat{\theta}) = \int_A [\{r_c(x,t;\hat{\theta})\}^2 - N(x,t)]dx$ the discrepancy function, where the function r is the raw residual process from [6]. He shows that $T(\hat{\theta}) = \{D_c(\hat{\theta}) - \text{bias}(\hat{\theta})\}/\sigma_c(\hat{\theta})$ follows a standard normal, where $\sigma_c^2(\hat{\theta}) = 2\int_A \int_A \{\Lambda_c(x,y;\hat{\theta})\}^2 dxdy$ and $\Lambda_c(x,t;\hat{\theta}) = \int_{B(x,y)\cap A} \lambda_c(u;\hat{\theta})du$ with λ a Poisson variable. The test statistic $T(\hat{\theta})$ is rejected based on critical values from the standard normal. The author contemplates a spatial model to describe cases of larynx and lung cancer near an incinerator in northwest England.

5.6 Adelfio and Chiodi (2009)

The paper adapts residual analysis methods for point processes by weighting the original point process by the inverse of its conditional intensity function to yield new second-order statistics. The authors pursue the integral transformed process (equivalently, rescaling) and the thinned residual process with the idea in mind that these estimated models will follow the null distribution. For a model that is temporal point process, the authors say

$$\tau_i = \Lambda(t_i) = \int_0^{t_i} \lambda(t)dt$$

yields a residuals process. If there is a good fit between model and data, the residuals τ_i will follow a standard Poisson process. Likewise, the thinning method randomly yields its residuals by randomly retaining points with probability $\frac{\lambda_{\min}}{\lambda(t_i,x_i,y_i,m_i)}$, where the numerator is the minimum intensity in the observation region. A model exemplifying a good fit will have residuals that follow a homogenous Poisson process. In addition, one may analyze the weighted version of second-order statistics rather than the residuals.

The authors enable the creation of an autocorrelation plot using data weighted by the model. Accordingly, the weighted correlation integral for a time point process can be written as

$$\hat{C}_W(\delta) = \frac{1}{(\lambda_{\min} T)^2} \sum_i^n \omega_i \sum_{j \neq i}^n \omega_j I(|t_i - t_j| \leqslant \delta)$$

with $\omega_k = \frac{\lambda_{\min}}{\lambda(t_k)}$. By extension the definition of the weighted correlation integral is

$$\hat{C}_W(\delta) = \frac{2}{n(n-1)} \sum_i \frac{1}{\lambda(s_i)} \sum_{j>i} \frac{1}{\lambda(s_j)} I(|t_i - t_j| \leqslant \delta).$$

The authors use weighted residual processes evaluate various ETAS models describing earthquakes in Sicily and California according to the weighted second-order statistics. If after weighting the data the autocorrelation diagnostic plot exhibits significant correlation, this is a sign of misestimation.

CHAPTER 6

Conclusion

This survey gives many reasons to favor spatial-temporal point process models. One striking aspect is the power of quantitative analysis in spatial problems. This can be illustrated by the intensity and contour plots. Also remarkable is the ability to consider a plurality of models to interpret data. A particular case of interest is the electric signals model discussed in Vere-Jones (2009). The development of the subject is made possible by considering various models.

In conclusion, this survey gives a mere opening to the field. The previous sample includes biologists, ecologists, epidemiologists, firefighters, and seismologists using space-time point processes. For further awareness of current developments, see Bailey (2001), Bonneu and Thomas-Agnan (2009), Brillinger et al. (03), Cho et al. (2012), Cronie and Sarkka (2011), Faenza et al. (2003), Felzer et al. (2002), Gabriel et al. (2013), Grillenzoni (2005), Helmstetter and Sornette (2002a and 2002b), Kagan (1997), Kagan and Jackson (2000), Louie et al. (2010), Lewis et al. (2012), Mohler et al. (2011), Parsons et al. (2008), Rathbun and Cressie (1994), Rhoades and Eivson (2004), Stoyan (2006), Waagepetersen and Guan (2009), van Lieshout (2011), Veen and Schoenberg (2005), Vere-Jones and Schoenberg (2004), Xu and Schoenberg (2011), Zhuang (2005), Zhuang et al. (2002), and Zhuang et al. (2004). Admittedly there exist openings for further studies.

References

[1] Adelfio, G. and Chiodi, M. (2009). Second-order diagnostics for space-time point processes with application to seismic events. *Environmetrics*, 20(8):895–911.

[2] Adelfio, G. and Ogata, Y. (2010). Hybrid kernel estimates of space-time earthquake occurrence rates using the epidemic-type aftershock sequence model. *Annals of the Institute of Statistical Mathematics*, 62(1):127–143.

[3] Adelfio, G. and Schoenberg, F.P. (2009). Point process diagnostics based on weighted second-order statistics and their asymptotic properties. *Annals of the Institute of Statistical Mathematics*, 61(4):929–948.

[4] Baddeley, A., Moller, J. and Pakes, A.G. (2008). Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics*, 60(3):627–649.

[5] Baddeley, A., Moller, J. and Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350.

[6] Baddeley, A., Turner, R., Moller, J. and Hazelton, M. (2005). Residual analysis for spatial point processes. *Journal of the Royal Statistical Society*, 67(5):617–666.

[7] Bailey, T.C. (2001). Spatial statistical methods in health. *Cad. Saude Publica*, 17(5):1083–1098.

[8] Bonneu, F. and Thomas-Agnan, C. (2009). Spatial point process models for location-allocation problems. *Computational Statistics and Data Analysis*, 53(8):3070–3081.

[9] Brillinger, D.R., Preisler, H.K. and Benoit, J.W. (2003). Statistics and science: a festschrift for terry speed. chapter Risk Assessment: A Forest Fire Example, pages 177–196. Institue of Mathematical Sciences, Beachwood, OH.

[10] Cho, Y., Galstyan, A., Brantingham, J. and Tita, G. (2012). Generative models for spatial-temporal processes with applications to predictive criminology.

[11] Choi, E. and Hall, P. (1999). Nonparametric approach to analysis of space-time data on earthquake occurrences. *Journal of Computational and Graphical Statistics*, 8(4):733–748.

[12] Console, R., Murru, M. and Lombardi, A.M. (2003). Refining earthquake clustering models. *Journal of Geophysical Research*, 108(B10):ESE 5–1–ESE 5–9.

[13] Cronie, O. and Sarkka, A. (2011). Some edge correction methods for marked spatio-temporal point process models. *Computational Statistics and Data Analysis*, 55:2209–220.

[14] Diggle, P.J., Guan, Y., Hart, A.C., Paize, F. and Stanton, M. (2010). Estimating individual-level risk in spatial epidemiology using spatially aggregated information on the population at risk. *Journal of the American Statistical Association*, 105(492):1394–1402.

[15] Diggle, P.J., Kaimi, I. and Abellana, R. (2010). Partial-likelihood analysis of spatio-temporal point-process data. *Biometrics*, 66(2):347–354.

[16] Faenza, L., Marzocchi, W. and Boschi, E. (2003). A non-parametric hazard model to characterize the spatio-temporal occurrence of large earthquakes; an application to the italian catalogue. *Geophysical Journal International*, 155(2):521–531.

[17] Felzer, K.R., Becker, T.W., Abercrombie, R.E. and Ekstrom, G. (2002). Triggering of the 1999 mw 7.1 hector mine earthquake by aftershocks of the 1992 mw 7.3 landers earthquake. *Journal of Geophysical Research*, 107(B9):ESE 6–1–ESE 6–13.

[18] Gabriel, E. and Diggle, P.J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica*, 63(1):43–51.

[19] Gabriel, E., Rowlingson, B. and Diggle, P.J. (2013). Stpp: an r package for plotting, simulating and analyzing spatio-temporal point patterns. *Journal of Statistical Software*, 53(2).

[20] Grillenzoni, C. (2005). Non-parametric smoothing of spatio-temporal point processes. *Journal of Statistical Planning and Inference*, 128(1):61–78.

[21] Guan, Y. (2006). Tests for independence between marks and points of a marked point process. *Biometrics*, 62(1):126–134.

[22] Guan, Y. (2008). A goodness-of-fit test for inhomogeneous spatial poisson processes. *Biometrika*, 95(4):831–845.

[23] Guan, Y. (2008). On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. *Journal of the American Statistical Association*, 103(483):1238–1247.

[24] Guan, Y. (2009). A minimum contrast estimation procedure for estimating the second-order parameters of inhomogeneous spatial point processes. *Statistics and Its Interface*, 2:91–99.

[25] Guan, Y. (2009). On nonparametric variance estimation for second-order statistics of inhomogeneous spatial point processes with a known parametric intensity form. *Journal of the American Statistical Association*, 104(488):1482–1491.

[26] Helmstetter, A. and Sornette, D. (2002). Diffusion of epicenters of earthquake aftershocks, omori's law, and generalized continuous-time. *Physical Review E*, 66(6):061104–1–061104–24.

[27] Helmstetter, A. and Sornette, D. (2002). Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research*, 107(B10):ESE 10–1–ESE 10–21.

[28] Kagan, Y.Y. (1997). Are earthquakes predictable? *Geophysical Journal International*, 131:503–525.

[29] Kagan, Y.Y. and Jackson, D.D. (2000). Probabilistic forecasting of earthquakes. *Geophysical Journal International*, 143:438–453.

[30] Lewis, E., Mohler, G., Brantingham, P.J. and Bertozzi, A.L. (2012). Self-exciting point process models of civilian deaths in iraq. *Security Journal*, 25:244–264.

[31] Louie, K., Masaki, M. and Allenby, M. (2010). A point process model for simulating gang-on-gang violence.

[32] Marsan, D. and Lengline, O. (2008). Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079.

[33] Marsan, D. and Lengline, O. (2010). A new estimation of the decay of aftershock density with distance to the mainshock. *Journal of Geophysical Research*, 115(2010).

[34] Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P. and Tita, G.E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108.

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

[36] Ogata, Y. and Zhuang, J. (2006). Space-time etas models and an improved extension. *Tectonophysics*, 413(1-2):13–23.

[37] Parsons, N., Smith, J., Thonnes, E., Wang, L. and Wilson, R. (2008). Rotationally invariant statistics for examining the evidence from the pores in fingerprints. *Law, Probability and Risk*, 7(1):1–14.

[38] 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.

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

[40] Rathbun, S.L. and Cressie, N. (1994). A space-time survival point process for a longleaf pine forest in southern georgia. *Journal of the American Statistical Association*, 89(428):1164–1174.

[41] Rhoades, D.A. and Evison, F.F. (2004). Long-range earthquake forecasting with every earthquake a precursor according to scale. *pure and applied geophysics*, 161(1):47–72.

[42] Schlather, M., P. J. Ribeiro, J. and Diggle, P.J. (2004). Detecting dependence between marks and locations of marked point processes. *Journal of the Royal Statistical Society B*, 66(Part 1):79–93.

[43] Schoenberg, F.P. (2003). Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98(464):789–795.

[44] Schoenberg, F.P. (2004). Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference*, 128(1):79–83.

[45] Schoenberg, F.P. (2004). Testing separability in spatial-temporal marked point processes. *Biometrics*, 60(2):471–481.

[46] Schoenberg, F.P., Brillinger, D.R. and Guttorp, P. (2002). Point processes, spatial-temporal. *Encyclopedia of Environmentrics*, 3:1573–1577.

[47] Stoyan, D. (2006). On estimators of the nearest neighbour distance distribution function for stationary point processes. *Metrika*, 64:139–150.

[48] van Lieshout, M.N.M. (2011). A j–function for inhomogeneous point processes. *Statistica Neerlandica*, 65(2):183–201.

[49] Veen, A. and Schoenberg, F.P. (2005). *Case Studies in Spatial Point Process Models*, chapter Assessing spatial point process models for california earthquakes using weighted K-functions, pages 293–306. Springer, NY.

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

[51] Vere-Jones, D. and Schoenberg, F.P. (2004). Rescaling marked point processes. *Australian and New Zealand Journal of Statistics*, 46(1):133–143.

[52] Waagepetersen, R. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society B*, B(71):685–702.

[53] Xu, H. and Schoenberg, F.P. (2011). Point process modeling of wildfire hazard in los angeles county, california. *The Annals of Applied Statistics*, 5(2A):684–704.

[54] Zhuang, J. (2005). Multi-dimensional second-order residual analysis of space-time point processes and its applications in modelling earthquake data.

[55] Zhuang, J., Ogata, Y. and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380.

[56] Zhuang, J., Ogata, Y. and Vere-Jones, D. (2004). Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research*, 109(B5):B05301.