Point process diagnostics based on weighted second-order statistics and their asymptotic properties

Giada Adelfio · Frederic Paik Schoenberg

Abstract A new approach for point process diagnostics is presented. The method is based on extending second-order statistics for point processes by weighting each point by the inverse of the conditional intensity function at the point's location. The result is generalized versions of the spectral density, R/S statistic, correlation integral and K-function, which can be used to test the fit of a complex point process model with an arbitrary conditional intensity function, rather than a stationary Poisson model. Asymptotic properties of these generalized second-order statistics are derived, using an approach based on martingale theory.

Keywords Residual analysis · Point process · Second-order analysis · Conditional intensity function

1 Introduction

A major goal in the analysis of spatial-temporal point processes is the description of the second-order properties of the point process, such as its covariance density, K-function, spectrum, or measures of self-similarity and long-range dependence. Many tests have been proposed which are designed to test whether the second-order properties of an observed point pattern are consistent with the stationary Poisson process (Saw 1975; Davies 1977; Liebetrau 1978; Ripley 1977; Albrecht 1982; Sundt 1982; Diggle 1983; Lisek and Lisek 1985; Arsham 1987; Lawson 1988; Heinrich 1991).

G. Adelfio (✉)
Dipartimento di Scienze Statistiche e Matematiche "S. Vianelli", University of Palermo, Palermo, Italy
e-mail: adelfio@dssm.unipa.it

F. P. Schoenberg
Department of Statistics, University of California, Los Angeles, CA, USA
e-mail: frederic@ucla.stat.edu

More recently, attention has focused on methods and tests where the null hypothesis is a more general point process model rather than the stationary Poisson. For instance, Zhuang (2006) assesses the second-order residuals for various general space-time branching processes such as the epidemic-type aftershock sequence (ETAS) model. Another approach is to construct a stationary Poisson residual process by randomly rescaling (Meyer 1971; Schoenberg 1999) or thinning (Schoenberg 2003), and investigating whether the second-order properties of the observed residuals are consistent with those of the stationary Poisson process, as in Ogata (1988) or Schoenberg (2003). An alternative approach is to construct a weighted second-order statistic, where essentially to each observed point is given a weight inversely proportional to the conditional intensity at that point. This method was adopted by Veen and Schoenberg (2005) in constructing a weighted version of the K-function of Ripley (1977); the resulting weighted statistic is in many cases more powerful than residual methods (Veen 2006).

In this paper, we extend the weighting approach of Veen and Schoenberg (2005) to a variety of second-order statistics, proving that some weighted second-order statistics behave as the corresponding ones (not weighted) of a homogeneous Poisson process: departures suggest the unsuitability of the conditional intensity function used in the weighting scheme. In particular, results on the asymptotic equivalence of distributions of these residual and weighted estimates are derived here using martingale techniques. Investigations of convergence rates and of the sample sizes required to reach the asymptotic regime are not discussed here, however, and are considered important issues for future work.

Following a brief introduction of spatial-temporal point processes (Sect. 2) and some of their second-order characteristics (Sect. 3), the diagnostic method is introduced in Sect. 4 and weighted versions of those statistics are proposed in Sect. 4.1. This method is defined for general point processes, requiring only the existence of the conditional intensity at each location in space-time. Section 5 provides some concluding remarks and directions for future study.

2 Point processes and conditional intensity function

A *spatial-temporal point process* is a random point pattern defined by time and location of every single event. Point processes are introduced here by a mathematical approach that uses the definition of a counting measure on a set $X \subseteq \mathbb{R}^d$, $d \geq 1$, with positive values in $\mathbb{Z}$: for each Borel set B this $\mathbb{Z}_+$-valued random measure gives the number of events falling in B.

This section reviews some basic definitions related to point processes, reported to introduce the notation used throughout the paper. For further elaboration and references, please see Daley and Vere-Jones (2003).

Definition 1 (Point process) Let $(\Omega, \mathscr{A}, P)$ be a probability space and Φ a collection of locally finite counting measures on $X \subset \mathbb{R}^d$. Define $\mathscr{X}$ as the Borel σ-algebra of X and let $\mathscr{N}$ be the smallest σ-algebra on Φ, generated by sets of the form $\{\phi \in \Phi : \phi(B) = n\}$ for all $B \in \mathscr{X}$. A point process N on X is a measurable mapping of $(\Omega, \mathscr{X})$ into $(\Phi, \mathscr{N})$. A point process defined over $(\Omega, \mathscr{A}, P)$ induces a probability measure $\Pi_N(Y) = P(N \in Y), \forall Y \in \mathscr{N}$ (Cressie 1991).

Given a point process N defined on the space $(X, \mathscr{X})$ and a Borel set B, the number of points $N(B)$ in B is a random variable with first moment defined by:

$$\mu_N(B) = E[N(B)] = \int_\Phi \phi(B)\Pi_N(\mathrm{d}\phi)$$

that is a measure on $(X, \mathscr{X})$. The measure μ_N is called the mean measure or first moment measure of N (Cressie 1991). The second moment measure of N is given by:

$$\mu_N^{(2)}(B_1 \times B_2) = E[N(B_1)N(B_k)] = \int_\Phi \phi(B_1)\phi(B_k)\Pi_N(\mathrm{d}\phi),$$

with $B_1, B_2 \in \mathscr{X}$. If it is finite in $\mathscr{X}^{(2)}$ the process is second-order.

Let ds and du be small regions located at s and $u \in X$, and let $\ell(x)$ be the Lebesgue measure of x. The first-order intensity is defined by:

$$\eta(s) = \lim_{\ell(ds)\to 0} \frac{\mu_N(ds)}{\ell(ds)};$$

the second-order intensity is defined by:

$$\eta_2(s, u) = \lim_{\substack{\ell(ds)\to 0 \\ \ell(du)\to 0}} \frac{\mu_N^{(2)}(ds \times du)}{\ell(ds)\ell(du)}.$$

Let N be a point process on a spatial-temporal domain $X = \mathbb{R}^2 \times \mathbb{R}_+$; its conditional intensity function is defined by:

$$\lambda(t, \boldsymbol{x}|\mathscr{H}_t) = \lim_{dt,d\boldsymbol{x}\to 0} \frac{E[N([t, t+dt) \times [\boldsymbol{x}, \boldsymbol{x} + d\boldsymbol{x})|\mathscr{H}_t)]}{\ell(dtd\boldsymbol{x})}, \tag{1}$$

where $\mathscr{H}_t$ is the space–time occurrence history of the process up to time t, i.e. the σ-algebra of events occurring at times up to but not including t; $dt, d\boldsymbol{x}$ are time and space increments, respectively, and $E[N([t, t+t)\times[\boldsymbol{x}, \boldsymbol{x}+d\boldsymbol{x})|\mathscr{H}_t)]$ is the history-dependent expected value of occurrence in the volume $\{[t, t+\mathrm{d}t) \times [\boldsymbol{x}, \boldsymbol{x}+\mathrm{d}\boldsymbol{x})\}$. The conditional intensity function is a function of the point history and it is itself a stochastic process depending on the past up to time t. Assuming such a limit exists for each point $(t, \boldsymbol{x})$ in the space–time domain and the point process is simple, the conditional intensity process uniquely characterizes the finite-dimensional distributions of N (Daley and Vere-Jones 2003). If the conditional intensity function is independent of the past history, but dependent only on the current time and the spatial locations, (1) identifies a nonhomogeneous Poisson process. A constant conditional intensity characterizes a stationary Poisson process.

2.1 Compensator of point process

The integral of the conditional intensity function is the compensator of the point process. Indeed, the conditional intensity function of a space-time point process is a non-negative $\mathcal{H}_t$-predictable process $\lambda(t, \boldsymbol{x}|\mathcal{H}_t)$, such that for each Borel set B

$$N(B \times [0, t]) - \iint_B \int_0^t \lambda(t, \boldsymbol{x}|\mathcal{H}_t)\mathrm{d}\ell(t)\mathrm{d}\ell(\boldsymbol{x}) \tag{2}$$

is a $\mathcal{H}_t$-martingale, where $\ell(\cdot)$ denotes Lebesgue measure. If N is $\mathcal{H}_t$-measurable for all t, then it is said to be $\mathcal{H}$-adapted (Daley and Vere-Jones 2003).

An important theorem for the convergence of martingales (Hall and Heyde 1980) is here introduced and will be used in the following sections.

Theorem 1 (Central limit theorem for martingales) *If the martingale difference stochastic process $\{X_n, \mathcal{H}_n\}_{n=1}^\infty$, with $\mathcal{H}_n = \sigma(X_1, X_2, \ldots, X_n)$ and $E[X_i|\mathcal{H}_{i-1}] = 0$, $i = 1, 2, \ldots$, satisfies the following conditions:*

1. *Lindeberg condition: $E[(X_i)^2] < \infty$, $i = 1, 2, \ldots$ such that for any $\varepsilon > 0$:*

$$\lim_{n \to \infty} \left(\frac{1}{S_n^2} \sum_{i=1}^n E\left[X_i^2 I_{\{|X_i| > \varepsilon S_n\}} \right] \right) = 0,$$

   with $S_n^2 = \mathrm{var}\left[\sum_{i=1}^n X_i \right] \to \infty$ as $n \to \infty$;
2. *$E[E[X_i^2|\mathcal{H}_{i-1}]] = \sigma_i^2$, $i = 1, 2, \ldots$*
   then $\frac{1}{S_n} \sum_{i=1}^n X_i \xrightarrow{d} N(0, 1)$.

3 Second-order statistics for point processes

Some of the second-order statistics useful to describe observed point patterns are here listed. Statistics mainly used to describe point processes that exhibit long-range dependence and self-similarity are given in Sect. 3.1, which also provides theoretical results developed in subsequent sections.

In Sect. 3.2, the definition of the reduced second measure of a point process and fractal dimension are given, being useful quantities in the description of attractive and repulsive features for spatial point processes.

3.1 The R/S statistic

The rescaled range was introduced by Hurst, who monitored the Nile river flow and the minimum size for the river to not overflow or run dry over a given period of time (Hurst 1951). The R/S statistic was introduced as a graphical technique (Mandelbrot and Wallis 1969) for the study of long-range dependence properties of temporal processes. For the R/S analysis in point processes see for instance Ogata and Abe (1991), who

used this statistic to investigate the long-range dependence of Japanese earthquakes occurrence or Daley (1999) who studied long-range dependence for renewal point processes.

Consider a step function on $(t, t + \delta)$ (supposing the time is span in an arbitrary interval $(t, t + \delta)$ of length δ, $t + \delta \leq T$) for a sample configuration of a point process which jumps at each occurrence time, with a size $X(u) = Z(t + u) - Z(t)$. Let $Z(t)$ be the cumulative of jumps sizes at time t from the time origin and define:

$$D(u, t, \delta) = [Z(u + t) - Z(t)] - \frac{u}{\delta}[Z(t + \delta) - Z(t)] \tag{3}$$

the deviation of the cumulative from the average increase of the step function in the interval $(t, t + \delta)$, assuming the width of the time interval is such that $X(t) = 0$ if no event occurs in a neighbor of t and 1 otherwise (i.e. the process is simple). Then, the rescaled range statistic (R/S statistic) is defined by:

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

where $R(t; \delta)$ is the range of the data aggregated (by simple summation) over blocks of length δ, given by:

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

and $S^2(t; \delta)$ is the sample variance of the data aggregated at the same scale:

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

The log–log plot of R/S versus δ should have a constant slope H as δ becomes large and this slope is named H-constant or Hurst's number (Clegg 2006).

There are several methods for estimating H; the most used is the R/S plot (Pox diagram). It is a plot of the rescaled range statistic R/S against d on a log–log scale; for large d R/S is scattered around a straight line with slope H. To obtain a parametric estimate of H, the use of an approximate MLE is required, because of the slow changing of the covariance function in case of long-range dependence and the singularity of the covariance matrix of the estimates. The large sample theory for the maximum spectral likelihood has been developed by Whittle (1962).

Under the hypothesis of independence and identical distribution, the R/S statistic is asymptotically distributed as the range of a Brownian bridge (Mandelbrot 1975). On the basis of the extended continuous mapping theorem and the Herrnordf's theorem (Billingsley 1968; Lo 1991) showed that, $\frac{1}{\sqrt{n}}R/S$, for short-range dependent point processes (with mixing properties), converges to the range of a Brownian bridge. This result is easily generalized below for a point process; in the following theorem $X_j \; \forall j$, denotes the multiplicity of the j-th point, i.e. $X_j = N(dt_j)$ and let $\mu = E[N(dt)]$.

Theorem 2 (Asymptotic distribution of R/S, for short-dependent process) *Let* $\{X_n\}$ *be a stochastic process defined by* $X_n = \mu + \xi_n$, *where* μ *is a fixed arbitrary parameter and* ξ_n *is a zero mean random variable following assumptions of Herrndorf's theorem; then, as n increases:*

$$
\begin{aligned}
&1.\ \max_{1 \leq u \leq n} \frac{1}{S\sqrt{n}} \left\{ \sum_{j=1}^{u} (X_j - \bar{X}_n) \right\} \xrightarrow{d} \max_{0 \leq \tau \leq 1} B^o(\tau); \\
&2.\ \min_{1 \leq u \leq n} \frac{1}{S\sqrt{n}} \left\{ \sum_{j=1}^{u} (X_j - \bar{X}_n) \right\} \xrightarrow{d} \min_{0 \leq \tau \leq 1} B^o(\tau); \quad\quad (4)\\
&3.\ \frac{1}{\sqrt{n}} R/S \xrightarrow{d} \max_{0 \leq \tau \leq 1} B^o(\tau) - \min_{0 \leq \tau \leq 1} B^o(\tau).
\end{aligned}
$$

3.2 The correlation integral and the K-function

If the generator process presents self-similarity or scaling behavior, the fractal dimension provides some information about the scaling parameter or self-similarity index. Many fractal processes possess a self-similarity property: i.e. a small part of the surface, magnified, resembles a larger part of the surface. In this case, for a random self-similar fractal process in $\mathbb{R}^d$, the inference about its spatial scale can be carried out on a smaller scale, keeping in mind that when scaled, point processes are typically only self-scaling in a statistical sense; i.e. portions of the process resemble scaled versions of the entire process in distribution. Indeed, fractal processes are classified according to their degree of self-similarity, distinguishing between three different types of self-similarity for a fractal set. The strongest type of self-similarity is observed when a fractal process appears identical at different scales. A weaker form of self-similarity is yielded by fractals that appear approximately (but not exactly) identical at different scales. Finally, the weakest type of self-similarity is observed for fractals with statistical measures which are preserved across scales (statistical self-similarity). For instance, fractal dimension itself is a numerical measure which is preserved across scales.

The first definition of fractal was given by Mandelbrot (1977), who defined a fractal as a set for which the Hausdorff–Besicovitch dimension strictly exceeds its topological dimension.

Fractals are often characterized by their correlation dimension D_{corr}:

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

where $C(\delta)$ is the number of points which have a smaller (Euclidean) distance than a given distance δ. This measure is widely used because it is easy to evaluate for observed data. The correlation dimension is equivalent to the second-order Rényi point centered dimension (D_2) (Harte 2001). Define the *correlation integral*:

$$
C_2(\delta) = \int_{\mathscr{X}} \mu[S_\delta(x)]\mu(\mathrm{d}x) = Pr\{|X_1 - X_2| \leq \delta\},
$$

where X_1 and X_2 are d-dimensional independent random vectors sampled with respect to the probability measure μ, and $|\cdot|$ denotes the Euclidean distance.

Given a random process $X_1, X_2, \ldots, X_n$ defined on $\mathbb{R}^d$, the estimator of C_2 proposed by Grassberger and Procaccia (1983) is:

$$\hat{C}_2(\delta) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(|X_i - X_j| \le \delta),$$

with $I(\cdot)$ the Bernoulli indicator variable. The correlation dimension is estimated as the slope of the plot of $\log \delta$ versus $\log \hat{C}_2(\delta)$ for sufficiently small δ. Harte (1998) used the Hill (1975) estimator to estimate the correlation dimension.

As shown by Denker and Keller (1986), $\hat{C}_2(\delta)$ is a consistent estimator of $C_2(\delta)$ provided a general regularity condition is met. They proved the asymptotic normality of $\hat{C}_2^{(m)}(\delta)$ for the i.i.d. case and under mixing conditions by using the U-statistic estimators properties (Lee 1990).

The asymptotic distribution of the correlation integral can be easily provided also showing its strict relation with the K-function (Ripley 1976), which is a measure of the distribution of the inter-point distances and captures the spatial dependence between different regions of a point process. It is defined as the expected number of further events within distance δ of any given point divided by the overall rate λ, that is:

$$K(\delta) = \lambda^{-1} E[\text{no. of extra events within distance } \delta \text{ of an arbitrary event}], \quad \delta \ge 0 \tag{5}$$

Let N be a point process defined on $A \subseteq \mathbb{R}^2$ and $\{X_1, \ldots, X_n\}$ be n realizations of the process on A. A simple estimator of $K(\delta)$ is defined by:

$$\hat{K}(\delta) = \hat{\lambda}^{-1} \sum_i \sum_{i \ne j} I(|X_i - X_j| \le \delta)/n,$$

where $\hat{\lambda}$ is the estimator of the overall intensity given by $n/\ell(A)$, i.e. the observed number of events per unit of the area $\ell(A)$. Stoyan and Stoyan (2000) used $\tilde{\lambda}^2 = \frac{n(n-1)}{\ell(A)^2}$ as an estimator of the squared intensity function, and then:

$$\tilde{K}(\delta) = \frac{2\ell(A)}{n(n-1)} \sum_i^n \sum_{j>i}^n I(|X_i - X_j| \le \delta). \tag{6}$$

The properties of the K-function are well understood. Let N be a homogeneous Poisson process with rate λ with values in the subset A of $\mathbb{R}^2$ with finite area $\ell(A)$, and let the distances δ between the $\binom{n}{2}$ pairs of points be small compared with $\ell(A)$, as the area $\ell(A)$ tends to infinity:

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

(Cressie 1991, p. 642).

Therefore for planar process defined on $A \subseteq \mathbb{R}^2$ with finite area $\ell(A)$, on the basis of (6), it is possible to write:

$$\hat{C}(\delta) \approx \frac{\hat{K}(\delta)}{\ell(A)} \tag{7}$$

and therefore for a Poisson point process:

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

Hence, for a homogeneous Poisson process in $\mathbb{R}^2$ the correlation dimension is 2, that is the exponent of the power law growth of $C(\delta)$.

4 Second-order residuals

When a model is fitted to data, usually second-order statistics estimated for the residual process (i.e. the result of a thinning or rescaling procedure) are analyzed. Indeed, in the diagnostic theory of point processes, often two steps are needed: the transformation of data into residuals [thinning or rescaling Schoenberg 2003] and the use of tests to assess the consistency of the residuals with the homogeneous Poisson process. For instance, an estimate of the autocorrelation function of residuals could indicate the amount of dependence of data which is not described by the fitted model.

To interpret the goodness of fit of models to spatial-temporal point patterns, a new diagnostic tool is provided here. It is based on the interpretation of a weighted version of second-order statistics (such as autocorrelation, K-function, spectrum, fractal dimension and R/S statistic), calculated after weighting each point by the inverse of the conditional intensity. Weighted second-order statistics directly apply to data without assuming homogeneity or transforming data into residuals, eliminating the sampling variability due to the use of a transforming procedure. The probability used in the thinning method to retain points are replaced here by weights in order to offset the inhomogeneity of the process, with the advantage to include all the observed points rather than only the ones retained after the application of a random thinning. Moreover, this method can be applied to processes of any dimension, provided the statistics discussed here can be computed, and allows such second-order statistics to play a primary role in the diagnostic procedure, so that features such as clustering and inhibition can easily be interpreted.

4.1 The weighted process and its second-order properties

Let N be a point process defined on $S \in \mathbb{R}^d$, $d \geq 1$. For any point s in S, let $\lambda(s|\mathscr{F})$ be the conditional intensity function of the process with respect to some filtration $\mathscr{F}$ on S, for simplicity denoted by $\lambda(s)$. Then:

$$\lambda(s)\delta \approx E[N(s, s + \delta)|\mathscr{F}].$$

Since $\lambda(s)$, related to the probability that a point occurs near s, depends on the information on which the conditioning is based, it is a random process. In the proofs below λ is assumed to be positive and bounded away from zero.

N_w is defined as a real-valued random measure such that for any set S, $N_w(S) = \int_S \frac{1}{\lambda_*(s)} dN$ holds, with $\frac{1}{\lambda_*(s)} = \frac{\lambda_{\min}}{\lambda(s)}$ and assuming the existence of the positive constant $\lambda_{\min} \leq \inf\{\lambda(s); s \in S\}$.

The main purpose of this paper is to show, by formal proofs, that the main second-order statistics of $N_w(\cdot)$ behave similarly to those of a homogeneous Poisson process. For this purpose, first some results relative to the main features of N_w are provided. These properties will be used in proofs of subsequent results.

For simplicity, let N be a simple space–time process, such that $\mathcal{H}_i$ is the history of the process up to the ith event and with probability 1 all its points are distinct. Then, since $\lambda(s_i)$ is measurable with respect to $\mathcal{H}_i$:

$$\lim_{\delta \to 0} \frac{E[N_w(s_i, s_i + \delta)]}{\delta} = \lambda_{\min} \lim_{\delta \to 0} E\left[\frac{1}{\lambda(s_i)} \frac{E\left[N(s_i, s_i + \delta) \big| \mathcal{H}_i\right]}{\delta}\right] = \lambda_{\min}.$$

Assuming that the original process N is orderly, the second moment of N_w may be derived similarly:

$$\lim_{\delta \to 0} \frac{E[N_w(s_i, s_i + \delta)^2]}{\delta}$$
$$= \lambda_{\min}^2 \lim_{\delta \to 0} E\left[E\left[\left(\frac{1}{\lambda(s_i)} N(s_i, s_i + \delta)\right)^2 \Big| \mathcal{H}_i\right]\right] \frac{1}{\delta}$$
$$= \lambda_{\min}^2 \lim_{\delta \to 0} E\left[\frac{1}{(\lambda(s_i))^2} \frac{E\left[N^2(s_i, s_i + \delta) \big| \mathcal{H}_i\right]}{\delta}\right]$$
$$= \lambda_{\min}^2 E\left[\frac{1}{\lambda(s_i)}\right]. \tag{8}$$

Suppose $i < j$; since $\lambda(s_i)$ and $\lambda(s_j)$ are both $\mathcal{H}_j$-measurable and $\lambda(s_i)$ is measurable with respect to $\mathcal{H}_i$, the mixed moment of the process $N_w(S) = \frac{N(S)}{\lambda_*(s)}$ is:

$$\lim_{\delta \to 0} \frac{E[N_w(s_i, s_i + \delta) N_w(s_j, s_j + \delta)]}{\delta^2}$$
$$= \lambda_{\min}^2 \lim_{\delta \to 0} E\left[E\left[\frac{1}{\lambda(s_i)\lambda(s_j)} N(s_i, s_i + \delta) N(s_j, s_j + \delta) \Big| \mathcal{H}_j\right]\right] \frac{1}{\delta^2}$$
$$= \lambda_{\min}^2 \lim_{\delta \to 0} E\left[\frac{1}{\lambda(s_i)\lambda(s_j)} N(s_i, s_i + \delta) \frac{E\left[N(s_j, s_j + \delta) \big| \mathcal{H}_j\right]}{\delta}\right] \frac{1}{\delta}$$

$$= \lambda_{\min}^2 \lim_{\delta \to 0} E\left[E\left[\frac{1}{\lambda(s_i)} N(s_i, s_i + \delta) \Big| \mathscr{H}_i \right] \right] \frac{1}{\delta}$$

$$= \lambda_{\min}^2 \lim_{\delta \to 0} E\left[\frac{1}{\lambda(s_i)} \frac{E\left[N(s_i, s_i + \delta) \Big| \mathscr{H}_i \right]}{\delta} \right]$$

$$= \lambda_{\min}^2$$

(see also Appendix B in Zhuang (2006)). Thus the covariance is:

$$
\begin{aligned}
&\text{cov}[N_w(s_i, s_i + \delta), N_w(s_j, s_j + \delta)] \\
&= E[N_w(s_i, s_i + \delta) N_w(s_j, s_j + \delta)] - E[N_w(s_i, s_i + \delta)] E[Nw(s_j, s_j + \delta)] \\
&= 0.
\end{aligned}
\tag{9}
$$

If N is a Poisson process with intensity function λ its power spectrum is:

$$f_N(\omega) = \frac{1}{2\pi} \lambda$$

(Bartlett 1964). From (9) the spectral density of the weighted process (that is the spectral density calculated on the weighted points by using the inverse of the conditional intensity function multiplied by its minimum value) reduces to

$$f_{N_w}(\omega) = \frac{\lambda_{\min}}{2\pi},$$

which is the power spectrum of Poisson process with constant rate $\lambda_{\min}$.

Moreover, from the results above a martingale is obtained by the weighted process N_w. As expressed by (2), it is known that $N(S \times [0, t]) - \iint_S \int_0^t \lambda(s, t | \mathscr{H}_t) \mathrm{d}s \mathrm{d}t$ is a martingale.

Theorem 3 (Martingale characterization of N_w) *Let N be a space–time process defined on $S \times [0, t]$ with conditional intensity function $\lambda = \lambda(\cdot | \mathscr{H}_t)$ positive and bounded away from zero, with $\mathscr{H}_t$ the past history of the process up to time t and $N_w = N \frac{\lambda_{\min}}{\lambda}$, with $\lambda_{\min} \leq \inf\{\lambda\}$.*

The process defined by:

$$E[M(S \times [0, t + \delta]) | \mathscr{H}_t] = M(S \times [0, t]) \tag{10}$$

is a $\mathscr{H}$-martingale, for each $\delta > 0$.

4.1.1 Weighted R/S statistic

In this section, the distributional theorem introduced in Sect. 3.1 is generalized for a weighted version of the R/S statistic. On the basis of Theorem 2 and the results above, a distributional theorem is derived.

Let $N_w(t)$ be the weighted cumulative process at time t from the time origin, with $N_w(t, t + \delta) = \frac{\lambda_{\min}}{\lambda(t)}$ if N has an event in a neighbor of t and 0 otherwise and:

$$D_w(u, t, \delta) = [N_w(u + t) - N_w(t)] - \frac{u}{\delta}[N_w(t + \delta) - N_w(t)]$$

the weighted version of the deviation from the average in the interval $(t, t + \delta)$ [see (3)]. The weighted rescaled range is:

$$R_w(t; \delta) = \max_{0 \leq u \leq \delta} D_w(u, t, \delta) - \min_{0 \leq u \leq \delta} D_w(u, t, \delta)$$

and, for the simple process assumption, the sample variance of the random variable $N_w(t, t + d) = N_w(t + d) - N_w(t)$ around the sample average is:

$$S_w^2(t; \delta) = \frac{N_w(t + \delta) - N_w(t)}{\delta} - \left[\frac{N_w(t + \delta) - N_w(t)}{\delta} \right]^2 .$$

Thus, as a diagnostic tool, the weighted (or residual) R/S statistic can be defined:

$$R/S_w = \frac{R_w(t; \delta)}{S_w(t; \delta)} .$$

To show the consistency of the R/S_w statistic with the R/S statistic of a homogeneous Poisson process, assumptions of Theorem 1 are proved.

Theorem 4 (Asymptotic distribution of the statistic R/S_w) *Let N be a temporal point process with conditional intensity function $\lambda(t|\mathcal{H})$ positive and bounded away from zero. Assume that $E[N^2(\delta t)|\mathcal{H}_t] = E[N^2(t, t + \delta)|\mathcal{H}_t] < \infty$ and that there exists $\alpha > 0$ such that $E[N^{2+\alpha}(\delta t)|\mathcal{H}_t] = E[N^{2+\alpha}(t, t + \delta)|\mathcal{H}_t]$ is bounded.*

Then the R/S statistic of the weighted process $N_w(t, t + \delta) = N(t, t + \delta)\frac{\lambda_{\min}}{\lambda(t)}$, with $\lambda_{\min} \leq \inf\{\lambda(t); t \in [0, T]\}$, weakly converges to the range of a Brownian bridge.

Proof It has been shown that $N_w(t, t + \delta) - \delta\lambda_{\min}$ is a martingale difference process [see Theorem 3 and Eq. (10)]. Define increments X_i of the weighted process as:

$$X_i = N_w(t_i, t_i + \delta) - \delta\lambda_{\min},$$

with $N_w(t_i, t_i + \delta) = N(t_i, t_i + \delta)\frac{\lambda_{\min}}{\lambda(t_i)}$. From the previous results relative to the weighted process N_w, the following holds:

$$S_n^2 = \text{var}\left[\sum_i^n X_i \right] = \sum_i^n \text{var}[X_i] = \sum_i^n E[X_i^2].$$

Hence $\varepsilon S_n \to \infty$ as $n \to \infty$.

To prove condition *1* (Lindeberg condition) of Theorem 1, write:

$$\sum_{i=1}^{n} E\left[X_i^2 I_{\{|X_i|>\varepsilon S_n\}}\right]$$

$$\leq \sum_{i=1}^{n} \left(E\left[|X_i|^{2+\alpha}\right]\right)^{2/(\alpha+2)} P(|X_i| > \varepsilon S_n)^{\alpha/(\alpha+2)}$$

$$\leq \left(\sum_{i=1}^{n} E\left[|X_i|^{2+\alpha}\right]^{2/(\alpha+2)}\right)\left(\sum_{i=1}^{n} P(|X_i| > \varepsilon S_n)\right)^{\alpha/(\alpha+2)}$$

from the Hölder inequality. Since from the Chebyshev inequality:

$$\sum_i^n P(|X_i| \geq \varepsilon S_n) \leq \frac{E(X_i)^2}{\varepsilon^2 S_n^2} = \frac{1}{\varepsilon^2}$$

and $E\left[X_i^{2+\alpha}\right]$ remains bounded (from the assumption about $E[N^{2+\alpha}(\delta t)|\mathcal{H}_t]$), therefore $\sum_{i=1}^{n} E\left[X_i^2 I_{\{|X_i|>\varepsilon S_n\}}\right]$ grows less fast than S_n^2 and the Lindeberg condition holds.

Since $E[N^2(t, t + \delta)|\mathcal{H}_t]$ is bounded, condition 2 of Theorem 1 is implied by (8).

The asymptotic Normal distribution and, therefore, the independent increments of the process N_w for each t imply that the sample path obtained posing $\sigma^2 = E[X_i^2]\forall i$, converges to a Wiener process.

Hence, as in Theorem 2, the weak convergence of the weighted R/S statistic to the range of a Brownian bridge is proved, based on simple assumptions about the second moments of the point process. These assumptions are not restrictive for a general point process; indeed they hold for any non-explosive model, such as a Hawkes process with a branching ratio less than one.

4.1.2 Weighted correlation integral and weighted K-function

Let N be a planar point process defined on A, a subregion of $\mathbb{R}^2$ with Lebesgue measure $\ell(A)$, with $I(\cdot)$ the Bernoulli indicator variable, s_k, $\forall k$, points of the state space and $\delta > 0$.

We define the weighted correlation integral by:

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

where $\lambda(s)$ is the conditional intensity function of the process with respect to some filtration $\mathcal{F}$ on A; note that the weighted correlation integral can be approximated by:

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

To prove the asymptotic distributional properties of the weighted correlation integral for inhomogeneous Poisson processes, the analogy with the weighted K-function (Baddeley et al. 2000) can be considered, as in Sect. 3.2 for the unweighted version of these statistics. Let $\lambda_0(x, y)$ be the intensity function of the model that under the null hypothesis describes the observed process on an interval $A \subseteq \mathbb{R}^2$ of area $\ell(A)$. Then the weighted K-function is defined by:

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

where $\lambda_{\inf} = \inf\{\lambda_0(s); (s) \in A\}$ is the minimum of the conditional intensity over the observed region under the null hypothesis and, for each k, $\omega_k = \lambda_{\inf}/\lambda_0(s_k)$, with $\lambda_0(s_k)$ the conditional intensity at the point s_k of A under H_0 and $\delta > 0$.

Combining the point process residual analysis techniques and the use of the K-function as a diagnostic tool applied to residual processes, Veen and Schoenberg (2005) provided theorems on the distributional properties of the K-function and its weighted variant, considering, as the null hypothesis models, homogeneous and inhomogeneous Poisson processes. Formally, let $N^{(m)}$, $m = 1, 2, \ldots, M$ be a sequence of inhomogeneous Poisson processes with intensities $\lambda^{(m)}$ and K-functions $K_W^{(m)}(\delta)$, defined on the subsets $A^{(m)} \in \mathbb{R}^2$ of areas $\ell(A)^{(m)}$, such that:

- each $A^{(m)}$, $\forall m$, is obtained as the union of disjoint subsets $A_1^{(m)}, \ldots, A_M^{(m)}$ of areas $\ell(A)_1^{(m)}, \ldots, \ell(A)_M^{(m)}$, with M that tends to infinity,
- the conditional intensity within each subset $A_k^{(m)}$, denoted by $\lambda_i^{(m)}$, is approximately constant,
- δ is such that $\sup_k \frac{\pi \delta^2}{\ell(A)_k^{(m)}} \to 0$,
- properties of regularity of A hold.

From the central limit theorem, and from the fixed assumptions, as $m \to \infty$, the authors provided the following result:

$$\hat{K}_W^{(m)}(\delta) \xrightarrow{d} N\left(\pi\delta^2, \frac{2\pi\delta^2}{\ell(A)^{(m)} H((\lambda^{(m)})^2)}\right), \tag{11}$$

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

Therefore, from (11) and equation (7), it follows that:

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

Since the previous result can be proved only under some restrictive assumptions, like the approximate constancy of the conditional intensity on discs of area $\pi\delta^2$, a more general result is provided below by using martingales theory.

To show the asymptotic normality of $\hat{C}_W(\delta)$, the martingale theorem is used, first considering the temporal domain. The weighted correlation integral, for a time point process N with realizations $t_1, t_2, \ldots, t_n$ on $[0, T] \in \mathbb{R}$ with Lebesgue measure T, can be written as:

$$\hat{C}_W(\delta) \approx \frac{1}{(\lambda_{min}T)^2} \sum_i^n \omega_i \sum_{j \neq i}^n \omega_j I(|t_i - t_j| \leq \delta), \tag{12}$$

with $\omega_k = \frac{\lambda_{min}}{\lambda(t_k)}, \forall k$ and $\lambda(t)$ the conditional intensity function of the process with respect to some filtration $\mathscr{H}_t$ on $[0, T]$.

Define:

$$I_{uv}^w(\delta)|_T = 1_{\{0<|u-v|<\delta,T\}} \frac{\lambda_{min}^2}{\lambda(u)\lambda(v)} \tag{13}$$

with $1_{\{0<|u-v|<\delta,T\}}$ the indicator function that is 1 if the distance between the two points of the pair $\{t_u, t_v\}$, such that $u \neq v$ and $\{t_u, t_v\}_{u \neq v} \in [0, T]$, is both strictly greater than zero and less than δ, and zero otherwise. The following result holds:

$$E\left[\int_{\mathbb{R}\times\mathbb{R}} I_{uv}^w(\delta)|_T \, dN(u)dN(v)\right]$$

$$= E\left[\int_{\mathbb{R}\times\mathbb{R}} 1_{\{0<|u-v|<\delta,T\}} \frac{\lambda_{min}^2}{\lambda(u)\lambda(v)} dN(u)dN(v)\right]$$

taking conditional expectations on $\mathscr{H}_u$ and on $\mathscr{H}_v$

$$= E\left[\int_{\mathbb{R}\times\mathbb{R}} 1_{\{0<|u-v|<\delta,T\}} \frac{\lambda_{min}^2}{\lambda(u)\lambda(v)} \lambda(u)\lambda(v) d\ell(u)d\ell(v)\right]$$

$$= E\left[\lambda_{min}^2 \int_{\mathbb{R}\times\mathbb{R}} 1_{\{0<|u-v|<\delta,T\}} d\ell(u)d\ell(v)\right]$$

$$= \lambda_{min}^2 T\delta.$$

Therefore:

$$E\left[\hat{C}_W(\delta)\right] = \frac{1}{T^2}[T\delta] = \frac{\delta}{T}.$$

Let $I_{uv}^w(\delta)|_{t_i}$ and $I_{uv}^w(\delta)|_{t_i-t_{i-1}}$ be defined as $I_{uv}^w(\delta)|_T$, conditioning on the points that occur before t_i and in the interval $[t_{i-1}, t_i]$, for any $i = 1, 2, \ldots, n$, respectively.

Let us define, for any i, $\int_{\mathbb{R}\times\mathbb{R}} I_{uv}^w(\delta)|_{\tau_i} dN(u)dN(v)$ as the number of pairs of points with elements occurring both up to τ_i, that is the last point less than or equal to t_i such that no points are in $(\tau_i, \tau_i + \delta)$: $\tau_i = \sup\{\tau : N(\tau) = 1, N(\tau, \tau + \delta) = 0, \tau + \delta < t_i\}$

letting $\tau_i = 0$ if no such τ exists. In other words, τ_i is the left end-point of the last gap prior to t_i of size at least δ.

Define $Z(t_i) = \int_{\mathbb{R} \times \mathbb{R}} I_{uv}^w(\delta)|_{\tau_i} \mathrm{d}N(u) \mathrm{d}N(v) - \lambda_{\min}^2 \delta \tau_i$, for any i, assuming that such kind of gaps exists. The existence of a gap of size δ corresponding to each t_i guarantees that, conditioning on $Z(t_{i-1})$, no pair of points within distance δ crosses t_{i-1} and therefore the knowledge about the past up to any point t_{i-1} does not give any information about t_i. Note that for each i, τ_i is an $\mathcal{H}_{t_i}$-stopping time, and thus $Z(t_i)$ is measurable [see e.g. Corollary A3.4.VIII on p. 430 of Daley and Vere-Jones (2003)].

Theorem 5 (Martingale characterization) *Let N be a temporal point process with conditional intensity function $\lambda(t|\mathcal{H})$ positive and bounded away from zero. Let $Z(t_i) = \int_{\mathbb{R} \times \mathbb{R}} I_{uv}^w(\delta)|_{\tau_i} \mathrm{d}N(u) \mathrm{d}N(v) - \lambda_{\min}^2 \delta \tau_i$ for any $i = 1, \ldots, n$, $\delta > 0$, with $\int_{\mathbb{R} \times \mathbb{R}} I_{uv}^w(\delta)|_{\tau_i} \mathrm{d}N(u) \mathrm{d}N(v)$ the number of pairs of points with elements occurring both up to τ_i and $\tau_i = \sup\{\tau : N(\tau) = 1, N(\tau, \tau + \delta) = 0, \tau + \delta < t_i\}$.*

Then $Z(t_i)$ is a martingale with respect to a filtration $\mathcal{H}_i$, i.e.:

$$E[Z(t_i) - Z(t_{i-1})|Z(t_{i-1})] = 0. \qquad (14)$$

Proof Note that:

$$
\begin{aligned}
&E[Z(t_i) - Z(t_{i-1})|Z(t_{i-1})] \\
&= E\left[\left(\int_{\mathbb{R} \times \mathbb{R}} I_{uv}^w(\delta)|_{\tau_i} \mathrm{d}N(u)\mathrm{d}N(v) - \int_{\mathbb{R} \times \mathbb{R}} I_{uv}^w(\delta)|_{\tau_{i-1}} \mathrm{d}N(u)\mathrm{d}N(v)\right)\right. \\
&\qquad \left. -\lambda_{\min}^2 \delta \left(\tau_i - \tau_{i-1}\right) \left| Z(t_{i-1})\right. \right] \\
&= E\left[\int_{\mathbb{R} \times \mathbb{R}} I_{uv}^w(\delta)|_{(\tau_i - \tau_{i-1})} \mathrm{d}N(u)\mathrm{d}N(v) - \lambda_{\min}^2 \delta \left(\tau_i - \tau_{i-1}\right) \left| Z(t_{i-1})\right.\right] \\
&= 0.
\end{aligned}
$$

Theorem 6 (Asymptotic distribution of $\hat{C}_W(\delta)$) *Let N be a temporal point process with conditional intensity function $\lambda(t|\mathcal{H})$ positive and bounded away from zero, such that $E[N^2(\mathrm{d}t_i)|\mathcal{H}_i] = E[N^2(t_i, t_i + \mathrm{d}t_i)|\mathcal{H}_i]$ $\forall i$ is bounded and there exists an $\alpha > 0$ such that $E[N^{2+\alpha}(\mathrm{d}t_i)|\mathcal{H}_i] = E[N^{2+\alpha}(t_i, t_i + \mathrm{d}t_i)|\mathcal{H}_i]$ $\forall i$ is bounded. Moreover, assume that for any i there exists $\tau_i = \sup\{\tau : N(\tau) = 1, N(\tau, \tau + \delta) = 0, \tau + \delta < t_i\}$, $\delta > 0$.*

Then the weighted correlation dimension defined as in (12) *is asymptotically normally distributed.*

Proof In Theorem 5 we have already proved that $Z(t_i)$ is a martingale and, therefore, the validity of the assumptions of Theorem 1 are proved.

Define, for any i, the martingale difference process associated to Z by:

$$X_i = \int_{\mathbb{R} \times \mathbb{R}} I_{uv}^w(\delta)|_{(\tau_i - \tau_{i-1})} \mathrm{d}N(u)\mathrm{d}N(v) - \lambda_{\min}^2 \delta \left(\tau_i - \tau_{i-1}\right).$$

Then $E[X_i^2]$ is obtained:

$$
E\left[\left(\int_{\mathbb{R}\times\mathbb{R}} I_{uv}^w(\delta)|_{\tau_i - \tau_{i-1}} \mathrm{d}N(u)\mathrm{d}N(v) - \lambda_{\min}^2 \delta(\tau_i - \tau_{i-1})\right)^2\right]
$$

$$
= E\left[\left(\int_{\mathbb{R}\times\mathbb{R}} \frac{\lambda_{\min}^2}{\lambda(u)\lambda(v)} 1_{\{0<|u-v|<\delta, \tau_i-\tau_{i-1}\}} \mathrm{d}N(u)\mathrm{d}N(v) - \lambda_{\min}^2 \delta(\tau_i - \tau_{i-1})\right)^2\right]
$$

$$
= E\left[\left(\int_{\mathbb{R}\times\mathbb{R}} \frac{\lambda_{\min}^2}{\sqrt{\lambda(u)\lambda(v)}\sqrt{\lambda(u)\lambda(v)}} 1_{\{0<|u-v|<\delta, \tau_i-\tau_{i-1}\}} \mathrm{d}N(u)\mathrm{d}N(v)\right)^2\right]
$$

$$
- \lambda_{\min}^4 \delta^2 (\tau_i - \tau_{i-1})^2
$$

for the Cauchy–Schwartz inequality

$$
\leq E\left[\int_{\mathbb{R}\times\mathbb{R}} \frac{\lambda_{\min}^2}{\lambda(u)\lambda(v)} \mathrm{d}N(u)\mathrm{d}N(v) \int_{\mathbb{R}\times\mathbb{R}} \frac{\lambda_{\min}^2}{\lambda(u)\lambda(v)} 1_{\{0<|u-v|<\delta, \tau_i-\tau_{i-1}\}} \mathrm{d}N(u)\mathrm{d}N(v)\right]
$$

$$
- \lambda_{\min}^4 \delta^2 (\tau_i - \tau_{i-1})^2
$$

$$
\leq E\left[\left(\int_{\mathbb{R}\times\mathbb{R}} \frac{\lambda_{\min}^2}{\lambda(u)\lambda(v)} \mathrm{d}N(u)\mathrm{d}N(v)\right)^2\right] - \lambda_{\min}^4 \delta^2 (\tau_i - \tau_{i-1})^2
$$

$$
\leq E\left[\left(\int_{\mathbb{R}\times\mathbb{R}} \frac{\lambda_{\min}^2}{\lambda(u)\lambda(v)} \mathrm{d}N(u)\mathrm{d}N(v)\right)^2\right]
$$

$$
= E\left[\sum_{u\neq v}^{n}\left(\frac{\lambda_{\min}^2 N(u)N(v)}{\lambda(u)\lambda(v)}\right)^2\right] + 2E\left[\sum_{u\neq v}^{n} \frac{\lambda_{\min}^2 N(u)N(v)}{\lambda(u)\lambda(v)} \sum_{u'\neq v'}^{n} \frac{\lambda_{\min}^2 N(u')N(v')}{\lambda(u')\lambda(v')}\right]
$$

from the assumptions

$$
= E\left[\sum_{u\neq v}^{n}\left(\frac{\lambda_{\min}^4}{(\lambda(u)\lambda(v))^2} E_u^2 E_v^2\right)\right] + 2\lambda_{\min}^4 (\tau_i - \tau_{i-1})^2, \tag{15}
$$

with $E_i^2 = E[N^2(\mathrm{d}t_i)|\mathcal{H}_i]$, $\forall\, i$. Since λ is bounded, the summand in (15) is bounded and, then, the expected number of points is finite.

If the process N is simple, (15) becomes:

$$
E\left[\sum_{u\neq v}^{n}\left(\frac{\lambda_{\min}^4}{\lambda(u)\lambda(v)}\right)\right] + 2\lambda_{\min}^4 (\tau_i - \tau_{i-1})^2.
$$

Thus, $E[X_i^2] < \infty \ \forall i$. If N is simple the Lindeberg condition is immediately proved; indeed, since:

$$S_n^2 = \text{var}\left[\sum_i^n X_i\right] = \sum_i^n \text{var}\,[X_i]$$

and $\varepsilon S_n \to \infty$ as $n \to \infty$, for any $\alpha > 0$ $E[N^{2+\alpha}(\mathrm{d}t_i)|\mathscr{H}_i] = E[N(\mathrm{d}t_i)|\mathscr{H}_i] \approx \lambda(t_i)\mathrm{d}t_i$ and $E[X_i]^{2+\alpha}$ is surely bounded. Therefore:

$$\lim_{n\to\infty}\left(\frac{1}{S_n^2}\sum_{i=1}^n E\left[X_i^2 I_{\{|X_i|>\varepsilon S_n\}}\right]\right) = 0$$

If N is not simple, from the made assumptions, then:

$$\sum_{i=1}^n E\left[X_i^2 I_{\{|X_i|>\varepsilon S_n\}}\right]$$

from the Hölder inequality

$$\leq \sum_{i=1}^n \left(E\left[|X_i|^{2+\alpha}\right]\right)^{2/(\alpha+2)} P(|X_i| > \varepsilon S_n)^{\alpha/(\alpha+2)}$$

$$\leq \left(\sum_{i=1}^n E\left[|X_i|^{2+\alpha}\right]^{2/(\alpha+2)}\right)\left(\sum_{i=1}^n P(|X_i| > \varepsilon S_n)\right)^{\alpha/(\alpha+2)}.$$

Since from the Chebyshev inequality:

$$\sum_i^n P(|X_i| \geq \varepsilon S_n) \leq \frac{E(X_i)^2}{\varepsilon^2 S_n^2} = \frac{1}{\varepsilon^2}$$

and $E[N^{2+\alpha}(\mathrm{d}t_i)|\mathscr{H}_i]$ is bounded for all i, the Lindeberg condition holds.

Therefore, we have shown that:

$$\frac{1}{S_n}Z \to N(0, 1) \tag{16}$$

with $Z = \sum_{i=1}^n X_i$.

In order to prove the asymptotic normality for the weighted correlation integral, the process $Y(t_i)$ is introduced, counting the number of pairs up to time t_i. To prove the same convergence for $Y(t_i)$ as in (16), the following condition is proved:

$$\frac{1}{S_n}|Y - Z| \xrightarrow{P} 0, \tag{17}$$

where the differences $|Y(t_i) - Z(t_i)|$, for each i, consist of all the points occurring between τ_i and t_i, where τ_i is the left-side point of the last gap of size not less than δ before t_i. Since $E[|Y - Z|] = 0$, from (15) then $E\left[(Y(t_i) - Z(t_i))^2\right] < \infty$ (easily proved replacing τ_{i-1} with τ_i and τ_i with t_i, respectively); from the made assumptions and the Chebyshev inequality:

$$P(|Y - Z| \geq \varepsilon) \leq \frac{\text{var}[Y - Z]}{\varepsilon^2} = \frac{E\left[(Y - Z)^2\right]}{\varepsilon^2},$$

convergence in (17) is proved.

If a two-dimensional space is considered, to define a filtration $\mathscr{F}$ on A, some ordering is necessary. For $s' = (x', y')$ and $s'' = (x'', y'') \in \mathbb{R}_+^2$, say $s' \leq s''$ if the Euclidean distance from s' to the origin (s_0) is less than the Euclidean distance from s'' to the origin, i.e:

$$s' \leq s'' \iff |\overline{s's_0}| \leq |\overline{s''s_0}|.$$

For the sake of simplicity, it is assumed that the point process N vanishes on the axes. Define a filtration $\mathscr{F}(s)$ on the complete probability space $(\Omega, \mathscr{A}, P)$, as the increasing family of sub-σ-algebras of $\mathscr{A}$ such that if $s' \leq s''$ then $\mathscr{F}(s') \subseteq \mathscr{F}(s'')$. Let D be the mapping from Ω to the closed subsets of $\mathbb{R}_+^2$ such that for $s' < s''$, $s'' \in D(\omega)$ implies $s' \in D(\omega)$ and $\{\omega : D(\omega) \leq s\} \in \mathscr{F}(s)$, $\forall s \in \mathbb{R}_+^2$.

As in Merzbach and Nualart (1986), let ζ be a random curve (for instance an arc of circle, the sides of a square of length depending on the maximum coordinate of s) such that for each pair $s' \leq s''$ then $s' \notin \zeta(\omega)$ or $s'' \notin \zeta(\omega)$ and $\{\omega : s \in \zeta(\omega)\} \in \mathscr{F}(s)$. $s' \leq \zeta$ implies the existence of a random point s'' such that $s' \leq s''$ and $s'' \in \zeta$. The stopping line C determines the stopping set $D(\zeta) = \{(\omega, s) : s \in \zeta(\omega)\}$, and conversely. Therefore, D_s is the random closed subset of $A \subseteq \mathbb{R}^2$ bounded by the axes and ζ_s, induced by all the points that are less than s, with measure $\ell(A)_s$.

Let $I_{pq}^w(\delta)|_A = \frac{1_{\{0 < |p-q| < \delta, A\}}}{\lambda(p)\lambda(q)}$ with $1_{\{0 < |p-q| < \delta, A\}}$ the indicator function that is 1 if the distance (in terms of the defined order on A) between the two points of the pair $\{s_p, s_q\}$, such that $p \neq q$ and $\{s_p, s_q\}_{p \neq q} \in A$, is both strictly greater than zero and less than δ, and zero otherwise. Moreover, let $I_{pq}^w(\delta)|_{A_h}$ be the same as $I_{pq}^w(\delta)|_A$ conditioning on the points that occur in the space area A_h. From the defined ordering, and assuming N is simple, it is possible to move from the two-dimensional space $\mathbb{R}_+^2$ to $\mathbb{R}$, and therefore, to extend the results provided for temporal processes to the spatial case.

5 Conclusion

The proposed method for residuals analysis in point processes is an useful tool for the comprehension of point patterns properties, when attraction or dependence features are present. The method, in comparison with others, has the advantage to analyze features of data without requiring a random depletion of data or a rescaling along a fixed domain. As a consequence it eliminates the sampling variability due to a random thinning and

extends the field of application to point processes of any dimension. Moreover, the method incorporates the use of the second-order statistics of point processes, allowing an immediate interpretation of attractive or repulsive characteristics of observed point patterns.

The diagnostic tool here introduced is the result of an attempt to generalize the method of defining a residual measure for point processes (Baddeley et al. 2000) as described in Veen and Schoenberg (2005). While we have focused on second-order characteristics for purely temporal or planar point processes, it must be emphasized that these techniques should be applicable for spatial-temporal point processes, processes in higher dimensions, and processes in more general metric spaces. In particular, further attention should focus on space–time metrics and statistics useful for describing attractive or repulsive features in space–time point processes. In addition, weighted versions could be developed for other statistics aside from those considered here, including for instance high-order moments and high-order point-centered Renyi dimensions.

Moreover, although asymptotic distributional results are provided in this paper, we did not focus on issues related to rates of convergence, which are important in real applications, especially those where the sample size is small, and therefore would be an important topic for future work.

Another important direction for future work would be a comparison between the various residual second-order statistics proposed here as well as comparison with other methods of residual analysis such as the first-order residuals of Baddeley et al. (2000) or the second-order statistics of the residual processes such as those in Schoenberg (2003).

References

Albrecht, P. (1982). Testing the goodness-of-fit of a mixed Poisson process. *Insurance: Mathematics and Economics*, *1*(1), 27–33.

Arsham, H. (1987). A modified one-sided *k-s* confidence region narrower in one tail. *Communication in Statistics. Theory and Methods*, *16*, 17–28.

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

Bartlett, M. S. (1964). The spectral analysis of two dimensional point processes. *Biometrika*, *51*(3/4), 299–311.

Billingsley, P. (1968). *Convergence of probability measures*. New York: Wiley.

Clegg, R. (2006). A practical guide to measuring the hurst parameter. *International Journal of Simulation: Systems, Science and Technology*, *7*(2), 3–14.

Cressie, N. (1991). *Statistics for spatial data*. New York: Wiley series in probability and mathematical statistics.

Daley, D. J. (1999). The hurst index of long-range dependent renewal processes. *Annals of Probability*, *27*(4), 2035–2041.

Daley, D. J., Vere-Jones, D. (2003). *An introduction to the theory of point processes*. (2nd ed). New York: Springer.

Davies, R. (1977). Testing the hypothesis that a point process is Poisson. *Advances in Applied Probability*, *9*, 724–746.

Denker, M., Keller, G. (1986). Rigorous statistical procedures for data from dynamical systems. *Journal of Statistical Physics*, *44*, 67–94.

Diggle, P. J. (1983). *Statistical analysis of spatial point patterns*. London: Academic Press.

Grassberger, P., Procaccia, I. (1983). Measuring the strangeness of a strange attractor. *Physica D*, *9*, 189–208.

Hall, P., Heyde, C. C. (1980). *Martingale limit theory and its application*. New York: Academic Press.

Harte, D. S. (1998). Dimension estimates of earthquake epicentres and hypocentres. *Journal of Nonlinear Science*, 8, 581–618.

Harte, D. S. (2001). *Multifractals: theory and aplications*. Boca Raton: Chapman & Hall/CRC.

Heinrich, L. (1991). Goodness-of-fit tests for the second moment function of a stationary multidimensional Poisson process. *Statistics*, 22, 245–278.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163–1174.

Hurst, H. E. (1951). Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116, 770–808.

Lawson, A. B. (1988). On tests for spatial trend in a non-homogeneous Poisson process. *Journal of Applied Statistics*, 15(2), 225–234.

Lee, A. J. (1990). *U-statistics: theory and practice*. New York: Marcel Dekker.

Liebetrau, A. M. (1978). The weak convergence of a class of estimators of the variance function of a two-dimensional Poisson process. *Journal of Applied Probability*, 15(2), 433–439.

Lisek, B., Lisek, M. (1985). A new method for testing whether a point process is Poisson. *Statistics*, 16, 445–450.

Lo, A. (1991). Long-term memory in stock market prices. *Econometrica*, 59(5), 1279–1313.

Mandelbrot, B. B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift ftr. Wahrscheinlichkeitstheorie. und Verwandte Gebiete*, 31, 271–285.

Mandelbrot, B. B. (1977). *Fractals; form chance and dimension*. San Francisco: W. H. Freeman.

Mandelbrot, B. B., Wallis, J. R. (1969). Robustness of the rescaled range r/s in the measurement in non-cyclic long run statistical dependence. *Water Resources Research*, 5, 967–988.

Merzbach, E., Nualart, D. (1986). A characterization of the spatial Poisson process and changing time. *Annals of Probabability*, 114, 1380–1390.

Meyer, P. (1971). Démonstration simplifée d'un théorème de knight. *Lecture Notes in Mathematics*, 191, 191–195.

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

Ogata, Y., Abe, K. (1991). Some statistical features of the long-term variation of the global and regional seismic activity. *International Statistical Review*, 59(2), 139–161.

Ripley, B. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, Series B*, 39, 172–212.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13(2), 255–266.

Saw, J. (1975). Tests on the intensity of a Poisson process. *Communication in Statistics*, 4, 777–782.

Schoenberg, F. P. (1999). Transforming spatial point processes into Poisson processes. *Stochastic Processes and their Applications*, 81(2), 155–164.

Schoenberg, F. P. (2003). Multi-dimensional residual analysis of point process models for earthquake occurrences. *Journal American Statistical Association*, 98(464), 789–795.

Stoyan, D., Stoyan, H. (2000). Improving ratio estimators of second-order point process characteristics. *Scandinavian Journal of Statistics*, 27, 641–656.

Sundt, B. (1982). On the problem of testing whether a mixed Poisson process is homogeneous. *Insurance: Mathematics and Economics*, 1(4), 253–254.

Veen, A. (2006). *Some methods of assessing and estimating point processes models for earthquake occurrences*. Ph.D. thesis, UCLA.

Veen, A., Schoenberg, F. P. (2005). Assessing spatial point process models using weighted *k*-functions: Analysis of california earthquakes. *UC's eScholarship Repository*.

Whittle, P. (1962). Gaussian estimation in stationary time series. *Bulletin of the International Statistical Institute*, 39, 105–129.

Zhuang, J. (2006). Second-order residual anaysis of spatio-timporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society, Series B*, 68(4), 635–653.