Multiresolution analysis of point processes and statistical thresholding for Haar wavelet-based intensity estimation:

Supplementary Material

Youssef Taleb and Edward A. K. Cohen*

Department of Mathematics, Imperial College London, London, UK

S1 Further Discussion on Likelihood Ratio Tests

S1.1 Null parameters on the boundary of the parameter space

In Theorems 1 and 2, we establish in particular the asymptotic distribution of the modified likelihood ratio statistic R under J-th level homogeneity and L-th level innovation, respectively. These results require that the true but unknown values of the parameters λ_c^J and μ_k^{pair} under the null hypothesis of each LRT are in the interior of the global parameter space. The boundaries of the parameter spaces when testing J-th level homogeneity and L-th level innovation contain the parameter vectors satisfying $\lambda_k^J = 0$ and $\mu_k = 0$ for one or more dyadic translation indexes $k \in 0, ..., 2^J - 1$ and $k \in 0, ..., 2^{L+1} - 1$, respectively. If $\lambda_c^J = 0$ under J-th level homogeneity for any resolution level J, then we are in the trivial case where the intensity is the zero function on $[0, T)$. This would lead us to never observe any event almost surely, thus we can exclude the value $\lambda_c^J = 0$ from the null parameter space of our likelihood ratio model. From any other data, Theorem 1 can be used provided its other conditions are met.

*Correspondence should be addressed to E.A.K.C. (e.cohen@imperial.ac.uk).

However, point processes with non-zero intensities may not possess L-th level innovation for some level L but still have parameters for that test on the boundary. For example, if λ is non-zero and constant on $[0, T/2)$ and zero otherwise, then there is no level 1 innovation (in the Haar sense) but $\mu_2^{pair} = 0$. Therefore, we need to detail a further analysis in order to propose decision rules when M is large and some MLEs of the parameters take value zero. This analysis will be done under the general setting of Section 3.4.1.

Under the null hypothesis of the model leading to Theorem 2, we denote U the number of true parameters μ_i^{pair} equal to zero. By definition, we have $0 \leq U < P$, the case $U = P$ being excluded since under this condition no data points would be observed. From some data $\mathbb{X}$, we also denote $\bar{U}$ the number of pairs of MLEs $(\bar{\mu}_{2i-1}, \bar{\mu}_{2i})$ that are equal to $(0, 0)$, which is equivalent to $\bar{\mu}_i^{pair} = 0$. If $\bar{U} = 0$, then $U = 0$ since $U \leq \bar{U}$ and therefore the true parameter vector does not lie on the boundary of the parameter space, which allows us to apply Wilks' theorem provided its other conditions are met. Now consider the case where $\bar{U} > 0$ and thus $U > 0$ is a possibility. Let us place ourselves under the null hypothesis H, which we recall states $\mu_{2i-1} = \mu_{2i} = \mu_i^{pair},\ 1 \leq i \leq P$. From the proof in Appendix S2.5, statistic R would have the same value if the data $\mathbb{X}$ was instead obtained from a multivariate Poisson random variable of dimension $2P - 2U$ where we exclude the U pairs of components that have zero mean. Therefore, under the conditions of Theorem 2, R is asymptotically χ^2_{P-U} distributed.

Since the value of U is hidden and $U \leq \bar{U}$, we have $\bar{U} + 1$ possible distributions for R under the null and hence $\bar{U} + 1$ possible critical values. Each critical value is denoted $z_{P-u,\alpha},\ 0 \leq u \leq \bar{U}$, and is the upper $100(1-\alpha)\%$ point of the χ^2_{P-u} distribution. We propose three choices of critical values which yield different type 1 error bounds for the LRT. The asymptotic type 1 error of the LRT is denoted ϵ_1 and the cumulative distribution function of the chi-squared distribution with d degrees of freedom is denoted $\mathcal{F}_d$.

1. The first choice is to use the critical value $z_{P,\alpha}$, which is equivalent to assuming $U = 0$, i.e. all mean parameters μ_i^{pair} are non-zero. This places us in the most conservative setting since $z_{P,\alpha} = \max\left\{z_{P-u,\alpha},\ 0 \leq u \leq \bar{U}\right\}$. The type 1 error of the LRT in this case satisfies $1 - \mathcal{F}_{P-\bar{U}}(z_{P,\alpha}) \leq \epsilon_1 \leq \alpha$. This reduction in type 1 error is accompanied by a loss of power.

2. The second choice is to use the critical value $z_{P-\bar{U},\alpha}$, which is equivalent to assuming $U = \bar{U}$, i.e $\mu_i^{pair} = 0$ if and only if $\bar{\mu}_i^{pair} = 0$. Therefore, this is the maximum likelihood decision. It leads to a gain of

power when $U < \bar{U}$ since $z_{P-\bar{U},\alpha} = \min\left\{z_{P-u,\alpha},\ 0 \leq u \leq \bar{U}\right\}$. However, the type 1 error of the LRT in this case now satisfies $\alpha \leq \epsilon_1 \leq 1 - \mathcal{F}_P(z_{P-\bar{U},\alpha})$.

3. The third choice of critical value is motivated by an attempt to strike a balance between $z_{P,\alpha}$ and $z_{P-\bar{U},\alpha}$. We propose the intermediate value $z_{P-\left\lceil\frac{\bar{U}}{2}\right\rceil,\alpha}$. This provides a scheme for balancing the type 1 error/power trade-off. The type 1 error now satisfies the inequality

$$1 - \mathcal{F}_{P-\bar{U}}(z_{P-\left\lceil\frac{\bar{U}}{2}\right\rceil,\alpha}) \leq \epsilon_1 \leq 1 - \mathcal{F}_P(z_{P-\left\lceil\frac{\bar{U}}{2}\right\rceil,\alpha}).$$

Figure S1 gives a potential criterion on how to choose $z_{P-u,\alpha}$. If both $\bar{U}/2^J$ is close to 0 and $\bar{U}$ is low (top-left plot), then the choice of critical value does not lead to considerably significant differences in the behaviour of the hypothesis test and hence this choice mostly depends on a preference for a slightly more conservative or power favoring setting. If alternatively $\bar{U}$ increases or $\bar{U}/2^J$ is closer to 1, then the two extreme choices $z_{P,\alpha}$ and $z_{P-\bar{U},\alpha}$ lead to an effective type 1 error that will be significantly different than α when this choice is erroneous, and could yield a test that is almost always accepting or rejecting the null. In this situation, the intermediate value $z_{P-\left\lceil\frac{\bar{U}}{2}\right\rceil,\alpha}$ gives more reasonable error boundaries.

S1.2 Maximizing the J-th level homogeneity test statistic

For a point process that is not level J homogeneous, and for large M, we can still encounter situations where one or several MLEs $\bar{\lambda}_k^J$ are equal to zero, whether the corresponding true parameters λ_k^J are zero or not. Considering the general setting of the LRT in Section 3.2.1, we derive the situation under which R, the test statistic defined in Proposition 3, is maximized.

Proposition S1.1 *Let $c > 0$, $P \geq 1$ and $f^P : \mathbb{R}_+^P \to \mathbb{R}$ given by $f^P(x) = \sum_{i=1}^{P} x_i \log(\frac{x_i}{c})$, where we assume the continuous extension of the function $x \log x$ from $\mathbb{R}_+^*$ to $\mathbb{R}_+$, i.e. $f^P(0) = 0$. Let $\Omega_{c,P}$ be the subset of $[0, Pc]^P$ defined as*

$$\Omega_{c,P} = \left\{(x_1, ..., x_P) \in [0, Pc]^P,\ \frac{1}{P}\sum_{i=1}^{P} x_i = c\right\}.$$

Then the restriction of f^P on $\Omega_{c,P}$ attains its maximum for any element in $\Omega_{c,P}$ of the form $(0, ..., x_i = Pc, ..., 0), 1 \leq i \leq P$.

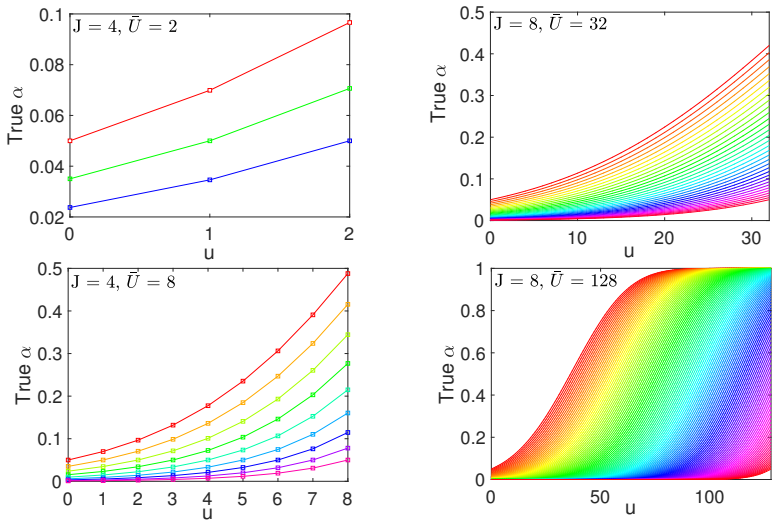


Figure S1: *True type 1 error depending on the choice of critical value* $z_{P-u,\alpha}$, $0 \leq u \leq \bar{U}$. *Each line represents the situation when there are truly* U *parameters* μ_i^{pair} *equal to zero, with* $0 \leq U \leq \bar{U}$.

See proof in Appendix S2.8. In our setting, Proposition S1.1 has an interesting interpretation. If we impose that the MLE $\bar{\mu}_c$ takes some value $c > 0$, then statistic R is maximized by the data $\mathbb{X}$ that produces one MLE $\bar{\mu}_i$ with value Pc and all other MLEs with value zero. Proposition S1.1 illustrates a scenario of maximum inhomogeneity in the likelihood ratio sense, which is characterized by a maximum distance between the lowest and highest values among the MLEs $\bar{\mu}_i, 1 \leq i \leq P$. A similar result can be formulated for L-th level innovation.

S2 Proofs

S2.1 Proof of Remark 1

W.l.o.g. we prove this result with $T = 1$. Since $\alpha_{J,k} = \langle \lambda, \phi_{J,k} \rangle$, we have:

$$\begin{aligned}
\alpha_{J,k} = \int_0^1 \lambda(t)\phi_{J,k}(t)dt &= \int_{k/2^J}^{(k+1)/2^J} 2^{J/2}\lambda(t)\mathrm{d}t \\
&= \tfrac{1}{\sqrt{2}}\left[\int_{2k/2^{J+1}}^{(2k+1)/2^{J+1}} 2^{(J+1)/2}\lambda(t)\mathrm{d}t + \int_{(2k+1)/2^{J+1}}^{(2k+2)/2^{J+1}} 2^{(J+1)/2}\lambda(t)\mathrm{d}t\right] \\
&= \frac{1}{\sqrt{2}}(\alpha_{J+1,2k} + \alpha_{J+1,2k+1}).
\end{aligned}$$

S2.2 Proof of Proposition 1

From the multiresolution setting defined in Section 2.2.2, we know that $\widehat{\lambda}^J(t) = \widehat{\lambda}_k^J \cdot \mathbb{1}_{s_k^J}(t)$ where $\widehat{\lambda}_k^J$ is the value of the J-th level wavelet reconstruction estimator on the subinterval $s_k^J \in S_J$. Using $\widehat{\alpha}_{J,k} = \sum_{\tau_i} \phi_{J,k}(\tau_i)$, for Haar wavelets we have $\widehat{\alpha}_{J,k} = \frac{2^{J/2}}{\sqrt{T}} x_k^J$, where x_k^J is the event count in the corresponding subinterval $s_k^J \in S_J$. Therefore $\widehat{\lambda}_k^J = \frac{2^J}{T} x_k^J$. Since N is a Poisson process and Haar wavelets have disjoint supports across all translations for a fixed scale J, we have:

1. each event count x_k^J is Poisson distributed with mean $\mu_k^J = \int_{s_k^J} \lambda(t)\mathrm{d}t$,
2. all event counts x_k^J, $0 \le k \le 2^J - 1$, are independent.

Therefore the intensity estimators $\widehat{\lambda}_0^J, ..., \widehat{\lambda}_{2^J-1}^J$ are independent random variables distributed as

$$\widehat{\lambda}_k^J \sim \frac{2^J}{T}\mathrm{Pois}(\mu_k^J).$$

S2.3 Proof of Proposition 2

Left to right: This direction is proved using the definition of Haar wavelet coefficients. We know there exists $\lambda_0 \ge 0$ such that $\lambda(t) = \lambda_0$ a.e. (in the Lebesgue sense) on $[0, T)$. Let $J \ge 0$ and consider the subdivision

$S_J = \{s_k\}_{k=0}^{2^J-1}$ defined in Section 2.2.2. For all $0 \leq k \leq 2^J - 1$ we have:

$$\alpha_{J,k} = \int_0^T \lambda(t)\phi_{J,k}(t)\mathrm{d}t = \int_{s_k} \lambda(t)\phi_{J,k}(t)\mathrm{d}t = \int_{s_k} \lambda_0\phi_{J,k}(t)\mathrm{d}t = \frac{\sqrt{T}}{2^{J/2}}\lambda_0.$$

Hence for all $0 \leq k \leq 2^J - 1$ and $0 \leq k' \leq 2^J - 1$ we obtain $\alpha_{J,k} = \alpha_{J,k'}$. This is equivalent to $\lambda_k^J = \lambda_{k'}^J$ and $N \in H_J$.

Right to left: Since λ is locally square integrable, it is also locally integrable. Let $F : [0,T] \to \mathbb{R}_+$ given by $F(x) = \int_0^x \lambda(t)\mathrm{d}t$. Let $x, y \in [0,T)$ such that $0 < x < y$, and for each resolution J let $k_{x,J}, k_{y,J} \in \{0, ..., 2^J - 1\}$ such that $[k_{x,J}T/2^J, k_{y,J}T/2^J)$ be the smallest interval that includes $[x,y)$. Since $N \in H_J$ for all $J \geq 0$, we have

$$F(\frac{k_{x,J} + k_{y,J}}{2^{J+1}}T) = \frac{1}{2}\left(F(k_{x,J}T/2^J) + F(k_{y,J}T/2^J)\right). \tag{1}$$

The Lebesgue integral F is continuous and the interval $[k_{x,J}T/2^J, k_{y,J}T/2^J)$ shrinks to $[x,y)$ when $J \to \infty$ thus (1) becomes $F(\frac{x+y}{2}) = 1/2(F(x)+F(y))$ when $J \to \infty$. From the Lebesgue differentiation theorem F is differentiable almost everywhere on $[0,T)$, and for all $y \in [0,T)$ we have

$$\frac{\partial}{\partial x}F(\frac{x+y}{2}) = \frac{1}{2}F'(\frac{x+y}{2}) = \frac{1}{2}F'(x).$$

Taking $y = T - x$ gives $F'(T/2) = F'(x)$ hence $F' = \lambda$ is constant almost everywhere on $[0,T)$.

S2.4 Proof of Proposition 3

Let $\mathbb{X} = \{\mathbf{X}_m\}_{m=1}^M$ be a set of iid scaled Poisson random vectors, each with independent components of form $\mathbf{X}_m = (X_{m,i})_{i=1}^P$, $X_{m,i} \sim \delta\mathrm{Pois}(\mu_i)$. Therefore, for any non-negative integer $k_{m,i}$ we have $P(X_{m,i} = \delta k_{m,i}) = \exp(-\mu_i)\frac{\mu_i^{k_{m,i}}}{k_{m,i}!}$. The likelihood functions of $\mathbb{X}$ under the null and alternative hypotheses H and K are

$$\mathcal{L}_H(\mathbb{X}; \mu_c, ..., \mu_c) = \prod_{m=1}^M \prod_{i=1}^P \exp(-\mu_c)\frac{\mu_c^{k_{m,i}}}{k_{m,i}!} = \exp(-MP\mu_c)\prod_{i=1}^P \frac{\mu_c^{\sum_{m=1}^M k_{m,i}}}{\prod_{m=1}^M k_{m,i}!}$$

$$\text{and } \mathcal{L}_K(\mathbb{X}; \mu_1, ..., \mu_P) = \prod_{m=1}^M \prod_{i=1}^P \exp(-\mu_i)\frac{\mu_i^{k_{m,i}}}{k_{m,i}!} = \exp(-M\sum_{i=1}^P \mu_i)\prod_{i=1}^P \frac{\mu_i^{\sum_{m=1}^M k_{m,i}}}{\prod_{m=1}^M k_{m,i}!}.$$

To locate their maxima we consider the log-likelihood functions

$$\log \mathcal{L}_H(\mathbb{X}; \mu_c, ..., \mu_c) = -MP\mu_c + \sum_{i=1}^{P} \left[\log(\mu_c) \sum_{m=1}^{M} k_{m,i} - \sum_{m=1}^{M} \log(k_{m,i}!) \right]$$

$$\text{and } \log \mathcal{L}_K(\mathbb{X}; \mu_1, ..., \mu_P) = -M \sum_{i=1}^{P} \mu_i + \sum_{i=1}^{P} \left[\log(\mu_i) \sum_{m=1}^{M} k_{m,i} - \sum_{m=1}^{M} \log(k_{m,i}!) \right].$$

Differentiating each function with respect to its parameters gives:

$$\begin{cases} \dfrac{d \log \mathcal{L}_H}{d\mu_c} & = -MP + \frac{1}{\mu_c} \sum\limits_{i=1}^{P} \sum\limits_{m=1}^{M} k_{m,i} \\ \dfrac{\partial \log \mathcal{L}_K}{\partial \mu_i} & = -M + \frac{1}{\mu_i} \sum\limits_{m=1}^{M} k_{m,i}, \ \forall\ 1 \leq i \leq P. \end{cases}$$

Therefore, with $k_{m,i} = X_{m,i}/\delta$, the maximum values of $\mathcal{L}_H$ and $\mathcal{L}_K$ are respectively attained at $\bar{\mu}_c = \frac{1}{\delta MP} \sum\limits_{i=1}^{P} \sum\limits_{m=1}^{M} X_{m,i}$ and $\bar{\mu}_i = \frac{1}{\delta M} \sum\limits_{m=1}^{M} X_{m,i}$ for all $1 \leq i \leq P$. Statistic $\bar{\mu}_c$ is the MLE of μ_c, the constant intensity under the null hypothesis H, and $\bar{\mu}_i$ is the MLE for μ_i $(i = 1, ..., P)$ under the alternative hypothesis K. Since the likelihood ratio statistic r is

$$r = \frac{\sup\limits_{\mu_c > 0} \mathcal{L}(\mathbb{X}; \mu_c, ..., \mu_c)}{\sup\limits_{\{\mu_i\}_{i=1}^{P}, \sum \mu_i > 0} \mathcal{L}(\mathbb{X}; \mu_1, ..., \mu_P)},$$

applying the previous results yields

$$r = \exp\left(-M \left(P\bar{\mu}_c - \sum_{i=1}^{P} \bar{\mu}_i \right) \right) \prod_{i=1}^{P} \left(\frac{\bar{\mu}_c}{\bar{\mu}_i} \right)^{M\bar{\mu}_i} = \prod_{i=1}^{P} \left(\frac{\bar{\mu}_c}{\bar{\mu}_i} \right)^{M\bar{\mu}_i}.$$

We can now derive the test statistic R:

$$R = -2 \log(r) = 2M \sum_{i=1}^{P} \bar{\mu}_i \log\left(\frac{\bar{\mu}_i}{\bar{\mu}_c} \right).$$

S2.5 Proof of Proposition 4

Let $\mathbb{X} = \{\mathbf{X}_m\}_{m=1}^{M}$ be a set of iid Poisson random vectors, each with independent components of form $\mathbf{X}_m = (X_{m,i})_{i=1}^{2P}$, $X_{m,i} \sim \text{Pois}(\mu_i)$. Therefore, for any non-negative integer $k_{m,i}$ we have $P(X_{m,i} = k_{m,i}) = \exp(-\mu_i) \frac{\mu_i^{k_{m,i}}}{k_{m,i}!}$.

The likelihood functions of $\mathbb{X}$ under the null and alternative hypotheses H and K are

$$\begin{aligned}
\mathcal{L}_H(\mathbb{X}; \mu_1^{pair}, \mu_1^{pair}, ..., \mu_P^{pair}, \mu_P^{pair}) &= \prod_{m=1}^{M} \prod_{i=1}^{P} \exp(-2\mu_i^{pair}) \frac{(\mu_i^{pair})^{k_{m,2i-1}+k_{m,2i}}}{k_{m,2i-1}! k_{m,2i}!} \\
&= \exp(-2M \sum_{i=1}^{P} \mu_i^{pair}) \prod_{i=1}^{P} \frac{(\mu_i^{pair})^{\sum_{m=1}^{M} k_{m,2i-1}+k_{m,2i}}}{\prod_{m=1}^{M} k_{m,2i-1}! k_{m,2i}!} \\
\text{and } \mathcal{L}_K(\mathbb{X}; \mu_1, ..., \mu_{2P}) &= \prod_{m=1}^{M} \prod_{i=1}^{2P} \exp(-\mu_i) \frac{\mu_i^{k_{m,i}}}{k_{m,i}!} \\
&= \exp(-M \sum_{i=1}^{2P} \mu_i) \prod_{i=1}^{2P} \frac{\mu_i^{\sum_{m=1}^{M} k_{m,i}}}{\prod_{m=1}^{M} k_{m,i}!}.
\end{aligned}$$

Then similarly as in S2.4, the likelihood function $\mathcal{L}_H$ is maximized when each parameter μ_i^{pair} is equal to $\bar{\mu}_i^{pair} = \frac{1}{2M} \sum_{m=1}^{M} k_{m,2i-1} + k_{m,2i}$, and the likelihood function $\mathcal{L}_K$ is maximized when each parameter μ_i is equal to $\bar{\mu}_i = \frac{1}{M} \sum_{m=1}^{M} k_{m,i}$. We also immediately have $\bar{\mu}_i^{pair} = \frac{1}{2}(\bar{\mu}_{2i-1} + \bar{\mu}_{2i})$. Since the likelihood ratio statistic r is

$$r = \frac{\sup_{\{\mu_i^{pair}\}_{i=1}^{P},\ \sum \mu_i^{pair} > 0} \mathcal{L}(\mathbb{X}; \mu_1^{pair}, ..., \mu_P^{pair})}{\sup_{\{\mu_i\}_{i=1}^{2P},\ \sum \mu_i > 0} \mathcal{L}(\mathbb{X}; \mu_1, ..., \mu_{2P})},$$

applying the previous results yields

$$\begin{aligned}
r &= \exp\left(-M\left(2\sum_{i=1}^{P} \bar{\mu}_i^{pair} - \sum_{i=1}^{2P} \bar{\mu}_i\right)\right) \prod_{i=1}^{P} \frac{\left(\bar{\mu}_i^{pair}\right)^{2M\bar{\mu}_i^{pair}}}{\left(\bar{\mu}_{2i-1}\right)^{M\bar{\mu}_{2i-1}} \left(\bar{\mu}_{2i}\right)^{M\bar{\mu}_{2i}}} \\
&= \prod_{i=1}^{P} \frac{\left(\bar{\mu}_i^{pair}\right)^{2M\bar{\mu}_i^{pair}}}{\left(\bar{\mu}_{2i-1}\right)^{M\bar{\mu}_{2i-1}} \left(\bar{\mu}_{2i}\right)^{M\bar{\mu}_{2i}}} \\
&= \prod_{i=1}^{P} \frac{\left(\bar{\mu}_i^{pair}\right)^{M\bar{\mu}_{2i-1}} \left(\bar{\mu}_i^{pair}\right)^{M\bar{\mu}_{2i}}}{\left(\bar{\mu}_{2i-1}\right)^{M\bar{\mu}_{2i-1}} \left(\bar{\mu}_{2i}\right)^{M\bar{\mu}_{2i}}}.
\end{aligned}$$

We can now derive the test statistic R:

$$R = -2\log(r) = 2M\left[\sum_{i=1}^{P}\bar{\mu}_{2i-1}\log\left(\frac{\bar{\mu}_{2i-1}}{\bar{\mu}_i^{pair}}\right) + \sum_{i=1}^{P}\bar{\mu}_{2i}\log\left(\frac{\bar{\mu}_{2i}}{\bar{\mu}_i^{pair}}\right)\right].$$

S2.6 Proof of Theorem 1

The expression of R given in Proposition 3 can be rewritten as

$$R = 2\sum_{i=1}^{P}\left[\sum_{m=1}^{M}\frac{X_{m,i}}{\delta}\log\left(\frac{P\sum_{m=1}^{M}X_{m,i}}{\sum_{j=1}^{P}\sum_{m=1}^{M}X_{m,j}}\right)\right]$$

when replacing the MLEs by their actual value. Using the notation $Y_i^M = \sum_{m=1}^{M}X_{m,i}/\delta$, this becomes

$$R = 2\sum_{i=1}^{P}\left[Y_i^M\log\left(\frac{PY_i^M}{\sum_{j=1}^{P}Y_j^M}\right)\right].$$

Given Y_i^M is Poisson distributed with mean $\mu_c M$ under the null hypothesis H ($\mu_i M$ under the alternative hypothesis K), the distribution of R depends only on the product $\mu_c M$ (or $\mu_i M$). Therefore, the standard asymptotic results for R hold as $\mu_c M \to \infty$. This limit can be achieved either through $M \to \infty$, $\mu_c \to \infty$, or both. The null distribution of R is asymptotically χ^2 with $P-1$ degrees of freedom for a large $\mu_c M$. We thus reject H at significance level α if $R > c_\alpha$ where c_α, the critical value, is the upper $100(1-\alpha)\%$ point of the χ^2_{P-1} distribution.

S2.7 Proof of Theorem 2

Similarly as in the proof for Theorem 1, we go back to the expression of R given in Proposition 4:

$$R = 2M\left[\sum_{i=1}^{P}\bar{\mu}_{2i-1}\log\left(\frac{\bar{\mu}_{2i-1}}{\bar{\mu}_i^{pair}}\right) + \sum_{i=1}^{P}\bar{\mu}_{2i}\log\left(\frac{\bar{\mu}_{2i}}{\bar{\mu}_i^{pair}}\right)\right].$$

This can also be written as

$$R = 2\left[\sum_{i=1}^{P}\sum_{m=1}^{M}X_{m,2i-1}\log\left(\frac{2\sum_{m=1}^{M}X_{m,2i-1}}{\sum_{m=1}^{M}X_{m,2i-1}+X_{m,2i}}\right) + \sum_{i=1}^{P}\sum_{m=1}^{M}X_{m,2i}\log\left(\frac{2\sum_{m=1}^{M}X_{m,2i}}{\sum_{m=1}^{M}X_{m,2i-1}+X_{m,2i}}\right)\right]$$

when replacing the MLEs by their actual value. Using the notation $Y_i^M = \sum_{m=1}^M X_{m,i}$, this becomes

$$R = 2\left[\sum_{i=1}^{P} Y_{2i-1}^M \log\left(\frac{2Y_{2i-1}^M}{Y_{2i-1}^M + Y_{2i}^M}\right) + \sum_{i=1}^{P} Y_{2i}^M \log\left(\frac{2Y_{2i}^M}{Y_{2i-1}^M + Y_{2i}^M}\right)\right].$$

Given Y_i^M is Poisson distributed with mean $\mu_i^{pair} M$ under the null hypothesis H ($\mu_i M$ under the alternative hypothesis K), the distribution of R depends only on the product $\mu_i^{pair} M$ (or $\mu_i M$). Therefore, the standard asymptotic results for R hold as $\mu_i^{pair} M \to \infty$, for all $1 \leq i \leq P$. This limit can be achieved either through $M \to \infty$, $\mu_i^{pair} \to \infty$ for all $1 \leq i \leq P$, or both. The null distribution of R is asymptotically χ^2 with P degrees of freedom for all $\mu_i^{pair} M$ large. We thus reject H at significance level α if $R > c_\alpha$ where c_α, the critical value, is the upper $100(1-\alpha)\%$ point of the χ_P^2 distribution.

S2.8 Proof of Proposition S1.1

We prove this by mathematical induction on the number of parameters P. The result is obvious at $P = 1$ since $\Omega_{c,1}$ becomes the singleton $\{c\}$. We will therefore detail the case $P = 2$.

Base case $P = 2$:

We have $\Omega_{c,2} = \big\{(x_1, x_2) \in [0, 2c]^2, x_1 + x_2 = 2c\big\}$. This lets us write

$$f^2(x_1, x_2) = x_1 \log(\frac{x_1}{c}) + x_2 \log(\frac{x_2}{c}) = x_1 \log(\frac{x_1}{c}) + (2c - x_1) \log(\frac{2c - x_1}{c}).$$

Denoting $g : x_1 \mapsto x_1 \log(\frac{x_1}{c}) + (2c - x_1) \log(\frac{2c-x_1}{c})$, then g is differentiable with respect to x_1 on $(0, 2c)$ and for all $x_1 \in (0, 2c)$ we have:

$$g'(x_1) = \log(\frac{x_1}{c}) + 1 - \log(\frac{2c - x_1}{c}) - 1 = \log(\frac{x_1}{2c - x_1}).$$

Immediately, $g'(x_1) = 0$ when $x_1 = c$, $g'(x_1) \leq 0$ when $x_1 \leq c$ and $g'(x_1) \geq 0$ when $x_1 \geq c$. Hence g attains a local minimum at $x_1 = c$ and $\max_{x_1 \in [0,2c]} g(x_1) = g(0) = g(2c) = 2c \log(2)$. Similarly, the restriction of f^2 on $\Omega_{c,2}$ is minimized at $(x_1, x_2) = (c, c)$ and maximized at $(2c, 0)$ and $(0, 2c)$.

Inductive step:

Assume $P \geq 2$ and the restriction of f^P on $\Omega_{c,P}$ is maximized at any vector $(x_1, ..., x_P)$ of the form $(0, ..., x_i = Pc, ..., 0), 1 \leq i \leq P$. Let $x_{P+1} \in [0, (P+1)c]$ and $c_{x_{P+1}} = \frac{1}{P}((P+1)c - x_{P+1})$. For any $(x_1, ..., x_P) \in \Omega_{c_{x_{P+1}},P}$, we have $\sum_{i=1}^{P} x_i + x_{P+1} = Pc_{x_{P+1}} + x_{P+1} = (P+1)c$, hence $(x_1, ..., x_P, x_{P+1}) \in \Omega_{c,P+1}$. Since the converse is also true we have $\Omega_{c,P+1} = \bigcup_{x_{P+1} \in [0,(P+1)c]} \Omega_{c_{x_{P+1}},P} \times \{x_{P+1}\}$, where $A \times B$ is the cartesian product of the sets A and B. We know from the initial assumption that with a fixed value of x_{P+1} the restriction of $h^P : (x_1, ..., x_P) \mapsto f^{P+1}((x_1, ..., x_P), x_{P+1})$ on $\Omega_{c_{x_{P+1}},P}$ is maximized when $(x_1, ..., x_P)$ is a vector belonging to the set $\big\{(0, ..., x_i = Pc_{x_{P+1}}, ..., 0), 1 \leq i \leq P\big\}$. We now want to find the values $\tilde{x}_{P+1}$ that satisfy:

$$\tilde{x}_{P+1} = \underset{x_{P+1} \in [0,(P+1)c]}{\arg\max} \max_{\Omega_{c_{x_{P+1}},P}} h^P(x_1, ..., x_P).$$

Denoting $g_i : x_{P+1} \mapsto f^{P+1}(0, ..., x_i = Pc_{x_{P+1}}, ..., 0, x_{P+1})$, then g_i is differentiable with respect to x_{P+1} on the open interval $(0, (P+1)c)$ and for all $x_{P+1} \in (0, (P+1)c)$ we have:

$$g_i'(x_{P+1}) = \log\left(\tfrac{x_{P+1}}{c}\right) + 1 - \log\left(\tfrac{(P+1)c - x_{P+1}}{c}\right) - 1 = \log\left(\tfrac{x_{P+1}}{(P+1)c - x_{P+1}}\right).$$

Similarly as in the base case, g_i attains a local minimum in the open interval $(0, (P+1)c)$ when $x_{P+1} = (P+1)c/2$. It also attains a maximum on $[0, (P+1)c]$ at $x_{P+1} = 0$, giving $c_{x_{P+1}} = \frac{P+1}{P}c$, and at $x_{P+1} = (P+1)c$, giving $c_{x_{P+1}} = 0$. Therefore the restriction of f^{P+1} on $\Omega_{c,P+1}$ is maximized when $(x_1, ..., x_P, x_{P+1}) \in \{(0, ..., x_i = (P+1)c, ..., 0), 1 \leq i \leq P+1\}$.

S3 Influence of Different Parameters on the MRISE and MIAE

In this section, we provide an extension of the simulation study proposed in Section 4.4. We want to study the effect of parameters such as the coarse resolution j_0, the finest resolution J, the magnitude A_0 and the significance level α on the ranking between the thresholding methods and linear estimator defined in Section 4.4. The statistical validity of this study relies on bootstrapped 95% confidence intervals for the MRISE and the MIAE, obtained from 10000 simulations of the different intensity models.

S3.1 Influence of j_0

Table 1: *Bootstrapped 95% confidence intervals for the MRISE and MIAE with $A_0 = 10000, j_0 = 3, J = 7, M = 1$ and significance level $\alpha = 0.05$. The number in bold indicates the best performing method for each intensity model.*

	Linear	DM-L	LRT-L	LRT-I	LRT-G
Blocks MRISE	2317 ([2315,2319])	1495 ([1493,1497])	1607 ([1604,1609])	**1483 ([1481,1486])**	1782 ([1776,1788])
Blocks MIAE	1838 ([1836,1839])	948 ([945,950])	1062 ([1059,1064])	**939 ([936,941])**	1354 ([1349,1360])
Bumps MRISE	3061 ([3059,3063])	3091 ([3089,3094])	3226 ([3223,3229])	**2957 ([2954,2959])**	3060 ([3058,3061])
Bumps MIAE	2100 ([2099,2102])	1603 ([1601,1605])	1675 ([1673,1677])	**1512 ([1510,1514])**	2100 ([2099,2102])
TriangleSine MRISE	2267 ([2265,2269])	1561 ([1560,1563])	1484 ([1483,1484])	1530 ([1528,1531])	**1360 ([1355,1365])**
TriangleSine MIAE	1809 ([1807,1811])	1347 ([1347,1348])	1307 ([1306,1307])	1281 ([1280,1282])	**1087 ([1083,1091])**

Table 2: *Bootstrapped 95% confidence intervals for the MRISE and MIAE with $A_0 = 10000, j_0 = 0, J = 7, M = 1$ and significance level $\alpha = 0.05$. The number in bold indicates the best performing method for each intensity model.*

	Linear	DM-L	LRT-L	LRT-I	LRT-G
Blocks MRISE	2317 ([2315,2319])	1530 ([1528,1533])	1637 ([1635,1640])	**1493 ([1490,1495])**	1770 ([1765,1776])
Blocks MIAE	1838 ([1836,1839])	999 ([996,1001])	1116 ([1113,1119])	**953 ([950,955])**	1342 ([1337,1347])
Bumps MRISE	3059 ([3057,3060])	3125 ([3122,3128])	3241 ([3238,3244])	**2971 ([2968,2973])**	3065 ([3063,3066])
Bumps MIAE	2100 ([2099,2102])	1683 ([1681,1685])	1758 ([1755,1760])	**1543 ([1541,1545])**	2103 ([2101,2104])
TriangleSine MRISE	2266 ([2264,2268])	1541 ([1540,1542])	1458 ([1457,1458])	1514 ([1513,1515])	**1303 ([1299,1308])**
TriangleSine MIAE	1809 ([1807,1811])	1335 ([1335,1336])	1294 ([1294,1295])	1269 ([1268,1270])	**1042 ([1039,1046])**

S3.2 Influence of J and A_0

Table 3: *Bootstrapped 95% confidence intervals for the MRISE and MIAE with $A_0 = 100000, j_0 = 3, J = 7, M = 1$, and significance level $\alpha = 0.05$. The number in bold indicates the best performing method for each intensity model.*

	Linear	**DM-L**	**LRT-L**	**LRT-I**	**LRT-G**
Blocks MRISE	8798 ([8792,8803])	**6896 ([6891,6902])**	7037 ([7032,7043])	7116 ([7110,7123])	8795 ([8789,8800])
Blocks MIAE	6374 ([6369,6379])	**3302 ([3296,3308])**	3380 ([3374,3387])	3381 ([3374,3387])	6371 ([6366,6376])
Bumps MRISE	21812 ([21809,21815])	21585 ([21581,21588])	21610 ([21607,21614])	**21516 ([21513,21519])**	21812 ([21809,21815])
Bumps MIAE	10060 ([10056,10065])	8237 ([8232,8242])	8248 ([8243,8254])	**8226 ([8221,8232])**	10060 ([10055,10065])
TriangleSine MRISE	7324 ([7317,7330])	8297 ([8288,8306])	8697 ([8687,8708])	6873 ([6866,6879])	**6860 ([6845,6875])**
TriangleSine MIAE	5847 ([5842,5852])	6499 ([6491,6507])	6829 ([6820,6838])	**5352 ([5346,5357])**	5492 ([5479,5504])

Table 4: *Bootstrapped 95% confidence intervals for the MRISE and MIAE with $A_0 = 100000, j_0 = 3, J = 9, M = 1$, and significance level $\alpha = 0.05$. The number in bold indicates the best performing method for each intensity model.*

	Linear	**DM-L**	**LRT-L**	**LRT-I**	**LRT-G**
Blocks MRISE	14435 ([14429,14441])	**6371 ([6362,6380])**	6840 ([6831,6849])	6661 ([6651,6671])	11216 ([11181,11252])
Blocks MIAE	11457 ([11452,11463])	**3236 ([3229,3242])**	3433 ([3426,3439])	3299 ([3292,3306])	8821 ([8791,8852])
Bumps MRISE	16220 ([16212,16229])	14805 ([14792,14818])	15780 ([15767,15793])	**13562 ([13547,13576])**	16217 ([16208,16225])
Bumps MIAE	12071 ([12066,12077])	6904 ([6897,6910])	7161 ([7155,7167])	**6553 ([6546,6559])**	12072 ([12066,12077])
TriangleSine MRISE	14314 ([14308,14320])	8568 ([8558,8577])	10045 ([10033,10057])	**6916 ([6908,6923])**	6987 ([6960,7015])
TriangleSine MIAE	11414 ([11409,11420])	6617 ([6609,6626])	8052 ([8040,8064])	**5372 ([5366,5377])**	5566 ([5543,5589])

S3.3 Influence of α

Table 5: *Bootstrapped 95% confidence intervals for the MRISE and MIAE with $A_0 = 10000, j_0 = 3, J = 7$ and $M = 1$. The number in bold indicates the best choice of α for each method.*

	LRT-L, $\alpha = 0.01$	**LRT-L, $\alpha = 0.05$**	**LRT-I, $\alpha = 0.01$**	**LRT-I, $\alpha = 0.05$**	**LRT-G, $\alpha = 0.01$**	**LRT-G, $\alpha = 0.05$**
Blocks MRISE	1694 ([1692,1697])	**1607 ([1604,1609])**	1530 ([1529,1532])	**1483 ([1481,1485])**	**1648 ([1644,1652])**	1782 ([1776,1788])
Blocks MIAE	1137 ([1134,1139])	**1062 ([1059,1064])**	983 ([981,985])	**939 ([936,941])**	**1210 ([1205,1214])**	1354 ([1348,1359])
Bumps MRISE	3475 ([3472,3479])	**3226 ([3223,3229])**	3090 ([3088,3093])	**2957 ([2954,2959])**	3060 ([3058,3061])	**3060 ([3058,3061])**
Bumps MIAE	1811 ([1809,1813])	**1675 ([1673,1677])**	1591 ([1589,1593])	**1512 ([1510,1514])**	2101 ([2099,2102])	**2100 ([2099,2102])**
TriangleSine MRISE	**1480 ([1479,1480])**	1484 ([1483,1484])	**1521 ([1520,1522])**	1530 ([1528,1531])	**1310 ([1306,1313])**	1360 ([1355,1365])
TriangleSine MIAE	**1304 ([1304,1304])**	1307 ([1306,1307])	1293 ([1292,1294])	**1281 ([1280,1282])**	**1047 ([1044,1050])**	1087 ([1083,1091])

Table 6: *Bootstrapped 95% confidence intervals for the MRISE and MIAE with $A_0 = 10000, j_0 = 0, J = 7$ and $M = 1$. The number in bold indicates the best choice of α for each method.*

	LRT-L, $\alpha = 0.01$	**LRT-L, $\alpha = 0.05$**	**LRT-I, $\alpha = 0.01$**	**LRT-I, $\alpha = 0.05$**	**LRT-G, $\alpha = 0.01$**	**LRT-G, $\alpha = 0.05$**
Blocks MRISE	1744 ([1742,1747])	**1637 ([1635,1640])**	1553 ([1551,1555])	**1493 ([1490,1495])**	**1655 ([1651,1659])**	1770 ([1765,1776])
Blocks MIAE	1223 ([1220,1226])	**1116 ([1113,1118])**	1018 ([1015,1020])	**953 ([950,955])**	**1213 ([1208,1217])**	1342 ([1337,1347])
Bumps MRISE	3500 ([3497,3503])	**3241 ([3238,3244])**	3112 ([3109,3115])	**2971 ([2968,2973])**	3070 ([3069,3072])	**3065 ([3063,3066])**
Bumps MIAE	1917 ([1914,1919])	**1758 ([1755,1760])**	1649 ([1647,1651])	**1543 ([1541,1545])**	2110 ([2109,2112])	**2103 ([2101,2104])**
TriangleSine MRISE	**1453 ([1452,1453])**	1458 ([1457,1458])	**1499 ([1498,1500])**	1514 ([1513,1515])	**1269 ([1266,1272])**	1304 ([1299,1308])
TriangleSine MIAE	**1291 ([1291,1292])**	1294 ([1294,1295])	1279 ([1279,1280])	**1269 ([1268,1270])**	**1023 ([1021,1026])**	1042 ([1039,1046])

S3.4 Discussion

Decreasing j_0 from 3 to 0 is only slightly beneficial for the *TriangleSine* model and increases the MRISE in the two other intensity models. The amelioration observed for *TriangleSine* could be explained by the absence of innovation at levels 0 and 1, and therefore the truly zero coefficients from these scales are less likely to be kept.

When we increase the value of A_0 from 10000 to 100000, the effect of which is to increase the power of each individual LRT involved in the statistical thresholding strategies. For the *Blocks* model, we observe that DM-L is performing better than LRT-L, LRT-I and LRT-G. A study on the asymptotic evolution of the MRISE values as $A_0 \to \infty$ could be done to verify this change of ranking. We also look at the effect of increasing J from 7 to 9 while fixing $A_0 = 100000$. This leads to a significant decrease of the MRISE for all thresholding strategies in the *Bumps* model, as the peaks are located at very fine scales. As expected, it also increases the MRISE for Linear and LRT-G under the *Blocks* model as they keep a larger number of unnecessary coefficients, whereas the performance of DM-L, LRT-L and LRT-I is improved with this choice. However, a significant increase is observed for all thresholding strategies with the *TriangleSine* intensity, which indicates that high resolutions terms penalize the MRISE in this model.

Decreasing α from 0.05 to 0.01, and thus making the hypothesis tests more conservative, seems only interesting for LRT-G as it decreases its MRISE for *Blocks* and *TriangleSine* as well as maintaining similar performance for *Bumps*. For all other methods, choosing $\alpha = 0.01$ leads to a significant increase in MRISE for *Blocks* and *Bumps* whereas a slight decrease is observed for *TriangleSine*. Again, the effect of one parameter on the MRISE is very specific to each intensity model. This whole analysis holds for both choices of j_0.

When using the MIAE as an error measure, the rankings remain generally unchanged. The exceptions to this happen with LRT-I, which performs

better with $\alpha = 0.05$ rather than $\alpha = 0.01$ on *TriangleSine* and significantly better than LRT-G on *TriangleSine* when A_0 is increased to 100000.

S3.5 MATLAB Code

A MATLAB code for the five thresholding strategies defined in Section 4 has been written to be applied on both simulated and real Poisson process data. This code is available using the following link: https://github.com/YoussefT/Statistical-thresholding. The function to run is "*statistical_thresholding.m*", with instructions and details of the inputs specified in the file "*Instructions.html*".

S4 Extension of J-th level homogeneity and L-th level innovation to other wavelets

Since only Haar wavelets allow for a constant reconstruction, an extension of J-th level homogeneity to other wavelets should be based on a different criterion. For some resolution J, we denote $\widetilde{V_{J,K}}$ a subspace of the approximation space V_J spanned by a subset of $K+1$ father wavelets. This allows us to define a flexible generalization of J-th level homogeneity.

Definition S4.1 *Let N be a point process with a locally square integrable intensity λ and let (ϕ, ψ) be a compactly supported wavelet family associated with an MRA of $L^2(\mathbb{R})$. We say that N is* ***level J homogeneous of order K*** *under (ϕ, ψ) if there exists a subspace $\widetilde{V_{J,K}}$ of V_J such that the projection of the intensity on $\widetilde{V_{J,K}}$ admits identical wavelets coefficients. This means $\alpha_{J,k} = \langle \lambda, \phi_{J,k} \rangle = C_J$, for all $\phi_{J,k} \in \widetilde{V_{J,K}}$.*

If N is level J homogeneous of order K under (ϕ, ψ), then the projection of λ on $\widetilde{V_{J,K}}$ is equal to

$$C_J \sum_{\phi_{J,k} \in \widetilde{V_{J,K}}} \phi_{J,k}.$$

Under Haar wavelets and a point process on $[0, T)$, Definition 2 for J-th level homogeneity translates here into a J-th level homogeneity of order $2^J - 1$. From Definition S4.1, every point process for which the intensity is at least locally square integrable is level J homogeneous at order 0 for all compactly supported wavelets and all resolutions, similarly to Definition 2 from which every point process on $[0, T)$ is level 0 homogeneous under Haar wavelets. J-th level homogeneity of order K indicates the largest temporal region where

the intensity display a constant behavior in the sense of the corresponding wavelet family (ϕ, ψ). With Haar wavelets, this is easily interpretable as a constant integrated intensity across $\widetilde{V_{J,K}}$.

An extension of L-th level innovation to other wavelets could initially be thought as following.

Definition S4.2 *Let N be a point process with intensity $\lambda \in L^2(\mathbb{R})$ and let (ϕ, ψ) be a wavelet family associated with an MRA of $L^2(\mathbb{R})$. We then say that N possesses strictly no level L innovation under (ϕ, ψ) if and only if we have $\beta_{L,k} = \langle \lambda, \psi_{L,k} \rangle = 0$, for all $k \in \mathbb{Z}$.*

Although this view of L-th level innovation encompasses the entirety of the detail space W_L, it corresponds to an ideal theoretical case and has at least two limitations. First, it is in practice almost exclusively adapted to Haar wavelets. Indeed, a point process on $[0, T)$ with constant intensity has strictly no level L innovation under the Haar basis. However, its intensity has non zero mother wavelet coefficients for a Daubechies wavelet whose support is partially included in the support of the intensity or for a wavelet with infinite support. Second, Definition S4.2 makes hypothesis testing challenging when one has to test an infinite number of coefficients.

In the spirit of Definition S4.1, we now denote $\widetilde{W_{L,K}}$ the subspace of the detail space W_L generated by a subset of $K+1$ mother wavelets. This again allows us to define a flexible view for the absence of L-th level innovation.

Definition S4.3 *Let N be a point process with a locally square integrable intensity λ and let (ϕ, ψ) be a compactly supported wavelet family associated with an MRA of $L^2(\mathbb{R})$. If there exists a subspace $\widetilde{W_{L,K}}$ of W_L such that λ satisfies $\beta_{L,k} = \langle \lambda, \psi_{L,k} \rangle = 0$, for all $\psi_{L,k} \in \widetilde{W_{L,K}}$, we then say that N exhibits an absence of **level L innovation of order K** under (ϕ, ψ).*

With this perspective, Definition S4.1 and Definition S4.3 follow a consistent modelling. Definition 4 for L-th level innovation would translate here into an absence of level L innovation of order $2^L - 1$ under the Haar basis. One interest of Definition S4.3 is that we can avoid the consideration of non-zero wavelet coefficients that are solely due to boundary effects when the intensity has a compact support. Absence of L-th level innovation of order K has a meaning dependent on the wavelet family (ϕ, ψ). In the case of Haar wavelets, it has an intuitive interpretation as the absence of any change in the integrated intensity between the left and right hand sides of the Haar wavelet. Admittedly, such an interpretation becomes less intuitive with alternative wavelets.

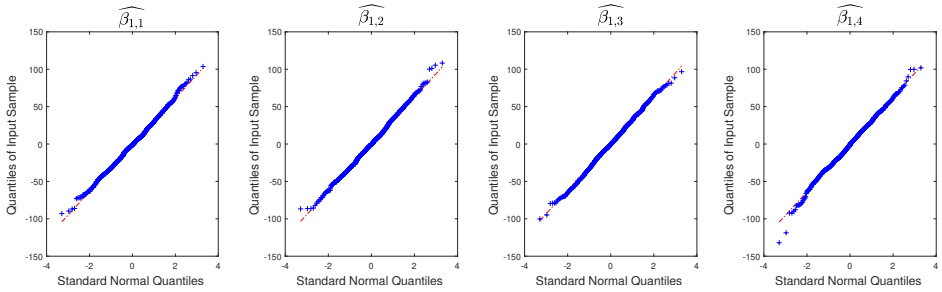


Figure S2: *QQ-plots for the coefficients estimates* $\widehat{\beta}_{1,k}, k = 1, ..., 4$ *with the Daubechies D4 wavelet and a homogeneous Poisson process with intensity* $\lambda_0 = 1000$.

S4.1 Intensity Estimation with Daubechies D4 Wavelets

S4.1.1 Linear estimator

The Daubechies D2Q wavelets (e.g. Daubechies, 1988; Härdle et al, 1998) have *supp* $\phi \subseteq [0, 2Q-1]$ and *supp* $\psi \subseteq [-Q+1, Q]$. When considering Daubechies D2Q wavelets with $Q > 1$, a closed form time domain approximation is needed as there does not exist an exact one. From a set of values obtained with the cascade algorithm (Mallat, 1989), we use a linear interpolation to approximate the mother and father wavelets. As *supp* $\phi \subseteq [0, 2Q-1]$, Daubechies D2Q wavelets do not have disjoint supports across all unit translations for a fixed scale. However *supp* ϕ is finite so we do have a finite number of coefficients that we estimate at each scale. For consistency between the different estimation methods, we desire that the interval $[0, T]$ coincide with the support of the Daubechies D2Q father wavelet at resolution 0. Taking $Q = 2$, this means rescaling the process N to $[0, 3]$, performing the estimation of its intensity, and rescaling this reconstruction back to $[0, T]$. We have the following linear estimator for the projection of the rescaled intensity onto V_J:

$$\widehat{\lambda}^J(t) = \sum_{k=-2}^{(3\times 2^J)-1} \widehat{\alpha}_{J,k}\phi_{J,k}(t).$$

S4.1.2 Coefficient-wise hypothesis test for local thresholding

In order to define thresholding strategies we need to derive the distribution of the mother wavelet coefficients. Consider the collection of mother

Daubechies D2Q wavelets $\{\psi_{L,k}, k \in \mathcal{K}_{\mathcal{L}}\}$ that describes $\widetilde{W_{L,\mathcal{K}_{\mathcal{L}}}}$, the subset of W_L for which all mother wavelets have their support included in $[0,T)$ at each scale L. Under the Daubechies D4 wavelet, we have $\mathcal{K}_{\mathcal{L}} = \{1, ..., (3 \times 2^L) - 2\}$ and hence $\widetilde{W_{L,\mathcal{K}_{\mathcal{L}}}} = \text{span}\{\psi_{L,k}; k = 1, ..., (3 \times 2^L) - 2\}$. As defined in Section 4, let $\widehat{\mathbf{B}}^L = (b_{m,k}) \in \mathbb{R}^{M \times K_L}$ where M is the number of independent realizations of the point process N, $K_L = |\mathcal{K}_L| = (3 \times 2^L) - 2$ and $b_{m,k} \equiv \widehat{\beta}^{(m)}_{L,k}$ is the estimator of the true wavelet coefficient $\beta_{L,k}$ obtained from N_m.

In order to extend the local thresholding scheme based on FDR control to Daubechies D4 wavelets, we need a hypothesis test for each single coefficient. The probability density function of the empirical coefficients for a compactly supported and continuous wavelet family is given in de Miranda (2008). However if the wavelet is non tractable in time domain then so is its density. QQ-plots in Figure S2 suggest that a Gaussian approximation is well suited when the coefficients are estimated using the stochastic integral $\widehat{\beta}_{j_0,k} = \int_{\mathbb{R}} \psi_{j,k}(t) \mathrm{d}N(t) = \sum_{\tau_i \in \mathcal{E}} \psi_{j,k}(\tau_i)$ and ψ is approximated as in S4.1.1. Also, a useful result from de Miranda and Morettin (2011) is $\widehat{\text{Var}}(\widehat{\beta}_{L,k}) = \int \psi^2_{L,k}(t) \mathrm{d}N(t) = \sum_{\tau_i \in \mathcal{E}} \psi^2_{j,k}(\tau_i)$ is an unbiased estimator for the variance of coefficient $\widehat{\beta}_{L,k}$. With $M \geq 1$ independent realizations of the point process N, the estimator of $\beta_{L,k}$ used in the final reconstruction of λ will be the sample mean $\frac{1}{M} \sum_{m=1}^{M} \widehat{\beta}^{(m)}_{L,k}$. Similarly, a variance estimator for $\widehat{\beta}_{L,k}$ is $\frac{1}{M} \sum_{m=1}^{M} \widehat{\text{Var}}(\widehat{\beta}^{(m)}_{L,k})$. Therefore, testing the hypothesis $H : \beta_{L,k} = 0$ against the alternative hypothesis $K : \beta_{L,k} \neq 0$ can be performed using $\frac{1}{M} \sum_{m=1}^{M} \widehat{\beta}^{(m)}_{L,k}$ as a test statistic. Under the null hypothesis, we assume that $\widehat{\beta}^{(m)}_{L,k} \sim \mathcal{N}(0, \sigma^2_{L,k})$. Since $\sigma^2_{L,k}$ is unknown we instead use $\frac{1}{M} \sum_{m=1}^{M} \widehat{\text{Var}}(\widehat{\beta}^{(m)}_{L,k})$ to estimate a confidence interval from a given significance level α. All estimators are consistent so the approximate null distribution converges to the true null distribution as $M \to \infty$.

S4.1.3 L-th level innovation hypothesis test for global thresholding

We now want to design a multivariate test for the null hypothesis $H : \boldsymbol{\mu}_L = \mathbf{0}_L$ where $\boldsymbol{\mu}_L$ is the mean vector of the coefficients $\widehat{\beta}_{L,k}, k \in \mathcal{K}_{\mathcal{L}}$. Given the approximate normality of the coefficients estimates $\widehat{\beta}_{L,k}, k \in \mathcal{K}_{\mathcal{L}}$ under

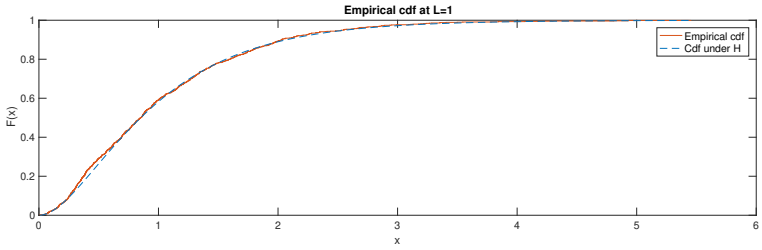


Figure S3: *Comparison between the empirical and null cumulative distribution functions of Hotelling's t-squared statistic with the Daubechies D4 wavelet and a homogeneous Poisson process at resolution* $L = 1$ *with* $\lambda_0 = 1000$ *and* $M = 50$.

the Daubechies D4 wavelet suggested in Figure S2, a possible choice of hypothesis test is the multivariate extension of the Student's t-test based on Hotelling's t-squared statistic. In our setting this statistic will be $t^2 = (\bar{\boldsymbol{\mu}}_L)^T \widehat{\Sigma_L} \bar{\boldsymbol{\mu}}_L$ where $\bar{\boldsymbol{\mu}}_L$ is the sample mean of the empirical coefficients and $\widehat{\Sigma_L}$ their sample covariance. If the estimators $\widehat{\beta}_{L,k}, k \in \mathcal{K}_{\mathcal{L}}$ form a multivariate Gaussian vector, then under the null hypothesis H this statistic is proportional to an F-distributed random variable with parameters M and K_L. The empirical cumulative distribution function of t^2 shown in Figure S3 seems to follow closely the desired distribution under the null hypothesis. However, this particular hypothesis test requires that the sample size M must always be greater than K_L, making it impossible to apply at higher resolutions for low values of M. We will therefore not develop this hypothesis test further in this work.

References

Daubechies I (1988) Orthonormal bases of compactly supported wavelets. Comm Pure Appl Math 41:909–996

Härdle W, Kerkyacharian G, Picard D, Tsybakov A (1998) Wavelets, Approximation, and Statistical Applications, Lecture Notes in Statistics, vol 129. Springer New York

Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(7):674–693

de Miranda JCS (2008) Probability Density Functions of the Empirical Wavelet Coefficients of Multidimensional Poisson Intensities. In: Functional and Operatorial Statistics, Physica-Verlag HD, pp 231–236

de Miranda JCS, Morettin PA (2011) Estimation of the intensity of non-homogeneous point processes via wavelets. Annals of the Institute of Statistical Mathematics 63(6):1221–1246