Statistical Coupling Between time Point-Processes
Samuel Kubler, Jean-Christophe Olivo-Marin, Thibault Lagache

To cite this version:

HAL Id: pasteur-03699883
https://pasteur.hal.science/pasteur-03699883
Submitted on 20 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License
STATISTICAL COUPLING BETWEEN TIME POINT-PROCESSES

Samuel Kubler, Jean-Christophe Olivo-Marin, Thibault Lagache

* Institut Pasteur, Université de Paris, CNRS UMR 3691, BioImage Analysis Unit F-75015 Paris, France.
Corresponding Author: samuel.kubler@pasteur.fr

ABSTRACT
The observation of physical phenomena often goes through the recording of discrete time series of events, that can be represented with marked point processes. The robust estimation of the correlation between two point processes can, therefore, unveil physical mechanisms underlying the observed phenomena. However, the robust estimation of correlation between two, or more, point-processes is hindered by the signal noise (leading to false and missing point detections), the important density of points, and possible time-shift between coupled points. We propose a statistical framework that uses hypothesis testing to estimate coupling between time point-processes. Using simulations, we show that our framework accurately estimates the coupling between two time point-processes even for noisy signal (with false point detections), for high density of points and in the presence of a time shift between coupled points. By applying our statistical framework to the recordings of neuron population activity in mouse visual cortex, we measure the functional coupling between individual neurons, and cluster neurons into functional ensembles.

Index Terms— Time point-processes - Ripley’s K function - Statistical coupling - Neuronal spikes - Functional connectivity

1. INTRODUCTION
In many fields, the observation of physical events can only be done through discrete time series of events. This is the case, for example, of volcanic eruptions [1] and earthquakes [2] in geology, molecules arrival and departures form specific processes sites in cellular biology (e.g. endocytosis and pathogen entry [3]), and neuronal activity through the firing of action potentials from individual neurons [4] which the study case of the experimental part of our article. The statistical characterization of relations (coupling) between two (or more) time series of events can unveil important mechanisms that underly the observed processes. For example, the observation of the sequential arrival of molecules at endocytic sites with fluorescence microscopy helped to unravel the mechanisms of cell entry [3]. Another example is given by the observation of the firing of individual neurons within a population that provides information on neuronal communication and coding [5].

Observed time series of events can often be modeled as marked point processes [6], with the point being the time location of an event and the mark its attributes (e.g. intensity, color, duration...). Therefore, the characterization of the correlation between different time series of events reduces to the estimation of the coupling between time point-processes. In the case of neuronal activity studies, the two main methods used to estimate the coupling between point-processes are either based on the estimation of underlying firing rates (i.e. the intensity of associated point-processes) [7] and the analysis of the correlation between estimated intensities, or coupling estimation is directly performed with colocalization analysis between discrete time point-processes [4]. While the methods using the estimation of firing rates are more robust to the missing and false detections of single point events, they depend on the robust estimation of firing rates and are not well-suited for detecting the synchronization of single point events. On the other hand, the colocalization methods are sensitive to false and missed point detections. Moreover, high point density can lead to fortuitous point colocalization and overestimation of processes’ coupling, whereas time shifts between coupled points can lead to coupling underestimation.

To tackle these technical issues in colocalization analysis, we developed a statistical method to robustly evaluate the coupling between time point-processes, even in the presence of a time-shift between individual coupled points. Our method uses the multi-distance Ripley’s K function [8] to measure the time-shifted accumulation of points from one time point-process relatively to the other. It corresponds to an adaptation of state-of-the-art statistics of point-processes to 1-dimensional temporal case of spiking events. This method has been developed to account for false inferred spike detections and potential time-shifts that results from spike deconvolution methods in calcium imaging [14]. As our method is based on statistical characterization of the Ripley’s K function and hypothesis testing, it is robust to noise (false point detections) and remains accurate even for high point density. We assess the robustness of our method with synthetic simulations, and show that it outperforms state-of-the-art colocalization metrics. Finally, we use our method to compute the functional relations within a population of neurons in the visual cortex of mice, from their individual spiking activity.

2. METHOD
2.1. Measuring the coupling between time point-processes with the Ripley’s K function
Ripley’s K function introduced by Brian Ripley in the 70’s [8] remains the gold-standard to measure the coupling between spatial point processes.

The idea of our temporal adaptation is to measure the coupling between two time point-processes \( s_1 = [t_1, \ldots, t_{n_1}] \) and \( s_2 = [t_1', \ldots, t_{n_2}] \) by creating a regular mesh grid whose topology is dependent on \( s_1 \) point locations in time and by comparing the effective distribution of \( s_2 \) points falling in each mesh to the distribution expected under a random distribution assumption through a statistical hypothesis rejection test. Thus, the Ripley’s vector \( \mathbf{G}_N = [G_0, \ldots, G_i, \ldots, G_{N-1}] \) embeds a 1-dimensional mathematical mesh grid implementation composed by an ensemble of \( N \) fixed size rings of radius \( [r_1, r_{k+1}] \) centered around \( s_1 \) points. Thus, Ripley’s function just corresponds to a statistical effective counter...
of $s_2$ points that fall into rings around $s_1$ objects with a boundary corrective term $w$ which corrects for the potential underestimation of neighbors to points that are close to the starting- and ending-points of the time study period. This correction is inspired by the 2-dimensional corrective term used by Ripley in [8].

$$G_i = \frac{|\Omega|}{n_1 \pi R^2} \sum_{t_k \in \Omega} \sum_{t'_l \in s_2} \mathbb{I}_{[r_i \leq |t_k - t'_l| \leq r_{i+1}]} w(t_k, t'_l)$$

with

$$w(t_k, t'_l) = 1 + \mathbb{I}_{(|t_k - t'_l| > |r_i - \delta t(t_k)|)}$$

where $|\Omega|$ is the length of the time period over which the two time point-processes are observed, $\Omega$ is an indicator function such that $\mathbb{I}(_A) = 1$ if $A$ is True, 0 otherwise. $\delta t$ is the coordinate of the closest boundary of the study domain $\delta t(t_k) = 0$ if $t_k \leq |\Omega|/2$, $\Omega$ otherwise.

To detect a significant coupling between time series (point-processes) $s_1$ and $s_2$, we design an hypothesis testing approach. We compare how far is the effective Ripley’s function counter in the rings regarding the number of points expected under a null hypothesis $H_0$ of Complete Spatial Randomness where points are located according to an homogeneous Poisson distribution for $s_2$ point-process. Analytical mean and standard deviation parameters are derived by calculating intersection of 1-dimensional volumes corresponding to the overlapping of rings. Under $H_0$, Ripley’s K function tends to a normal distribution [9] in accordance with the central limit theorem. Thus, the distribution of $G = \{G_i\}_{i=0...N-1}$ is fully characterized by its mean $M_N = \mathbb{E}\{G_i\} = \mu_i$ and its standard deviation $\Sigma_N = \mathbb{E}\{G_i^2\} - \mu_i^2 = \sigma_i^2$ for $i=0...N-1$. Using the CSR hypothesis for $s_2$ time points and the boundary correction (Eq. 2), we compute that

$$\mu_i = \frac{1}{n_1} \sum_{t_k \in \Omega} \int_{y \in \Omega} \mathbb{I}_{[r_i \leq |t_k - y| \leq r_{i+1}]} (1 + \mathbb{I}_{(|t_k - y| > |r_i - \delta t(t_k)|)}) dy$$

and

$$\sigma_i^2 = \frac{|\Omega|}{n_1^2 \pi R^2} \sum_{t_k \in \Omega} \left( I_{21}(t_k, r_i, r_{i+1}) + \sum_{t_j \in \Omega, t_j \neq t_k} I_{22}(t_k, t_j, r_i, r_{i+1}) \right)$$

with

$$I_{21}(t_k, r_i, r_{i+1}) = -\frac{\mu_i^2}{|\Omega|} + \int_{y \in \Omega} \mathbb{I}_{[r_i \leq |t_k - y| \leq r_{i+1}]} (1 + 3 \times \mathbb{I}_{(|t_k - y| > |r_i - \delta t(t_k)|)}) dy$$

$$I_{22}(t_k, t_j, r_i, r_{i+1}) \approx -\frac{\mu_j^2}{|\Omega|} + \int_{y \in \Omega} \mathbb{I}_{[r_i \leq |t_k - y| \leq r_{i+1}]} \mathbb{I}_{[|t_j - \delta t(t_j)| > r_{i+1}]}$$

$$\times \int_{y \in \Omega} \mathbb{I}_{[r_i \leq |t_k - y| \leq r_{i+1}]} \mathbb{I}_{[r_j \leq |t_j - y| \leq r_{j+1}]} dy$$

The size and number of the rings provide a maximum duration beyond which colocalization can no longer be detected and a temporal resolution to distinguish two close interactions.

### 2.2. Statistical test of time point-processes’ coupling

To build a statistical test of time point-processes’ coupling, we introduce the reduced statistics

$$\tilde{G} = A^{-1} G - \frac{M_N}{\Sigma_N}$$

with $A$ a correction matrix for ring’s overlapping [10]. Under the null hypothesis of $s_2$ randomness, $\tilde{G}$ is a standard normal vector (i.e. each of its component $\tilde{G}_i \sim N(0, 1)$). Therefore, a significantly high value of a vector component $\tilde{G}_i$ would indicate an accumulation of coupled points around reference points corresponding to a positive coupling (while a low value would indicate a depletion of $s_2$ points corresponding to a negative coupling).

Similarly to the statistical test introduced for spatial point-processes [10], we use the maximum component of reduced Ripley vector $\tilde{G}_{\text{max}} = \max_{0 \leq i \leq N-1} \tilde{G}_i$ to test if there is at least one ring $[r_i; r_{i+1}]$ where $s_2$ time points accumulate significantly. To compute the $p-value$ associated with the observed maximum component $\tilde{G}_{\text{max}}$, we compute that, $\forall x > 0$,

$$\Pr(\tilde{G}_{\text{max}} \geq x) = 1 - \Pr(\forall i \in [1...N-1], \tilde{G}_i < x) = 1 - (\Pr(N(0, 1) < x))^N = 1 - \text{cdf}^N(x),$$

where cdf$(x)$ is the cumulative density function of the standard normal law. Finally we obtain the $p-value$

$$p-value = 1 - \text{cdfs}(\tilde{G}_{\text{max}})$$

### 2.3. Quantitative characterization of time point-processes’ coupling

To further characterize the putative coupling between two time point-processes, we determine the components of Ripley’s reduced vector $\tilde{G}$ that are significantly high by using the universal threshold $T(N) = \sqrt{2 \log(N)}$ [11], which is widely used in image processing to determine the significant component of a signal corrupted with standard white noise. Thus, $\tilde{G}_i > T(N)$ indicates that there is a significant accumulation of $s_2$ points at a time shift comprised between $r_i$ and $r_{i+1}$ from $s_1$ points. This allows the detection of coupling between two time point-processes at different distances and not only the detection of points co-occurrence at the same time. Hence, our framework can handle coupling estimation with time shifts and varying delays. Finally, we convert the reduced Ripley vector components into a coupling probability between all the points $(t_k, t_l)$ of time point-processes $s_1$ and $s_2$

$$P((t_k, t_l)) = \sum_{i=0}^{N-1} \mathbb{I}_{[r_i \leq |t_k - t_l| \leq r_{i+1}]} \frac{\sigma_i \tilde{G}_i \mathbb{I}_{(\tilde{G}_i > T(N))}}{\tilde{G}_i}$$

and extract a global coupling metric on the entire time series

$$GC(s_1, s_2) = \frac{1}{n_1} \sum_{t_k \in s_1} \frac{1}{n_1} \sum_{t_l \in s_2} \sum_{P((t_k, t_l)) > 0} P((t_k, t_l))$$

where $n_1$ is the number of $t_k$ coupled points in $s_2$, i.e. such that their coupling probability with $t_k$ in $s_1$ is strictly positive.
To validate our proposed statistical framework we use simulations where the coupling characteristics between time point-processes are known, and we compare the results of our method to standard measures of signals' correlation.

3.1. Synthetic simulations

Using simulations of time point-processes with varying coupling level and time shifts, we compare the accuracy of our statistical method with standard correlation metrics (see Appendix for details):

1. the Pearson correlation coefficient, 2) F1 score and 3) the Cluster Core Index (C3I) [12]. To simulate coupled time-processes with effective coupling level \( p_c \), we first generate a reference homogeneous Poisson process over \( \Omega \), with \( \lambda_0 \) equal to [65%, 13%, 12%, 10%], meaning that 2/3 of the coupled

3. RESULTS

3.1. Robustness to variations of coupling level and time shift

3.1.1. Robustness to variations of coupling level and time shift

To simulate coupled time-processes, as well as when the intensity of processes is overall increased. The only sensitive parameter is the level of missing points that leads to an expected decreased of detected points’ couples.

3.2. Functional coupling between individual neurons

We apply our method to measure the functional coupling between individual neurons from their monitored spiking activity. We use the online dataset from [13], corresponding to two-photon imaging of neuron activity in mouse visual cortex (File M1d1AS in the dataset). From calcium fluorescence traces, the exact spiking times can be obtained using spike inference techniques with variable accuracy and robustness [14]. A representative rasterplot of neuron spiking activity obtained with the constrained FOOPSI deconvolution algorithm [15] is shown in Figure 1-A \( |\Omega| = 5 \) minutes, image acquisition rate = 12.3 Hz). We measure the coupling between individual neurons using the coupling index (eq. 8) (Figure 4-A). Size of used Ripley’s vector (eq. 2) is equal to \( N = 4 \), with identical time-shifts \( t_{i+1} - t_i = 1 \) frame (\( \sim 80 \) ms) for \( i = 0..3 \). Proportions of coupled spikes for the different time shifts are respectively equal to [65%, 13%, 12%, 10%], meaning that 2/3 of the coupled
spike times are co-occurring in the same time step $\Omega_t$, while 1/3 are time-shifted by more than one frame. These latter, time-shifted coupled spikes are completely missed by standard correlation techniques. Using the coupling information between individual neurons, we represent neurons’ couples with an undirected network graph, with edges corresponding to strictly positive coupling indexes (Figure 4-B). We identify neuronal ensembles with a Louvain clustering algorithm [16] which maximizes graph modularity, and obtain 3 neuronal ensembles (with $n \geq 10$ neurons), in agreement with [13].

We conclude that our statistical framework allows the robust estimation of functional coupling between individual neurons from the calcium imaging of their spiking activity. Contrary to standard correlation techniques, it allows the estimation of points’ coupling even in the presence of time-shifts, and does not require any thresholding of correlation coefficients for the network representation and clustering of neuronal connectivity.

4. CONCLUSION

We have proposed a statistical method for estimating the coupling between time point-processes that use the multi-distance Ripley’s $K$ function and hypothesis-testing framework. Our method is able to accurately estimate the coupling between time-shifted correlated point processes, and is robust to high intensity of point processes and false detections. The unique ability of our framework to compute the coupling between time-shifted point-processes is used to quantify the functional coupling between individual neurons imaged with fluorescent calcium indicators in mouse visual cortex [13].

5. APPENDIX

**Pearson correlation coefficient** For a discretized time period $\Omega = \{\Omega_t\}_{1 \leq t \leq T}$, we introduce the indicator functions $\delta_t(s_k) = 1 \{\exists t \in s_k | t \in \Omega_t\}$ for each time-process $(s_k)_{k=1,2}$ that determines whether at least one point of each time point-process falls into the time step $\Omega_t$. The Pearson correlation coefficient is then given by

$$r(s_1, s_2) = \frac{\sum_{1 \leq t \leq T} \delta_t(s_1)\delta_t(s_2) - T\bar{s}_1\bar{s}_2}{\sqrt{\sum_{1 \leq t \leq T} \delta_t(s_1)^2 - T\bar{s}_1^2} \sqrt{\sum_{1 \leq t \leq T} \delta_t(s_2)^2 - T\bar{s}_2^2}},$$

with $\bar{s}_k = T^{-1}\sum_{1 \leq t \leq T} \delta_t(s_k)$.

**F1 score** To compute the F1 score between time point-processes $s_1$ and $s_2$, we set a tolerance for point-matching ($tol = 2$ frames) and define true positive (TP) as $TP = \sum_{1 \leq t \leq T} \delta_t(s_1)\delta_{t \pm tol}(s_2)$ with $\delta_{t \pm tol}(s_2) = 1 \{\exists t \in s_2 | t \in \bigcup_{t-tol \leq \ell \leq t+tol} \Omega_t\}$. False positive (FP) and false negative (FN) are respectively given by $FP = \sum_{1 \leq t \leq T} \delta_t(s_2)(1 - \delta_{t \pm tol}(s_1))$ and $FN = \sum_{1 \leq t \leq T} \delta_t(s_1)(1 - \delta_{t \pm tol}(s_2))$. Finally F1 score is equal to

$$F1 = \frac{2 \times TP}{2 \times TP + (FP + FN)}.$$

6. COMPLIANCE WITH ETHICAL STANDARDS

This is a numerical simulation study for which no ethical approval was required.

7. ACKNOWLEDGMENTS

No funding was received for conducting this study. The authors have no relevant financial or non-financial interests to disclose.

8. REFERENCES


