



**HAL**  
open science

## Surrogate data methods based on a shuffling of the trials for synchrony detection: the centering issue

Mélanie Albert, Yann Bouret, Magalie Fromont, Patricia Reynaud-Bouret

### ► To cite this version:

Mélanie Albert, Yann Bouret, Magalie Fromont, Patricia Reynaud-Bouret. Surrogate data methods based on a shuffling of the trials for synchrony detection: the centering issue. *Neural Computation*, 2016, 28 (1), pp.2352-2392. 10.1162/NECO\_a\_00839 . hal-01154918v2

**HAL Id: hal-01154918**

**<https://hal.science/hal-01154918v2>**

Submitted on 12 Jul 2016

**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.

0001  
0002  
0003  
0004  
0005  
0006        Surrogate data methods based on a shuffling of  
0007 the trials for synchrony detection: the centering is-  
0008 sue

0009  
0010        **Mélisande Albert**<sup>1</sup>

0011        *Melisande.Albert@unice.fr*

0012        **Yann Bouret**<sup>2</sup>

0013        *Yann.Bouret@unice.fr*

0014        **Magalie Fromont**<sup>3</sup>

0015        *magalie.fromont@univ-rennes2.fr*

0016        **Patricia Reynaud-Bouret**<sup>1</sup>

0017        *reynaudb@unice.fr*

0018        <sup>1</sup>Univ. Nice Sophia Antipolis, CNRS, LJAD, UMR 7351, 06100 Nice, France.

0019        <sup>2</sup>Univ. Nice Sophia Antipolis, CNRS, LPMC, UMR 7336, 06100 Nice, France.

0020        <sup>3</sup>Univ. Européenne de Bretagne, CNRS, IRMAR, UMR 6625, 35043 Rennes Cedex,  
0021 France.

0022        **Keywords:** Unitary Events - Synchronization - Independence tests - Trial-Shuffling  
0023 - Permutation - Bootstrap - Point process - Multiple testing

0024  
0025  
0026  
0027  
0028  
0029  
0030  
0031  
0032  
0033  
0034  
0035  
0036  
0037  
0038  
0039  
0040  
0041  
0042  
0043  
0044  
0045  
0046

## Abstract

We investigate several distribution free dependence detection procedures, all based on a shuffling of the trials, from a statistical point of view. The mathematical justification of such procedures lies in the bootstrap principle and its approximation properties. In particular we show that such a shuffling has mainly to be done on centered quantities - that is quantities with zero mean under independence - to construct correct  $p$ -values, meaning that the corresponding tests control their False Positive (FP) rate. Thanks to this study, we introduce a method, named Permutation UE, which consists in a multiple testing procedure based on permutation of experimental trials and delayed coincidence count. Each involved single test of this procedure achieves the prescribed level, so that the corresponding multiple testing procedure controls the False Discovery Rate (FDR), and this with as few assumptions as possible on the underneath distribution, except independence and identical distribution across trials. The mathematical meaning of this assumption is discussed and it is in particular argued that it does not mean what is commonly referred in neuroscience to as cross-trials stationarity. Some simulations show moreover that Permutation UE outperforms the trial-shuffling of (Pipa & Grün, 2003) and the MTGAUE method of (Tuleau-Malot et al., 2014) in terms of single levels and FDR, for a comparable amount of False Negatives. Application to real data is also provided.

# 1 Introduction

The possible time dependence either between cerebral areas or between neurons, and in particular the synchrony phenomenon, has been vastly debated and investigated as a potential element of the neuronal code (Singer, 1993). To detect such a phenomenon at the microscopic level, multielectrodes are usually used to record the nearby electrical activity. After pretreatment, the time occurrences of action potentials (spikes) for several neurons are therefore available. One of the first steps of analysis is then to understand whether and how two simultaneously recorded spike trains, corresponding to two different neurons, are dependent or not.

Several methods have been used to detect synchrony (Perkel et al., 1967; Aertsen et al., 1989). Among the most popular ones, the Unitary Events (UE) method, due to Grün and collaborators (Grün, 1996; Grün et al., 2002a,b, 2010), has been applied in the last decade to a vast amount of real data (see, e.g., (Kilavik et al., 2009) and references therein). Two of its main features are at the root of its popularity: the UE method is not only able to give a precise location in time of the dependence periods, but also to quantify the degree of dependence by providing  $p$ -values for the independence tests.

One can decompose the method in three main steps:

(i) The first step consists in choosing a way to count coincidences. In the original UE method, the point processes modeling the data are binned and clipped at a rough level (see Figure 1.A for a more precise description), the bins being about 5 ms wide. However, it is proven in (Grün et al., 1999) that the binned coincidence count as a result of this preprocessing may induce a loss in synchrony detection of about 60% in certain cases (see also an illustration in Figure 1.A). The idea of (Grün et al., 1999) was

0070 therefore to keep the data at the initial resolution level despite its high dimension, but  
0071 to define the notion of multiple shift (MS) coincidence count, nicely condensing the  
0072 dependence feature that neurobiologists want to analyze without any loss in synchrony  
0073 detection. The delayed coincidence count is a generalization of this notion to non dis-  
0074 cretized process and which still does not suffer from any loss in synchrony detection  
0075 (see Figure 1.B). Other coincidence count notions have also been used such as the one  
0076 in (Louis et al., 2010b) which also holds for non discretized process.

0077 (ii) Once the coincidence count is fixed, one needs to understand what is the typical  
0078 behavior of this quantity under independence, so that independence is rejected if the  
0079 count is significantly unusual. To do so, the original method estimates the expected  
0080 number of coincidence under independence and assumes a Poisson distribution of the  
0081 count under independence. This assumption has been shown to be non completely ad-  
0082 equate in (Pipa et al., 2013; Tuleau-Malot et al., 2014) and the plug-in of estimates of  
0083 the underlying firing rates has also been discussed in (Gütig et al., 2001; Tuleau-Malot  
0084 et al., 2014). Notably in (Gütig et al., 2001) another method is introduced which uses  
0085 conditional distribution to avoid the misuse of plug-in estimates. However in all those  
0086 works, very strong assumptions on the distribution of the spikes are made: either bins  
0087 are assumed to be independent and identically distributed or the spike trains are as-  
0088 sumed to be Poisson or at least renewal processes. However conclusive experimental  
0089 evidence combined with many statistical and modeling studies show that those distri-  
0090 bution assumptions are not realistic - see (Nawrot et al., 2008; Farkhooi et al., 2009;  
0091 Pouzat & Chaffiol, 2009; Avila-Akerberg & Chacron, 2011) and the references therein.

0092 This is the reason why a lot of interest has been shown on surrogate data methods

0093 - see (Grün, 2009; Louis et al., 2010a) for a methodological review. These methods,  
0094 unlike the ones cited above, are not linked to a particular coincidence count and they  
0095 can be indifferently applied to any of the previous counts discussed above. The main  
0096 idea is to use the original observed data set and to combine it with a computer ran-  
0097 dom generator to produce new artificial data sets mimicking how the data set would  
0098 behave under independence. Thanks to these surrogate data sets, it is a priori possible  
0099 to estimate the distribution of the coincidence count under independence and therefore  
0100 to build reasonable  $p$ -values. This can usually be achieved in practice through parallel  
0101 programming and Monte-Carlo approximation (Louis et al., 2010a).

0102 There are mainly two trends in surrogate data methods. Either the trials are shuffled  
0103 (Pipa & Grün, 2003; Pipa et al., 2003), but it has been shown that this method suf-  
0104 fers from a non controlled False Positive rate when there is cross-trials non-stationarity  
0105 (Grün et al., 2003); or the spikes themselves are slightly moved as in the dithering  
0106 method - see (Louis et al., 2010b) and the references therein. This last method is more  
0107 able to cope with cross-trials non-stationarity. Indeed, and even under cross-trials non-  
0108 stationarity, several more or less technical variants of this method are able to reproduce  
0109 the mean intensity, also called profile or rate, and even the interspike interval distribu-  
0110 tion. However those methods cannot mimic the whole distribution of the coincidence  
0111 count under independence. As a consequence, the best dithering methods, in the sense  
0112 that these methods are able to control their False Positive rate even for highly non ho-  
0113 mogeneous processes in time, are much too conservative, as assessed in (Louis et al.,  
0114 2010b).

0115 (iii) The third step of a UE method is linked to the multiple testing aspect, seen by

0116 S. Grün herself (Grün, 2009) as "a useful side-effect" of the fact that the original UE  
0117 method needs homogeneity in time. Because of this drawback in the original procedure,  
0118 the UE tests described above are performed on small sliding windows on which the  
0119 homogeneity assumption is realistic. This allows, as a by-product, "a time resolved  
0120 analysis [which] shows potential modulation of synchrony". It has been proved however  
0121 in (Tuleau-Malot et al., 2014) that the procedure needs therefore to be corrected for the  
0122 multiplicity of the tests with, for instance, Benjamini and Hochberg procedure to control  
0123 the False Discovery Rate (Benjamini & Hochberg, 1995).

0124 We here focus on surrogate data methods based on a shuffling or resampling of the  
0125 trials. Such procedures are mathematically justified by the bootstrap principle that in-  
0126 deed provides several ways to shuffle trials, all able to reproduce the distribution of the  
0127 count under independence, if applied to centered quantities. So our main concern is to  
0128 warn people using methods based on a shuffling of the trials against a direct applica-  
0129 tion of these methods to rough coincidence counts, which are not correctly centered.  
0130 As a consequence of this study, we show that a permutation of the trials in line with  
0131 (Hoeffding, 1952; Romano, 1989; Romano & Wolf , 2005) is the most able to mimic  
0132 the correct distribution among the resampling approaches investigated here. We couple  
0133 it with the delayed coincidence count to avoid loss in synchrony detection and a Ben-  
0134 jamini and Hochberg procedure controlling the False Discovery Rate when considering  
0135 sliding windows, to obtain a new method named Permutation UE. Because resampling  
0136 methods are quite demanding in terms of computational cost, we also propose a fast  
0137 algorithm to compute the delayed coincidence count, with a computational cost equiv-  
0138 alent to the one of the binned coincidence count.

0139 A major assumption of the present work, due to the shuffling of the trials, is the  
0140 independence and the identical distribution between trials in the probabilistic sense.  
0141 However this mathematical notion does not mean stationarity across trials, as com-  
0142 monly expressed in the neuroscience literature (Arieli et al., 1996; Grün et al., 2003;  
0143 Churchland et al., 2010; Nawrot, 2010; Farkhooi et al., 2011; Churchland et al., 2011;  
0144 Litwin-Kumar & Doiron, 2012; Farkhooi et al., 2013). This point is clearly discussed  
0145 hereafter in Section 5.

0146 We begin with describing the mathematical framework in Section 2, by giving the  
0147 notation and the definition of binned and delayed coincidence counts together with a  
0148 detailed fast algorithm to compute the delayed one. In Section 3, we precisely discuss  
0149 the centering problem and its effect on the methods based on a shuffling of the trials.  
0150 In Section 4, we detail the Permutation UE method and apply it to real data. In Section  
0151 5, we discuss the limit of the methods in terms of both distribution free aspects, and  
0152 cross-trials stationarity and provide some open questions.

0153

## 0154 **2 Framework**

0155

0156 We start by giving some useful notation and reminders to understand the construction  
0157 and discussion of the dependence detection methods using a shuffling of the trials.

0158

### 0159 **2.1 Notation**

0160

0160 In all the sequel,  $X^1$  and  $X^2$  denote two point processes modeling the spike trains of  
0161 two simultaneously recorded neurons and  $X$  represents the couple  $(X^1, X^2)$ . The ab-

0162 breviation "i.i.d." stands for independent and identically distributed. In this sense, by  
0163 assuming that  $n$  independent and identically distributed trials are observed, the obser-  
0164 vation is modeled by an i.i.d. sample of size  $n$  of couples from the same distribution  
0165 as  $X$ , meaning  $n$  i.i.d. copies  $X_1, \dots, X_n$  of  $X$ . This sample is denoted in the sequel  
0166 by  $\mathbb{X}_n = (X_1, \dots, X_n)$ . The corresponding probability and expectation are respectively  
0167 denoted by  $\mathbb{P}$  and  $\mathbb{E}$ . For another random variable,  $Y$ , conditional probability and con-  
0168 ditional expectation given  $Y$  are respectively denoted  $\mathbb{P}(\cdot|Y)$  and  $\mathbb{E}[\cdot|Y]$ , they both are  
0169 random quantities that still depend on the value of  $Y$ .

0170 The notation  $\mathbf{1}_{X \in A}$  stands for a function whose value is 1 if  $X$  belongs to  $A$  and 0  
0171 otherwise. In particular note that

$$0172 \mathbb{P}(X \in A) = \mathbb{E}[\mathbf{1}_{X \in A}] = \mathbb{E}[\mathbb{E}[\mathbf{1}_{X \in A}|Y]] = \mathbb{E}[\mathbb{P}(X \in A|Y)],$$

0173 which amounts to integrate first in the conditional distribution of  $X$  given  $Y$  and then  
0174 to integrate in the distribution of  $Y$ .

0175 Since assessing dependence between  $X^1$  and  $X^2$  is the main focus of the present  
0176 work, the following notation is useful:  $X^\perp$  denotes a couple  $(X^{1,\perp}, X^{2,\perp})$  such that  
0177  $X^{1,\perp}$  (resp.  $X^{2,\perp}$ ) has the same distribution as  $X^1$  (resp.  $X^2$ ), but  $X^{1,\perp}$  is independent of  
0178  $X^{2,\perp}$ . In particular, the couple  $X^\perp$  has the same marginals as the couple  $X$ . Moreover,  
0179  $\mathbb{X}_n^\perp = (X_1^\perp, \dots, X_n^\perp)$ , with  $X_i^\perp = (X_i^{1,\perp}, X_i^{2,\perp})$ , denotes an i.i.d. sample of size  $n$  from  
0180 the same distribution as  $X^\perp$ , and  $\mathbb{P}_\perp$  and  $\mathbb{E}_\perp$  are the corresponding probability and  
0181 expectation.

0182 Note in particular that if the two observed neurons indeed behave independently,  
0183 then the observed sample  $\mathbb{X}_n$  has the same distribution as  $\mathbb{X}_n^\perp$ .

0184 Finally, for any point process  $X^j$  ( $j = 1, 2$ ),  $dN_{X^j}$  stands for its associated point

0185 measure, defined for all function  $f$  by:

$$0186 \int f(u) dN_{X^j}(u) = \sum_{T \in X^j} f(T),$$

0187

0188 and for any interval  $I$ ,  $N_{X^j}(I)$  denotes the number of points of  $X^j$  observed in  $I$ .

0189

## 0190 **2.2 Binned and delayed coincidence counts**

0191 Because of the way neurons transmit information through action potentials, it is com-  
0192 monly admitted that the dependence between the spike trains of two neurons is due to  
0193 temporal correlations between spikes produced by both neurons. Informally, a coinci-  
0194 dence occurs when two spikes (one from each neuron) appear with a delay less than  
0195 a fixed  $\delta$  (of the order of a few milliseconds). Several coincidence count functions  
0196 have been defined in the neuroscience literature, and among them the classical binned  
0197 coincidence count, introduced in (Grün et al., 2002a,b).

0198

**Definition 1** *The binned coincidence count between point processes  $X^1$  and  $X^2$  on the*  
0199 *interval  $[a, b]$  with  $b - a = M\delta$  for an integer  $M \geq 2$  and a fixed delay  $\delta > 0$  is given*  
0200 *by*

0201

$$0202 \psi_{\delta}^{coinc}(X^1, X^2) = \sum_{\ell=1}^M \mathbf{1}_{N_{X^1}(I_{\ell}) \geq 1} \mathbf{1}_{N_{X^2}(I_{\ell}) \geq 1},$$

0203

0204 where  $I_{\ell}$  is the  $\ell$ th bin of length  $\delta$ , i.e.  $[a + (\ell - 1)\delta, a + \ell\delta)$ .

0205

0206 More informally, the binned coincidence count is the number of bins that contain at  
0207 least one spike of each spike trains, as one can see on Figure 1.A.

0208

0209 The binned coincidence count computation algorithm is usually performed on al-  
0210 ready binned and clipped data (see Figure 1). Therefore, given two sequences of 0 and

0211

0208 1 of length  $M = (b - a)\delta^{-1}$ , the number of operations needed to compute the binned  
 0209 coincidence count is  $O(M) = O((b - a)\delta^{-1})$  (without taking the binning preprocessing  
 0210 into account).

0211 The more recent notion of delayed coincidence count, introduced in (Tuleau-Malot  
 0212 et al., 2014), is a generalization of the multiple-shift coincidence count, defined in (Grün  
 0213 et al., 1999) for discretized point processes, to non necessarily discretized point pro-  
 0214 cesses.

0215 **Definition 2** *The delayed coincidence count between point processes  $X^1$  and  $X^2$  on*  
 0216 *the interval  $[a, b]$  is given by*

$$0217 \varphi_{\delta}^{coinc}(X^1, X^2) = \int_a^b \int_a^b \mathbf{1}_{|u-v| \leq \delta} dN_{X^1}(u) dN_{X^2}(v),$$

0219 More informally,  $\varphi_{\delta}^{coinc}(X^1, X^2)$  is the number of couples of spikes (one spike from  $X^1$   
 0220 and one from  $X^2$ ) appearing in  $[a, b]$  with delay at most equal to  $\delta$ . A visual example  
 0221 is given on Figure 1.B. Note in particular that two coincidences are discarded by the  
 0222 binned coincidence count on this particular example: one because of the clipping effect  
 0223 in the third bin and one because of the effect of adjacent bins in the seventh and eighth  
 0224 bins. Both of them are counted in the delayed coincidence count.

0225 A rather naive algorithm to compute delayed coincidence count would test whether  
 0226 for any pair  $(u, v)$  of a spike  $u$  in  $X^1$  and a spike  $v$  in  $X^2$ , the delay  $|u - v|$  is less than  
 0227  $\delta$  and to count the number of hits. This would lead to an algorithm whose complexity  
 0228 is in the product of the number of points in each spike train. If one assumes both spike  
 0229 trains to be Poisson with intensity  $\lambda_1$  and  $\lambda_2$ , this algorithm has an average cost of  
 0230 order  $O(\lambda_1 \lambda_2 (b - a)^2)$  and is therefore quadratic in the length of the interval. One can

0231 actually drastically improve this rate thanks to the following algorithm for which the  
 0232 result  $c := \varphi_\delta^{coinc}(X^1, X^2)$  is the delayed coincidence count.

0233 **Delayed coincidence count algorithm**

0234 Given two sequences  $x_1$  and  $x_2$  of ordered points with respective lengths  $n_1 =$   
 0235  $N_{X^1}([a, b])$  and  $n_2 = N_{X^2}([a, b])$ , representing the observations of two point pro-  
 0236 cesses  $X^1$  and  $X^2$ ,

0237 - Initialize  $j = 1$  and  $c = 0$ .

0238 - For  $i = 1, \dots, n_1$ ,

0239 1. Assign  $x_{low} = x_1[i] - \delta$ .

0240 2. While  $j \leq n_2$  and  $x_2[j] < x_{low}$ ,  $j = j + 1$ .

0241 3. If  $j > n_2$ , stop.

0242 4. Else (here necessarily,  $x_2[j] \geq x_{low}$ ),

0243 4.a Assign  $x_{up} = x_1[i] + \delta$  and  $k = j$ .

0244 4.b While  $k \leq n_2$  and  $x_2[k] \leq x_{up}$ ,  $c = c + 1$  and  $k = k + 1$ .

0245 This algorithm is slightly more intricate but the computational complexity is much  
 0246 smaller than the previous one. Figure 1.C gives a visualization of the algorithm on  
 0247 a very simple example. The main point is that the index  $j$  in Step 2 cannot decrease  
 0248 and therefore it is not making a double full loop on all the indices of both sequences  
 0249  $x_1$  and  $x_2$ . A pseudo double loop is made thanks to the index  $k$  in step 4.b which  
 0250 indeed can take several times the same value but whose range is only governed by the  
 0251 number of points that appear in an interval of length  $2\delta$ , namely  $[x_{low}, x_{up}]$ , which is  
 0252 usually much smaller than the total length of the sequence  $x_2$ . More precisely, the  
 0253 complexity of the algorithm is therefore upper bounded, up to a constant, by  $n_1$  (for  
 steps 1, 3 and 4.a), plus  $n_2$  (for all steps 2 on all points of  $x_1$ , that is the range of

0254 the index  $j$  which never decreases), and plus  $n_1$  times the number of points of  $x_2$  in a  
0255 segment (namely  $[x_{low}, x_{up}]$ ) of length  $2\delta$  (for step 4.b). On average, if  $X^1$  and  $X^2$  are  
0256 for instance independent homogeneous Poisson processes of respective intensities  $\lambda_1$   
0257 and  $\lambda_2$ , the complexity is of order  $O((\lambda_1 + \lambda_2 + \lambda_1\lambda_2\delta)(b - a))$ . As compared with  
0258 the binned coincidence count algorithm, whose complexity is of order  $O(\delta^{-1}(b - a))$ ,  
0259 the present delayed coincidence count algorithm is therefore advantageous as soon as  
0260  $\lambda_1\delta \ll 1$  and  $\lambda_2\delta \ll 1$ , conditions that are usually satisfied in practice (take for  
0261 instance  $\lambda_1 = \lambda_2 = 30\text{Hz}$  and  $\delta = 0.005\text{s}$ , which gives  $\lambda_1\delta = \lambda_2\delta = 0.15$ ). Even if  
0262 both algorithms are linear in  $(b - a)$ , the delayed coincidence count algorithm exploits  
0263 the sparsity of the spike trains through the constant  $(\lambda_1 + \lambda_2 + \lambda_1\lambda_2\delta)$  in its complexity,  
0264 instead of  $\delta^{-1}$  in the complexity of the binned coincidence count algorithm. In Figure 1  
0265 is given a more visual representation of this sparsity: notice for instance that the bins  
0266 with 0's do not even have to be taken into account in the present delayed coincidence  
0267 count algorithm.

0268 As explained in the introduction - point (ii) - all surrogate data methods (see (Louis  
0269 et al., 2010a)) could in principle be applied to this notion of delayed coincidence count,  
0270 at least when only two neurons are involved. In the sequel and for illustration purpose  
0271 in the simulations, we apply the different surrogate methods of trial-shuffling type to the  
0272 delayed coincidence count but the mathematical justification (Albert et al., 2015) and  
0273 therefore the described behaviors in Section 3 are the same whatever the coincidence  
0274 count that one would like to consider.

0275

0276

0277  
0278  
0279  
0280  
0281  
0282  
0283  
0284  
0285  
0286  
0287  
0288  
0289  
0290  
0291  
0292  
0293  
0294  
0295  
0296  
0297  
0298  
0299

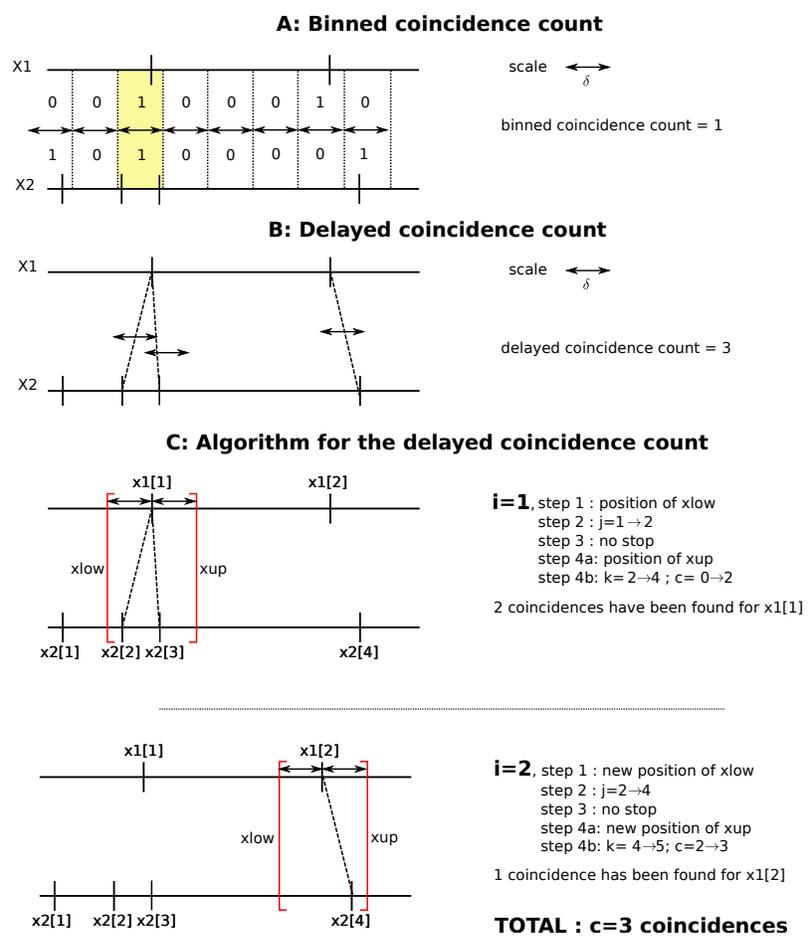


Figure 1: Coincidence counts. Part **A** gives an example of binned coincidence count on a couple of spike trains ( $X^1, X^2$ ) (the spikes corresponding to the respective dashes on the line): after binning the data into blocks of length  $\delta$ , one only keeps the information whether there is at least a spike or not in the bin (clipping). The binned coincidence count is then the number of times there is a "1" for each spike train in the same bin. Part **B** gives on the same example the number of delayed coincidence count, that is the number of pairs of points (one on each spike train) at distance less than  $\delta$ . Note that these two coincidence counts are on this particular example different. Part **C** provides a visualization of the first steps of the proposed algorithm. In particular, note that it exploits the sparsity of the data represented via the vector  $x_1$  and  $x_2$ : there is no computational time spent on the central part of the drawing corresponding to the 0's of the binned process.

### 3 Bootstrap and centering issues

Given an interval of time  $[a, b]$  and the observation of a sample  $\mathbb{X}_n = (X_1, \dots, X_n)$  corresponding to  $n$  different trials in this interval, we focus here on the problem of testing the null hypothesis:

$$(H_0) \text{ ”}X^1 \text{ and } X^2 \text{ are independent on } [a, b]\text{”}$$

against:

$$(H_1) \text{ ”}X^1 \text{ and } X^2 \text{ are not independent on } [a, b]\text{”}.$$

All existing UE methods are based on the total number of coincidences:

$$\mathbf{C} = \mathbf{C}(\mathbb{X}_n) = \sum_{i=1}^n \varphi(X_i^1, X_i^2),$$

where  $\varphi$  generically denotes either  $\psi_\delta^{coinc}$ , or  $\varphi_\delta^{coinc}$ , or other coincidence count functions that practitioners would like to use (see (Albert et al., 2015) for other choices).

To underline what is observed or not, when  $\mathbf{C}$  is computed on the observation of  $\mathbb{X}_n$ , it is denoted by  $\mathbf{C}^{obs}$ , the total number of observed coincidences.

In the following, several of these UE methods are described, which all rely on the same paradigm: ”reject  $(H_0)$  when  $\mathbf{C}^{obs}$  is significantly different from what is expected under  $(H_0)$ ”. More precisely, the independence null hypothesis  $(H_0)$  is rejected and the dependence is detected when a quantity, based on the difference between the observed coincidence count and what is expected under  $(H_0)$ , is smaller or larger than some critical values. Those critical values are obtained in various ways, each of them being peculiar to each method.

### 3.1 Importance of a centering step when parameters are unknown

Before explaining the various resampling methods based on a shuffling of the trials investigated here where the centering issue appears as a major point, we want to underline that such a centering issue also occurs in more naive methods, for which this problem is easier to understand. Informally, there is a *centering issue* if a method is able to reproduce the distribution of centered quantities (that is with zero mean under independence) but is not able to do so for non centered quantities.

Let us first look at a toy example. If the values of the expectation and the variance of  $\mathbf{C}$  under  $(H_0)$ , that is

$$c_0 = \mathbb{E}_{\perp}[\mathbf{C}] \text{ and } v_0 = \mathbb{E}_{\perp}[(\mathbf{C} - c_0)^2],$$

are precisely known, then the classical central limit theorem gives under independence that

$$\frac{\mathbf{C}(\mathbb{X}_n^{\perp}) - c_0}{\sqrt{v_0}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1). \quad (1)$$

This means in particular that when the number of trials  $n$  tends to infinity, the cumulative distribution function and the quantiles of  $(\mathbf{C}(\mathbb{X}_n^{\perp}) - c_0)/\sqrt{v_0}$  are tending to the ones of a standard Gaussian distribution,  $\mathcal{N}(0, 1)$ , that is a Gaussian variable of mean 0 and variance 1. Then, given  $\alpha$  in  $(0, 1)$ , the test which consists in rejecting  $(H_0)$  when  $(\mathbf{C}^{obs} - c_0)/\sqrt{v_0}$  is larger than  $z_{1-\alpha}$ , the  $1 - \alpha$  quantile of a standard Gaussian distribution, is asymptotically (in  $n$ , the number of trials) of False Positive (FP) rate  $\alpha$ . It means that, for this test, the probability of rejecting independence, whereas independence holds, is asymptotically equal to the prescribed  $\alpha$ .

In this particular case, we could rewrite the above procedure in a complete equiv-

0346 alent way, as follows: we reject the independence ( $H_0$ ) when  $\mathbf{C}^{obs}$  is larger than the  
 0347  $1 - \alpha$  quantile of  $\mathcal{N}(c_0, v_0)$ , Gaussian distribution of mean  $c_0$  and variance  $v_0$ . Another  
 0348 way to state this is that as long as  $c_0$  and  $v_0$  are known, approximating the distribu-  
 0349 tion of  $(\mathbf{C}(\mathbb{X}_n^\perp) - c_0)/\sqrt{v_0}$  by  $\mathcal{N}(0, 1)$  or approximating the distribution of  $\mathbf{C}(\mathbb{X}_n^\perp)$  by  
 0350  $\mathcal{N}(c_0, v_0)$  is completely equivalent: this is due to the scaling and shifting properties of  
 0351 the Gaussian distributions.

0352 However, if  $c_0$  and  $v_0$  are unknown, and it is always the case in practice even if one  
 0353 assumes Poissonian spike trains (since the firing rates are unknown), one would like  
 0354 to replace  $c_0$  and  $v_0$  by estimates, namely  $\hat{c}_0$  and  $\hat{v}_0$  and proceed as previously. It has  
 0355 been shown in (Tuleau-Malot et al., 2014) that we cannot do that. Indeed the plug-  
 0356 in step which consists in estimating the distribution of  $\mathbf{C}(\mathbb{X}_n^\perp)$  by  $\mathcal{N}(\hat{c}_0, \hat{v}_0)$  instead of  
 0357  $\mathcal{N}(c_0, v_0)$  does not work for the non centered quantity  $\mathbf{C}$ . Only the Gaussian approx-  
 0358 imation of the distribution of the centered quantity, namely  $\mathbf{C}(\mathbb{X}_n^\perp) - \hat{c}_0$ , holds and at  
 0359 the price of modified variance. Note that this plug-in issue is known in different terms  
 0360 since (Gütig et al., 2001), who advertise for the use of conditional distribution. However  
 0361 both (Tuleau-Malot et al., 2014) and (Gütig et al., 2001) still assume strong distribution  
 0362 assumption (such as Poissonian features) that can be avoided by surrogate data meth-  
 0363 ods. Can we show similar Gaussian approximations without such strong distribution  
 0364 assumptions?

0365 Firstly it is possible to estimate  $c_0$  without making any strong distribution assump-  
 0366 tion besides the fact that the trials are assumed to be i.i.d.

0367 Indeed note that

$$0368 \quad c_0 = \mathbb{E}_\perp \left[ \sum_{i=1}^n \varphi(X_i^{1,\perp}, X_i^{2,\perp}) \right] = n \mathbb{E}_\perp [\varphi(X^{1,\perp}, X^{2,\perp})],$$

0369 and that for  $i \neq i'$ , since  $X_i$  is always assumed to be independent of  $X_{i'}$ , the following  
 0370 equality holds

$$0371 \quad \mathbb{E} [\varphi(X_i^1, X_{i'}^2)] = \mathbb{E}_{\perp} [\varphi(X^{1,\perp}, X^{2,\perp})]. \quad (2)$$

0372 Thus,  $c_0$  can always be estimated in a distribution free manner by

$$0373 \quad \hat{C}_0(\mathbb{X}_n) = \frac{1}{n-1} \sum_{i \neq i'} \varphi(X_i^1, X_{i'}^2).$$

0375 The centered quantity of interest, in the sense that it has zero mean under independence,  
 0376 is therefore the difference:

$$0377 \quad \mathbf{U} = \mathbf{U}(\mathbb{X}_n) = \mathbf{C}(\mathbb{X}_n) - \hat{C}_0(\mathbb{X}_n), \quad (3)$$

0378 its observed version being denoted by  $\mathbf{U}^{obs}$ .

0380 The next step is to give the asymptotic distribution of  $\mathbf{U}$  (or a renormalized version  
 0381 of it) without making any distribution assumptions in the same spirit as (1) so that one  
 0382 has access to quantiles and critical values. The main mathematical difficulty is that now  
 0383  $\hat{C}_0(\mathbb{X}_n)$  is random and that therefore  $\mathbf{U}$  is not a simple sum over all the trials, but a sum  
 0384 on all the  $(i, i')$  pairs of trials <sup>1</sup>.

0385 Nevertheless, some asymptotic theorems close in spirit to central limit theorems  
 0386 and proven in (Albert et al., 2015), show that under mild conditions (always satisfied in  
 0387 practice in the present cases) the following convergence result holds:

$$0388 \quad \mathbf{Z}(\mathbb{X}_n^{\perp}) = \frac{\mathbf{U}(\mathbb{X}_n^{\perp})}{\sqrt{n\hat{\sigma}(\mathbb{X}_n^{\perp})}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1), \quad (4)$$

---

0389 <sup>1</sup>Double sum of this kind is usually called U-statistics of order 2. They are not sum of independent  
 0390 variables because of the double sum.

0391

0392 where

$$0393 \hat{\sigma}^2(\mathbb{X}_n) = \frac{4}{n(n-1)(n-2)} \sum_{i,j,k \text{ all different}} h(X_i, X_j)h(X_i, X_k),$$

0394 with

$$0395 h(x, y) = \frac{1}{2} \left[ \varphi(x^1, x^2) + \varphi(y^1, y^2) - \varphi(x^1, y^2) - \varphi(y^1, x^2) \right].$$

0396 This result means that one exactly has a distribution approximation of the same form as  
0397 the one of the toy example (1).

0398 As above for the toy example, denoting by  $\mathbf{Z}^{obs}$  the quantity  $\mathbf{Z}$  computed on the  
0399 observed sample, (4) implies that for some fixed  $\alpha$  in  $(0, 1)$ , the test that consists in  
0400 rejecting  $(H_0)$  when  $\mathbf{Z}^{obs} \geq z_{1-\alpha}$ , is asymptotically of level  $\alpha$ .

0401 Let us look more closely at the quality of the approximation (4) on Figure 2. Clearly,  
0402 one can see that the distribution approximation is good when  $n$  is large ( $n = 200$ ) as  
0403 expected, but not so convincing for small values of  $n$  ( $n = 20$ , or even  $n = 50$ ),  
0404 particularly in the tails of the distributions. However, as it is especially the tails of the  
0405 distributions that are involved in the test through the quantile  $z_{1-\alpha}$ , one can wonder, by  
0406 looking at Figure 2, if it may perform reasonably well in practice with a usual number  
0407 of a few tens of trials.

0408 However, unlike the toy example and in line with what happens in (Tuleau-Malot  
0409 et al., 2014), the fact that we have subtracted a random quantity  $\hat{\mathbf{C}}_0$  to  $\mathbf{C}$  makes the  
0410 approximation not valid for the uncentered quantity  $\mathbf{C}$ , as illustrated below. We cannot  
0411 go back and forth by using the scaling and shifting properties of the Gaussian distribu-  
0412 tions. This is what we call the *centering issue*, problem which is actually completely  
0413 related to the plug-in problem mentioned in (Gütig et al., 2001; Tuleau-Malot et al.,  
0414 2014). Indeed, looking informally at (4), and doing as if  $\hat{\mathbf{C}}_0$  was deterministic as for

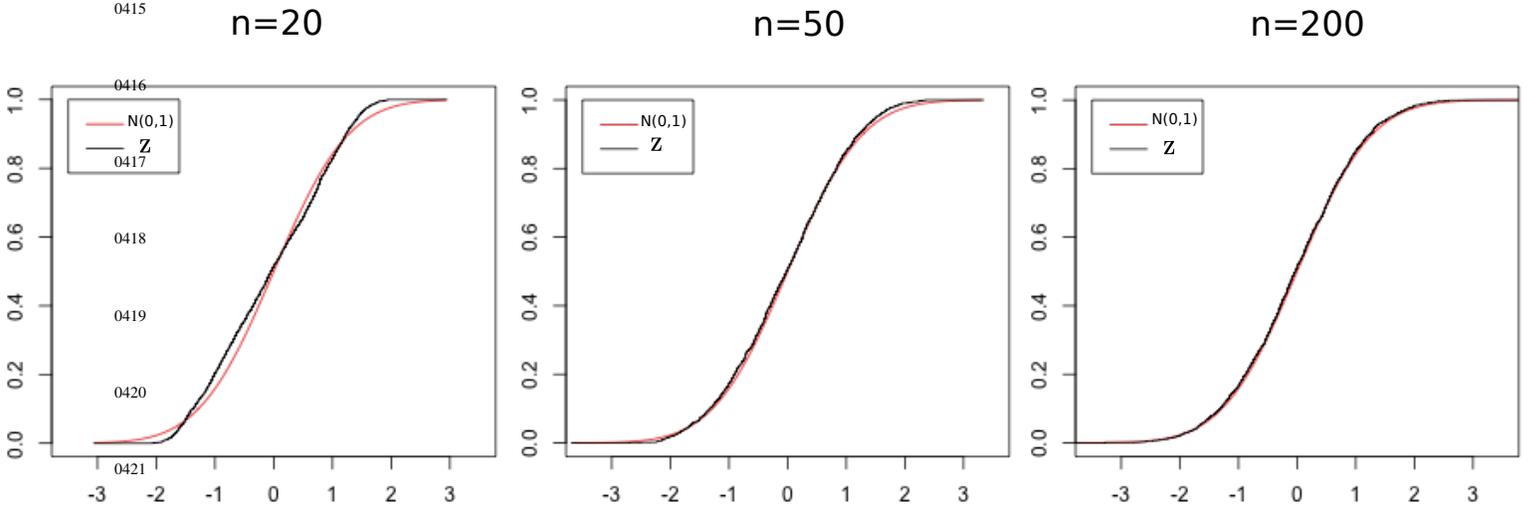


Figure 2: Gaussian approximation of the distribution of  $\mathbf{Z}$ . In plain black, cumulative distribution function (c.d.f.) of  $\mathbf{Z}$  under  $(H_0)$ , that is of  $\mathbf{Z}^\perp = \mathbf{Z}(\mathbb{X}_n^\perp)$  obtained with 2000 simulations of  $\mathbb{X}_n^\perp$ , for  $n = 20, 50$  or  $200$  trials of two independent Poisson processes of firing rate 30Hz, on a window of length 0.1s with  $\delta = 0.01$ s. The red line corresponds to the standard Gaussian c.d.f.

the toy example, if the scaling and shifting properties of the Gaussian distribution were still holding, one could imagine that

$$\mathbf{U}(\mathbb{X}_n^\perp) \underset{n \rightarrow \infty}{\overset{\mathcal{L}}{\approx}} \mathcal{N}(0, n\hat{\sigma}^2(\mathbb{X}_n^\perp)), \quad (5)$$

and

$$\mathbf{C}(\mathbb{X}_n^\perp) \underset{n \rightarrow \infty}{\overset{\mathcal{L}}{\approx}} \mathcal{N}(\hat{\mathbf{C}}_0(\mathbb{X}_n^\perp), n\hat{\sigma}^2(\mathbb{X}_n^\perp)). \quad (6)$$

This is illustrated on Figure 3.

Looking at the first line of Figure 3, one can see that the approximation formulated in (5) is actually conceivable for large values of  $n$ . Note that in practice, one cannot have access to  $\hat{\sigma}^2(\mathbb{X}_n^\perp)$  and it has to be replaced by  $\hat{\sigma}^2(\mathbb{X}_n)$ , meaning that it is computed

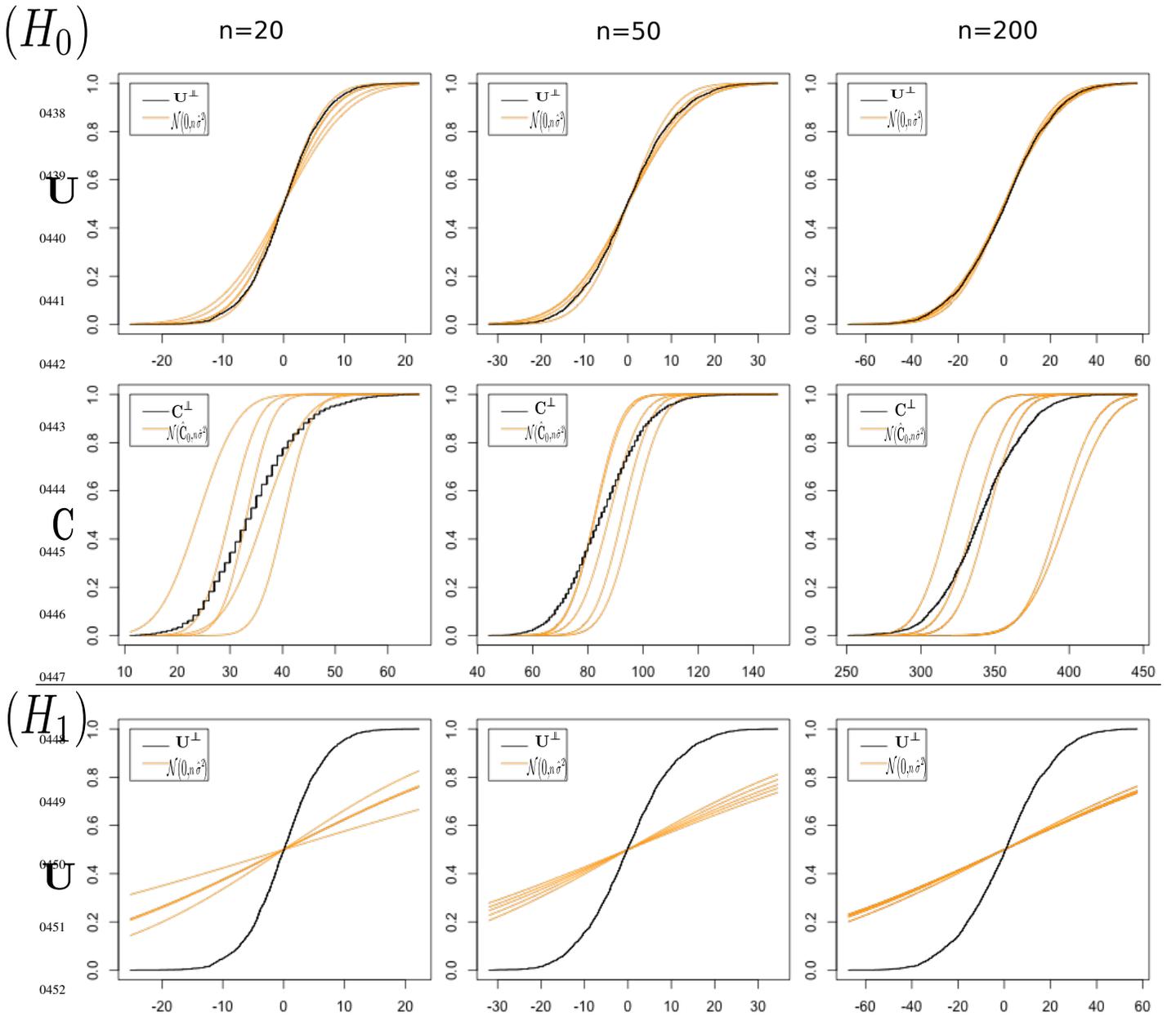


Figure 3: Other Gaussian distribution approximations. Two first lines: c.d.f. of  $U$  and  $C$  under  $(H_0)$ , obtained as in Figure 2. These c.d.f. are respectively compared with the Gaussian c.d.f. with mean 0 and standard deviation  $\sqrt{n}\hat{\sigma}(\mathbb{X}_n)$ , and the Gaussian c.d.f. with mean  $\hat{C}_0(\mathbb{X}_n)$  and standard deviation  $\sqrt{n}\hat{\sigma}(\mathbb{X}_n)$ , for five different simulations of  $\mathbb{X}_n$  under  $(H_0)$ . Third line: c.d.f. of  $U$  under  $(H_0)$  computed as above, compared with the centered Gaussian c.d.f. with standard deviation  $\sqrt{n}\hat{\sigma}(\mathbb{X}_n)$ , for five different simulations of  $\mathbb{X}_n$  under  $(H_1)$  (same marginals as in the first two lines but  $X^1 = X^2$ ).

0461 with the observed sample. This does not change anything under  $(H_0)$  since  $\mathbb{X}_n$  is in this  
 0462 case distributed as  $\mathbb{X}_n^\perp$ . Yet this is a particularly important sticking point if  $(H_0)$  is not  
 0463 satisfied as one can see on the third line of Figure 3: the distribution of  $\mathbf{U}(\mathbb{X}_n^\perp)$  does  
 0464 not look like a centered Gaussian distribution of variance  $n\hat{\sigma}^2(\mathbb{X}_n)$ , when  $\mathbb{X}_n$  does not  
 0465 satisfy  $(H_0)$ .

0466 More importantly for the centering issue, the second line of Figure 3 shows that the  
 0467 approximation formulated in (6) is in fact misleading. To understand why, one needs to  
 0468 take into account the two following points.

0469 (i)  $\hat{\mathbf{C}}_0(\mathbb{X}_n^\perp)$  moves around its expectation  $c_0$  (which is also the expectation of  $\mathbf{C}(\mathbb{X}_n^\perp)$ )  
 0470 with realizations of  $\mathbb{X}_n^\perp$ . These fluctuations have an order of magnitude of  $\sqrt{n}$  and are  
 0471 therefore perfectly observable on the distribution of  $\mathbf{C}(\mathbb{X}_n^\perp)$  whose variance is also of  
 0472 order  $\sqrt{n}$ .

0473 (ii)  $n\hat{\sigma}^2(\mathbb{X}_n^\perp)$  estimates the variance of  $\mathbf{U}(\mathbb{X}_n^\perp)$  and not the one of  $\mathbf{C}(\mathbb{X}_n^\perp)$  or  $\hat{\mathbf{C}}_0(\mathbb{X}_n^\perp)$ .  
 0474 This explains why not only the mean but also the variance are badly estimated in the  
 0475 second line of Figure 3. Two distinct kinds of randomness (the one coming from  $\mathbf{C}(\mathbb{X}_n^\perp)$   
 0476 and the one coming from  $\hat{\mathbf{C}}_0(\mathbb{X}_n^\perp)$ ) have to be taken into account to estimate the variance  
 0477 of  $\mathbf{U}(\mathbb{X}_n^\perp)$ .

0478 As a conclusion of this first naive approach, the test of purely asymptotic nature,  
 0479 which consists in rejecting  $(H_0)$  when  $\mathbf{Z}^{obs} > z_{1-\alpha}$  may work for  $n$  large enough,  
 0480 as the variance is here computed by considering the correctly recentered quantity  $\mathbf{U}$ ,  
 0481 and this even if the behavior of  $\mathbf{U}$  under  $(H_1)$  is not good. However, an ad hoc and  
 0482 more naive test, based on an estimation of the variance of non recentered quantity  $\mathbf{C}$   
 0483 directly and without taking into account the fact that the centering term  $\hat{\mathbf{C}}_0(\mathbb{X}_n)$  is also

0484 random, would not lead to a meaningful test with correct  $p$ -values: this is therefore a  
0485 first example of centering issue.

0486

### 0487 **3.2 The bootstrap approaches**

0488 In statistics, it is well known (Giné, 1997) that tests of purely asymptotic nature as  
0489 the one presented above are less accurate for small values of  $n$  than more involved  
0490 procedures. In this article, the focus is on bootstrap/resampling procedures that are  
0491 usually known to improve the performance from moderate to large sample sizes. Three  
0492 main procedures are investigated that are all three based on a shuffling of the trials: the  
0493 trial-shuffling introduced in (Pipa & Grün, 2003; Pipa et al., 2003), the full bootstrap  
0494 of independence and the permutation approach, the last two being more classical in  
0495 statistics (see e.g. (Romano, 1989)), but also already used on spike train data (see e.g.  
0496 (Ventura, 2010)).

0497 The main common paradigm of these three methods, as described in the sequel,  
0498 is that starting from an observation of the sample  $\mathbb{X}_n$ , they randomly generate, via a  
0499 computer, another sample  $\tilde{\mathbb{X}}_n$ , whose distribution should be close to the distribution of  
0500  $\mathbb{X}_n^\perp$  (see also Figure 4).

0501

0502

0503

0504

0505

0506

### Trial-shuffling

$$\tilde{\mathbb{X}}_n = \mathbb{X}_n^{TS} = ((X_{i^{TS}(1)}^1, X_{j^{TS}(1)}^2), \dots, (X_{i^{TS}(n)}^1, X_{j^{TS}(n)}^2)),$$

where the  $(i^{TS}(k), j^{TS}(k))$ 's are  $n$  i.i.d. couples drawn uniformly at random in  $\{(i, j) / i = 1, \dots, n, j = 1, \dots, n, i \neq j\}$ .

In particular, the corresponding bootstrapped coincidence count is

$$\mathbf{C}^{TS} = \mathbf{C}(\mathbb{X}_n^{TS}) := \sum_{k=1}^n \varphi(X_{i^{TS}(k)}^1, X_{j^{TS}(k)}^2).$$

This algorithm seems natural with respect to (2) because it avoids the diagonal terms of the square  $\{(i, j) / i = 1, \dots, n, j = 1, \dots, n\}$ . Hence as a result,

$$\mathbb{E}[\mathbf{C}^{TS}] = c_0 = \mathbb{E}_\perp[\mathbf{C}].$$

### Classical full bootstrap

$$\tilde{\mathbb{X}}_n = \mathbb{X}_n^* = ((X_{i^*(1)}^1, X_{j^*(1)}^2), \dots, (X_{i^*(n)}^1, X_{j^*(n)}^2)),$$

where the  $n$  couples  $(i^*(k), j^*(k))$  are i.i.d. and where  $i^*(k)$  and  $j^*(k)$  are drawn uniformly and independently at random in  $\{1, \dots, n\}$ .

In particular, the corresponding bootstrapped coincidence count is

$$\mathbf{C}^* = \mathbf{C}(\mathbb{X}_n^*) := \sum_{k=1}^n \varphi(X_{i^*(k)}^1, X_{j^*(k)}^2).$$

Note that this algorithm draws uniformly at random in the square  $\{(i, j) / i = 1, \dots, n, j = 1, \dots, n\}$  and therefore does not avoid the diagonal terms. The idea behind this algorithm is to mimic the independence under  $(H_0)$  of  $X_k^1$  and  $X_k^2$  by drawing the

indexes  $i^*(k)$  and  $j^*(k)$  independently. However

$$\mathbb{E}[\mathbf{C}^*] = n \left( \frac{1}{n} \mathbb{E}[\varphi(X^1, X^2)] + \frac{n-1}{n} \mathbb{E}_{\perp}[\varphi(X^{1,\perp}, X^{2,\perp})] \right).$$

Hence under  $(H_0)$ ,  $\mathbb{E}_{\perp}[\mathbf{C}^*] = c_0$  but, under  $(H_1)$ ,  $\mathbb{E}[\mathbf{C}^*]$  and  $c_0$  are only asymptotically equivalent.

### Permutation

$$\tilde{\mathbb{X}}_n = \mathbb{X}_n^{\Pi_n} = ((X_1^1, X_{\Pi_n(1)}^2), \dots, (X_n^1, X_{\Pi_n(n)}^2)),$$

where  $\Pi_n$  is a permutation drawn uniformly at random in the group of permutations  $\mathfrak{S}_n$  of the set of indexes  $\{1, \dots, n\}$ .

In particular, the corresponding bootstrapped coincidence count is

$$\mathbf{C}^* = \mathbf{C}(\mathbb{X}_n^{\Pi_n}) := \sum_{i=1}^n \varphi(X_i^1, X_{\Pi_n(i)}^2).$$

The idea is to use permutations to avoid picking twice the same spike train of the same trial. In particular under  $(H_0)$ , the sum in  $\mathbf{C}^*$  is still a sum of independent variables, which is not the case in both of the previous algorithms. However, under  $(H_1)$ , the behavior is not as limpid. As for the full bootstrap,

$$\mathbb{E}[\mathbf{C}^*] = n \left( \frac{1}{n} \mathbb{E}[\varphi(X^1, X^2)] + \frac{n-1}{n} \mathbb{E}_{\perp}[\varphi(X^1, X^2)] \right).$$

Hence under  $(H_1)$ ,  $\mathbb{E}[\mathbf{C}^*]$  and  $c_0$  are also only asymptotically equivalent.

To compare those three bootstrap/resampling algorithms, the first thing to wonder is whether, at least under  $(H_0)$ , the introduced extra randomness has not impacted the distribution. More precisely, as stated above, all three procedures satisfy

$$\mathbb{E}_{\perp}[\mathbf{C}(\tilde{\mathbb{X}}_n)] = \mathbb{E}_{\perp}[\mathbf{C}(\mathbb{X}_n^{\perp})] = c_0,$$

0553 but is the full unconditional distribution of  $C(\tilde{X}_n)$  the same as the one of  $C(X_n^\perp)$ ? See  
0554 Figure 4 for a more visual explanation of what is the unconditional distribution.

0555 The first line of Figure 5 shows as expected that the permutation does not change  
0556 the distribution of  $X_n^\perp$ , since, as said above, no spike train is picked twice. However,  
0557 clearly the trial-shuffling and the full bootstrap have not the same property, even if the  
0558 distributions are quite close.

0559 Nevertheless, this is not completely convincing. Indeed and as already mentioned in  
0560 Figure 4, the main particularity of surrogate data procedures is to be able for one current  
0561 observation of  $X_n$  to generate several surrogate data sets, that is several realizations of  
0562  $\tilde{X}_n$ , and to obtain not the unconditional distribution of  $C(\tilde{X}_n)$  but the conditional dis-  
0563 tribution of  $C(\tilde{X}_n)$  given  $X_n$ . What is important to emphasize is that this conditional  
0564 distribution (which is the one to which one has access in practice) actually depends on  
0565 the original data set. This is why on the second line of Figure 5, are given five real-  
0566 izations of the conditional cumulative distribution function: since this is a simulation,  
0567 we are able to produce 5 "original" data sets and to see how the conditional distribu-  
0568 tion fluctuates thanks to the Nature randomness as described in Figure 4. What we can  
0569 expect is that as a proxy, this conditional distribution, which is the only accessible one,  
0570 will be close to the one we would like to know, that is the distribution of  $C^\perp$ .

0571 However none of the three conditional distributions seems to fit the distribution of  
0572  $C(X_n^\perp)$ . One may eventually think that this is due to the Monte-Carlo approximation  
0573 of the conditional distributions, but for the trial-shuffling approach, Pipa and Grün de-  
0574 veloped an algorithm for exact computation of the conditional distribution (Pipa et al.,  
0575 2003): both Monte-Carlo and exact conditional distribution are so close that it is diffi-

0576 cult to make any difference between them.

0577 Hence there should be another explanation. In fact, the curves on the second line  
0578 of Figure 5 are similar to the ones on the second line of Figure 3. In both set-ups, one  
0579 wonders if the distribution of  $C(\mathbb{X}_n^\perp)$  can or cannot be approximated by a distribution  
0580 depending on the observation of  $\mathbb{X}_n$ : a very basic Gaussian distribution for Figure 3 and  
0581 a more intricate distribution using the bootstrap paradigm for Figure 5. In both cases,  
0582 the conditional c.d.f. are widely spread around the aim which is the distribution of  
0583  $C(\mathbb{X}_n^\perp)$ . Since the explanation for Figure 3 was a centering defect that can be corrected  
0584 by considering  $\mathbf{U}$ , the explanation here is a centering defect for the procedures based  
0585 on a shuffling of the trials too, and this can also be corrected as one can see below.

0586

### 0587 **3.3 Which centering for which bootstrap ?**

0588 To understand the centering issue of the procedures based on a shuffling of the trials,  
0589 one needs to understand more precisely the mathematical results on bootstrap.

0590 The precursor work of Bickel and Freedman (Bickel & Freedman, 1981) on the  
0591 bootstrap of the mean can be heuristically explained as follows. Given a  $n$  sample  
0592 of i.i.d. real random variables  $\mathbb{Y}_n = (Y_1, \dots, Y_n)$  with mean  $m$  and a corresponding  
0593 bootstrap sample  $\mathbb{Y}_n^*$ , it is not possible to estimate the distribution of the empirical mean  
0594  $\bar{Y} = (1/n) \sum_{i=1}^n Y_i$  directly. However one can estimate the centered distribution, i.e.  
0595 the distribution of  $\bar{Y} - m = \bar{Y} - \mathbb{E}[\bar{Y}]$ . To do so, it is sufficient to replace "empirical  
0596 mean" by "empirical bootstrap mean" and "expectation" by "conditional expectation".  
0597 More explicitly, denoting by  $\bar{Y}^*$  the empirical mean of the bootstrap sample  $\mathbb{Y}_n^*$ , the  
0598 distribution of  $\bar{Y} - \mathbb{E}[\bar{Y}]$  is approximated by the conditional distribution given  $\mathbb{Y}_n$  of

0599  $\bar{Y}^* - \mathbb{E}[\bar{Y}^* | \mathbb{Y}_n].$

0600 More generally, the bootstrap approaches that have been proved to work from a  
 0601 mathematical point of view are all based on centered quantities (Giné, 1997): this is  
 0602  $\bar{Y} - m$  in the previous example but this can also be centered U-statistics. However, this  
 0603 cannot be  $\mathbf{C}$ , which is not centered, as one can see in Figure 5.

0604 A suitable quantity in our context is  $\mathbf{U}$  given in (3), since it has zero mean under  
 0605  $(H_0)$ . Indeed, by the bootstrap paradigm recalled above, the distribution of  $\mathbf{U}(\mathbb{X}_n)$   
 0606 under  $(H_0)$ , that is of  $\mathbf{U}(\mathbb{X}_n^\perp)$  (which has zero mean), should be well approximated by  
 0607 the distribution of  $\mathbf{U}(\tilde{\mathbb{X}}_n) - \mathbb{E}[\mathbf{U}(\tilde{\mathbb{X}}_n) | \mathbb{X}_n]$ .

0608 For the trial-shuffling, since

0609 
$$\mathbf{U}(\mathbb{X}_n^{TS}) = \sum_{k=1}^n \varphi(X_{i^{TS}(k)}^1, X_{j^{TS}(k)}^2) - \frac{1}{n-1} \sum_{k \neq k'} \varphi(X_{i^{TS}(k)}^1, X_{j^{TS}(k')}^2),$$

0610 one can easily see that because the couple  $(i^{TS}(k), j^{TS}(k))$  is drawn uniformly at ran-  
 0611 dom in the set of the  $(i, j)$ 's such that  $i \neq j$  (set of cardinality  $n(n-1)$ ),

0612 
$$\begin{aligned} \mathbb{E}[\mathbf{U}(\mathbb{X}_n^{TS}) | \mathbb{X}_n] &= \frac{1}{n-1} \sum_{i \neq j} \varphi(X_i^1, X_j^2) - \frac{1}{n} \sum_{i,j} \varphi(X_i^1, X_j^2) \\ 0613 &= \frac{\hat{\mathbf{C}}_0(\mathbb{X}_n) - \mathbf{C}(\mathbb{X}_n)}{n} \\ 0614 &= -\frac{\mathbf{U}(\mathbb{X}_n)}{n}. \end{aligned}$$

0615 Hence the quantity that needs to be computed on the surrogate data set when applying  
 0616 the trial-shuffling method is

0617 
$$\tilde{\mathbf{U}}^{TS} = \tilde{\mathbf{U}}(\mathbb{X}_n^{TS}) = \mathbf{U}(\mathbb{X}_n^{TS}) + \frac{\mathbf{U}(\mathbb{X}_n)}{n}.$$

0618 Furthermore, similar computations show that the full bootstrap and the permutation  
 0619 satisfy  
 0620

0621 
$$\mathbb{E}[\mathbf{U}(\mathbb{X}_n^*) | \mathbb{X}_n] = \mathbb{E}[\mathbf{U}(\mathbb{X}_n^{\Pi_n}) | \mathbb{X}_n] = 0.$$

0622 Hence,  $\mathbf{U}(\mathbb{X}_n^*)$  and  $\mathbf{U}(\mathbb{X}_n^{\Pi_n})$  can be computed directly on the surrogate data sets when  
 0623 applying either the Full Bootstrap or the Permutation methods.

0624 Figure 6 shows the quality of approximation of the distribution of  $\mathbf{U}(\mathbb{X}_n^{\perp})$  by the  
 0625 conditional distribution given the observation of either  $\tilde{\mathbf{U}}^{TS}$ ,  $\mathbf{U}^{TS} = \mathbf{U}(\mathbb{X}_n^{TS})$ ,  $\mathbf{U}^* =$   
 0626  $\mathbf{U}(\mathbb{X}_n^*)$  or  $\mathbf{U}^* = \mathbf{U}(\mathbb{X}_n^{\Pi_n})$ . Contrary to Figure 5, the conditional distributions of  $\mathbf{U}^*$  and  
 0627  $\mathbf{U}^*$  do not spread widely around the target distribution but are accurate approximations  
 0628 not only under  $(H_0)$  but even if the observed sample is simulated under  $(H_1)$ , which is  
 0629 in complete accordance with the mathematical results of consistence proven in (Albert  
 0630 et al., 2015). The approximation is just as accurate when using the recentered quantity  
 0631  $\tilde{\mathbf{U}}^{TS}$  to mimic the distribution of  $\mathbf{U}^{\perp}$  but it is not for simply  $\mathbf{U}^{TS}$ , the difference between  
 0632 the conditional c.d.f. of  $\tilde{\mathbf{U}}^{TS}$  and the one of  $\mathbf{U}^{TS}$  being particularly visible under  $(H_1)$   
 0633 when  $X^1 = X^2$ . This means that one definitely need to recenter the quantities that are  
 0634 computed on the surrogate data set by subtracting their conditional expectation given  
 0635 the original data set to obtain a correct fit of the desired centered distribution under  
 0636 independence.

0637 Hence, as explained by the computations above, in a trial-shuffling approach, the  
 0638 correctly recentered version leads to the correct bootstrap distribution. Note finally that  
 0639 this corroborates the previous intuition: the reason why the approximation works for  
 0640  $\mathbf{U}$  and not for  $\mathbf{C}$  is a *centering issue*, that is exactly the same as for the first approach  
 0641 of Figure 3. The centering is indeed random as in Figure 3 (here it can be viewed as  
 0642  $\mathbb{E}[\mathbf{C}(\tilde{\mathbb{X}}_n)|\mathbb{X}_n]$ ) and one needs to take it into account to have a correct approximation.

0643 Finally an extra simplification holds in the permutation case, which may seem very  
 0644 surprising.

One can easily rewrite on the one hand,

$$\mathbf{U}(\mathbb{X}_n) = \left(1 + \frac{1}{n-1}\right) \mathbf{C}(\mathbb{X}_n) - \frac{1}{n-1} \sum_{i,j} \varphi(X_i^1, X_j^2)$$

and, on the other hand, for the permutation sample

$$\mathbf{U}(\mathbb{X}_n^{\Pi_n}) = \left(1 + \frac{1}{n-1}\right) \mathbf{C}(\mathbb{X}_n^{\Pi_n}) - \frac{1}{n-1} \sum_{i,j} \varphi(X_i^1, X_j^2).$$

Note that the sum  $\sum_{i,j} \varphi(X_i^1, X_j^2)$  is invariant by the action of the permutation. Hence if  $u_t^*$  denotes the quantile of order  $t$  of the conditional distribution of  $\mathbf{U}(\mathbb{X}_n^{\Pi_n})$  given  $\mathbb{X}_n$  and if  $c_t^*$  denotes the quantile of order  $t$  of the conditional distribution of  $\mathbf{C}(\mathbb{X}_n^{\Pi_n})$  given  $\mathbb{X}_n$ , this very simple relationship holds

$$u_t^* = \left(1 + \frac{1}{n-1}\right) c_t^* - \frac{1}{n-1} \sum_{i,j} \varphi(X_i^1, X_j^2).$$

Hence the test that rejects  $(H_0)$  when  $\mathbf{U}(\mathbb{X}_n) > u_{1-\alpha}^*$  is exactly the one that rejects  $(H_0)$  when  $\mathbf{C}(\mathbb{X}_n) > c_{1-\alpha}^*$ . Therefore despite the fact that the conditional distribution of  $\mathbf{C}(\mathbb{X}_n^{\Pi_n})$  is not close at all to the one of  $\mathbf{C}(\mathbb{X}_n^{\perp})$ , the test based on  $\mathbf{C}$  works, because it is equivalent to the test based on  $\mathbf{U}$ , for which the approximation of the conditional distribution works. Note however that this phenomenon happens only in the permutation approach, but not in the trial-shuffling or the full bootstrap approaches.

### 3.4 Practical testing procedures and $p$ -values

From the considerations given above, five different tests may be investigated, the first one based on a purely asymptotic approach, and the four other ones based on resampling approaches, with critical values approximated through a Monte-Carlo method. For each test, the corresponding  $p$ -values (i.e. the values of  $\alpha$  for which the test passes from acceptance to rejection) are given.

0668 **The naive test (N).** It consists in rejecting ( $H_0$ ) when

$$0669 \mathbf{Z}^{obs} \geq z_{1-\alpha}.$$

0670  
0671 The corresponding  $p$ -value is given by:

$$0672 1 - \Phi(\mathbf{Z}^{obs}),$$

0673 where  $\Phi$  is the c.d.f. of a standard Gaussian distribution.

0674  
0675 **The Trial-Shuffling test, version C (TSC).** It consists in rejecting ( $H_0$ ) when

$$0676 \mathbf{C}^{obs} \geq \hat{c}_{1-\alpha}^{TS},$$

0677  
0678 where  $\hat{c}_{1-\alpha}^{TS}$  is the empirical quantile of order  $(1 - \alpha)$  of the conditional distribution  
0679 of  $\mathbf{C}^{TS}$  given  $\mathbb{X}_n$ . This empirical quantile is estimated over  $B$  ( $B = 10000$  usually)  
0680 realizations  $\mathbf{C}_1^{TS}, \dots, \mathbf{C}_B^{TS}$  given the observed sample  $\mathbb{X}_n$ . The corresponding  $p$ -value is  
0681 given by:

$$0682 \frac{1}{B} \sum_{i=1}^B \mathbf{1}_{\mathbf{C}_i^{TS} \geq \mathbf{C}^{obs}}.$$

0683 Despite the centering defect of this method underlined in Section 3.3, we kept this test  
0684 in the present study since it corresponds to the one programmed in (Pipa & Grün, 2003)  
0685 and since it is widely applied in the neuroscience literature.

0686 **The Trial-Shuffling test, version recentered U (TSU).** It consists in rejecting ( $H_0$ )

0687 when

$$0688 \mathbf{U}^{obs} \geq \hat{w}_{1-\alpha}^{TS},$$

0689  
0690 where  $\hat{w}_{1-\alpha}^{TS}$  is the empirical quantile of order  $(1 - \alpha)$  of the conditional distribution  
of  $\tilde{\mathbf{U}}^{TS}$  (the correctly recentered quantity) given  $\mathbb{X}_n$ . This empirical quantile and the

0691 corresponding  $p$ -value are obtained in away similar to the above (TSC), based on  $B$   
 0692 realizations  $\tilde{U}_1^{TS}, \dots, \tilde{U}_B^{TS}$  of  $\tilde{U}(\mathbb{X}_n^{TS})$  given  $\mathbb{X}_n$ .

0693 **The Full Bootstrap test, version U (FBU).** It consists in rejecting ( $H_0$ ) when  
 0694

$$0695 \mathbf{U}^{obs} \geq \hat{u}_{1-\alpha}^*,$$

0696 where  $\hat{u}_{1-\alpha}^*$  is the empirical quantile of order  $(1 - \alpha)$  of the conditional distribution of  
 0697  $\mathbf{U}^*$  given  $\mathbb{X}_n$ . This empirical quantile and the corresponding  $p$ -value are obtained in a  
 0698 way similar to the above (TSC), based on  $B$  realizations  $\mathbf{U}_1^*, \dots, \mathbf{U}_B^*$  of  $\mathbf{U}(\mathbb{X}_n^*)$  given  
 0699  $\mathbb{X}_n$ .

0700  
 0701 **The permutation test (P).** The reader may think that it should consist in rejecting  
 0702 ( $H_0$ ) when

$$0703 \mathbf{C}^{obs} \geq \hat{c}_{1-\alpha}^*,$$

0704 where  $\hat{c}_{1-\alpha}^*$  is the empirical quantile of order  $(1 - \alpha)$  of the conditional distribution of  
 0705  $\mathbf{C}^*$  given  $\mathbb{X}_n$ . Yet the test by permutation is in fact directly defined by its  $p$ -value, which  
 0706 is slightly different here, equal to:

$$0707 \frac{1}{B+1} \left( 1 + \sum_{i=1}^B \mathbf{1}_{\mathbf{C}_i^* \geq \mathbf{C}^{obs}} \right),$$

0708  
 0709 where  $\mathbf{C}_1^*, \dots, \mathbf{C}_B^*$  are  $B$  realizations of  $\mathbf{C}(\mathbb{X}_n^{\Pi_n})$  given  $\mathbb{X}_n$ . The permutation test then  
 0710 consists in rejecting ( $H_0$ ) when this  $p$ -value is less than  $\alpha$ . Indeed, such a permutation  
 0711 test, with such a slightly different version of  $p$ -value, has been proved to be exactly of  
 0712 level  $\alpha$ , whatever  $B$  (Romano & Wolf , 2005), thanks to exchangeability properties of  
 0713 random permutations.

0714 Note however that such a slight correction does not work for full bootstrap or  
0715 trial-shuffling approaches, where the tests are only guaranteed to be asymptotically of  
0716 level  $\alpha$ .

0717 Saying that a test rejects at level  $\alpha$  (or that its False Positive (FP) rate is smaller than  
0718  $\alpha$ ) is exactly equivalent to saying that its  $p$ -value is less than  $\alpha$ . If a test is of level  $\alpha$   
0719 for any  $\alpha$  in  $(0, 1)$ , the c.d.f. of its  $p$ -values should therefore be smaller than the one  
0720 of a uniform variable (i.e. the diagonal) under  $(H_0)$ . Between several tests with this  
0721 guarantee, the less conservative one is the one for which the c.d.f of its  $p$ -values is the  
0722 closest to the diagonal. The left hand-side of Figure 7 shows the c.d.f. under  $(H_0)$  of  
0723 the corresponding  $p$ -values for the five considered testing procedures and focuses on  
0724 small  $p$ -values, which are the only ones usually involved in testing, to highlight the  
0725 main differences between the five methods. For the chosen small value of  $n$  ( $n = 20$ ),  
0726 the c.d.f. of the (TSU) and (FBU)  $p$ -values are almost identical and above the diagonal,  
0727 meaning that the corresponding tests do not guarantee the level. On the contrary, the  
0728 c.d.f. of the (N) and (TSC)  $p$ -values are clearly under the diagonal and far from it,  
0729 meaning that the corresponding tests are too conservative. As guaranteed by (Romano  
0730 & Wolf , 2005), the permutation approach guarantees the level of the test: the c.d.f. of  
0731 the (P)  $p$ -values is also under the diagonal, under  $(H_0)$ , but much closer to the diagonal  
0732 than the one of the (N) and (TSC)  $p$ -values.

0733 Furthermore, the behavior of the c.d.f. of the  $p$ -values under  $(H_1)$  gives an indi-  
0734 cation of the power of the test. Indeed this c.d.f associates to each  $\alpha$  in  $(0, 1)$ , the  
0735 (estimated) probability that the test rejecting independence when its  $p$ -value is less than  
0736  $\alpha$ , actually rejects independence. This probability is, under  $(H_1)$ , 1 minus the False

0737 Negative (FN) rate. It can also be seen as the power of the test. Hence among the tests  
0738 that guarantee the level, the permutation test (P) is the one with the smallest FN rate,  
0739 that is the most powerful one.

0740 Note that other simulations in more various cases have been performed in (Albert  
0741 et al., 2015) leading to the same conclusion.

0742 We have also performed some simulations for which the firing rate is not constant  
0743 across the trials. The results are displayed on Figure 8. It is important to note that the  
0744 independence is rejected (that is when the  $p$ -values are clearly small, with a c.d.f. clearly  
0745 above the diagonal) only when the rates of each components  $X^1$  and  $X^2$  progress in a  
0746 coordinate way (see Figure 8.A). If only one rate varies (see Figure 8.B), the  $p$ -values  
0747 are all close to diagonal except for (TSC), for which the distribution approximation does  
0748 not work as we showed above. The same appears in the set-up considered by (Grün  
0749 et al, 2003; Grün, 2009) (see Figure 8.C) with  $p$ -values even closer to the diagonal,  
0750 because the number of trials is larger. Note that this set-up was given in (Grün, 2009)  
0751 as the worst case scenario of non-stationarity across trials for the trial-shuffling method  
0752 and it was stated that this is due to a violation of the underlying assumption of non-  
0753 stationarity across trials. However, as shown by Figure 8 in those two last situations,  
0754 the  $p$ -values behave as under  $(H_0)$ , except maybe for (TSC), and we believe that this  
0755 is explained not by a violation of the i.i.d. assumption on the trials but by a centering  
0756 defect, as explained above. As announced in the introduction, cross-trials stationarity  
0757 is not equivalent to the i.i.d. assumptions on the trials and this explains also why the  
0758 correctly centered bootstrap methods work in this non-stationary case. We discuss in  
0759 more detail this behavior in the discussion (see Section 5).

0760 In the sequel, since the permutation method is the only one able to guarantee the  
0761 level of the test (that is to control the (FP) rate) even for a very small number of  
0762 observation (see Figure 7), we focus on the permutation approach, keeping also the  
0763 trial-shuffling version **C** approach, denoted by (TSC) on the graphs, as a variant of the  
0764 method developed in (Pipa & Grün, 2003).

## 0766 **4 Permutation UE**

### 0768 **4.1 Description of the complete multiple testing algorithm**

0769 To detect precise locations of dependence periods that can be matched to some ex-  
0770 perimental or behavioral events, the third step (point (iii) of the introduction) of a UE  
0771 method is classically to consider a family of windows  $\mathcal{W}$  of cardinal  $K$ , which is a col-  
0772 lection of potentially overlapping intervals  $[a, b]$  covering the whole interval  $[0, T]$  on  
0773 which trials have been recorded (Grün et al., 1999; Tuleau-Malot et al., 2014). Then,  
0774 some independence tests are implemented on each window of the collection. Here we  
0775 propose a complete algorithm which takes into account the multiplicity of the tests, and  
0776 which moreover enables to see if the coincidence count is significantly too large or too  
0777 small on each window as in (Tuleau-Malot et al., 2014).

### Permutation UE algorithm

Fix a real number  $q$  in  $(0, 0.5)$  and an integer  $B$  larger than 2.

- Do in parallel for each window  $W = [a, b]$  in  $\mathcal{W}$ :

\* Extract the points of the  $X_i^1$ 's and  $X_i^2$ 's in  $[a, b]$ .

\* For all  $(i, j)$  in  $\{1, \dots, n\}^2$ , compute  $a_{i,j} = \varphi_\delta^{coinc}(X_i^1, X_j^2)$  over  $[a, b]$

by the **delayed coincidence count algorithm**.

\* Draw at random  $B$  i.i.d. permutations  $\Pi_n^b$ ,  $1 \leq b \leq B$ , and compute  $C^b = \sum_i a_{i, \Pi_n^b(i)}$ .

(There is one full new set of  $B$  permutations for each window.)

\* Compute also  $C^{obs} = \sum_i a_{i,i}$ .

\* Return  $p_W^+ = \frac{1}{B+1} \left( 1 + \sum_{b=1}^B \mathbf{1}_{C^b \geq C^{obs}} \right)$  and  $p_W^- = \frac{1}{B+1} \left( 1 + \sum_{b=1}^B \mathbf{1}_{C^b \leq C^{obs}} \right)$ .

- Perform the BH procedure of (Benjamini & Hochberg, 1995) on the set of the above  $2K$   $p$ -values:

\* Sort the  $p$ -values  $p^{(1)} \leq \dots \leq p^{(2K)}$ .

\* Find  $k = \max\{l / p^{(l)} \leq lq / (2K)\}$ .

\* Return all the  $(W, \epsilon_W)$ 's, for which  $W$  is associated with one of the  $p$ -values  $p^{(l)}$  for  $l \leq k$ ,

with  $\epsilon_W = 1$  if  $p_W^+ \leq p^{(k)}$ , so the coincidence count is significantly too large on  $W$ ,

and  $\epsilon_W = -1$  if  $p_W^- \leq p^{(k)}$ , so the coincidence count is significantly too small on  $W$ .

The code has been parallelized in C++ and interfaced with R. The full corresponding

R-package is still a work in progress but actual codes are available at

<https://github.com/ybouret/neuro-stat>.

This algorithm corresponds to a slight variation of the multiple testing step of (Tuleau-

0806 Malot et al., 2014), but adapted to non necessarily symmetric distributions <sup>2</sup>. In several  
0807 applications, neuroscientists are interested in detecting dependence periods for which  
0808 the coincidence count is only significantly too large. In this case, one can use the re-  
0809 stricted set of the  $p_W^+$ 's.

0810 From a mathematical point of view, if the considered windows are disjoint and if  
0811 the spike trains are Poisson processes that are non necessarily stationary, the False Dis-  
0812 covery Rate (FDR) <sup>3</sup> of the above multiple testing procedure can be mathematically  
0813 proven<sup>4</sup> to be controlled by  $q$  for any  $B \geq 2$ . The problem of mathematically proving  
0814 that BH procedure guarantees an FDR smaller than  $q$  without those restrictions is very  
0815 difficult even in simple situations such as the Gaussian regression framework (Ben-  
0816 jamini & Yekutieli, 2001), while it is usually observed in practice that the FDR is still  
0817 controlled by  $q$ . However it has been proved in (Benjamini & Yekutieli, 2001) that for  
0818 any framework and therefore in particular for the most general setting of Permutation  
0819 UE, the FDR of BH is always smaller than  $q \sum_{\ell=1}^{2K} \ell^{-1} \simeq q \ln(2K)$ . Hence, for in-  
0820 stance, with 50 windows and  $q = 0.01$ , we are still mathematically guaranteeing that  
0821 the FDR of Permutation UE described in the above sidebar is whatever the underlying  
0822 distribution controlled by 0.052. Moreover the distributions that are reaching this rate  
0823 are so particular that it is often advised even by mathematical experts of multiple testing

---

0824 <sup>2</sup>Note in particular that for a fixed  $W$ , one cannot have both  $p_W^+ < 0.5$  and  $p_W^- < 0.5$  and therefore,  
0825 if a  $W$  is detected, it can only be because of one of the two situations,  $p_W^+ \leq p^{(k)}$  or  $p_W^- \leq p^{(k)}$ , which  
cannot happen simultaneously.

0826 <sup>3</sup>see (Tuleau-Malot et al., 2014) or Table 1 for a precise definition

0827 <sup>4</sup>The  $p_W^+$ 's are independent random variables such that  $\mathbb{P}_{\perp}(p_W^+ \leq \alpha) \leq \alpha$  for all  $\alpha$  in  $[0, 1]$  (Benjamini  
0828 & Yekutieli, 2001; Romano & Wolf, 2005).

0829 to do as if the control of the FDR by  $q$  holds as soon as typical simulations do not show  
0830 otherwise.

## 0831 **4.2 Comparison on simulations**

0832  
0833 Two sets of simulations have been carried out. The first one, namely Experiment 1,  
0834 combines different point processes encountered in the literature (homogeneous, and  
0835 inhomogeneous Poisson processes, Hawkes processes), and different kinds of depen-  
0836 dences. It is described in Figure 9.A. The second one, namely Experiment 2, consists  
0837 of simple independent homogeneous Poisson processes on the whole interval  $[0, 2]$ , as  
0838 described in Table 1. The corresponding results are described in Table 1 and one run of  
0839 simulation of the Permutation UE method is presented in Figure 9. Four methods have  
0840 been compared:

- 0841 • the MTGAUE method of Tuleau-Malot et al. (2014) which assumes both pro-  
0842 cesses to be homogeneous Poisson processes, with  $q = 0.05$ ,
- 0843  
0844 • the Trial-Shuffling, version C (TSC) which corresponds to the method of Pipa &  
0845 Grün (2003), which has been programmed with the delayed coincidence count  
0846 described above and which has not been corrected for multiplicity, that is with  
0847 level  $\alpha = 0.05$  on all windows,
- 0848 • the same as above but corrected by Benjamini and Hochberg procedure (TSC +  
0849 BH), that is with  $q = 0.05$ ,
- 0850 • the Permutation UE approach described above, with  $q = 0.05$ .

0852 In Figure 9.B, several  $\delta$ , that is several delays for the delayed coincidence counts, have  
0853 been tested and each line corresponds to a different value of  $\delta$ . We see that except for  
0854 very few false positives, the method is able to detect the correct dependence features  
0855 and that it is also able to distinguish between situations where there are too many co-  
0856 incidences (bands delimited by plain black lines and containing red crosses) or too few  
0857 coincidences (bands delimited by dotted black lines and containing blue crosses), the  
0858 bands being what should be detected and the crosses being what is indeed detected in  
0859 the simulation. Moreover one sees that even if there are some variations, detections  
0860 occur for all reasonable values of  $\delta$ .

0861 The permutation approach always guarantees an FDR less than the prescribed level  
0862 of 0.05 whereas MTGAUE does not when the homogeneous Poisson assumption fails  
0863 (Experiment 1). The classical trial-shuffling method (where dependence detection oc-  
0864 curs each time the  $p$ -value is less than 0.05) seems to have comparable results in terms  
0865 of both FDR and False Non Discovery Rate (FNDR) on Experiment 1 but fails to con-  
0866 trol the FDR on the most basic situation, namely purely independent processes (Ex-  
0867 periment 2). Adding a Benjamini-Hochberg step of selection of  $p$ -values to the trial-  
0868 shuffling makes it more robust but at the price of a much larger FNDR with respect to  
0869 the Permutation UE method, a fact which is consistent with the conservativeness shown  
0870 in Figure 7.

0871

### 0872 **4.3 Comparison on real data**

0873 **Behavioral procedure.** The data used in this theoretical article to test the dependence  
0874 detection ability of the four methods were already partially published in previous ex-

	Independ.	Depend.	Total
Rejected	$V$	$S$	$R$
Accepted	$U$	$T$	$m - R$
Total	$m_0$	$m - m_0$	$m$

	Experiment 1		Experiment 2	
	FDR	FNDR	FDR	FNDR
MTGAUE	0.10	0.17	0.04	0
TSC	0.01	0.26	0.25	0
TSC + BH	0	0.32	0	0
P	0.01	0.23	0.02	0

Table 1: False Discovery and Non Discovery Rates. On the left hand-side, the classical table for multiple testing adapted to our dependence framework, with a total number of tests  $m = 2K$ . On the right hand-side, estimated FDR and FNDR over 1000 runs, FDR being defined by  $\mathbb{E}[(V/R)\mathbf{1}_{R>0}]$  and FNDR being defined by  $\mathbb{E}[(T/(m - R))\mathbf{1}_{m-R>0}]$ . Experiment 1 is described in Figure 9, Experiment 2 consists of two independent homogeneous Poisson processes of firing rate 60 Hz on  $[0, 2]$ . The set of windows is as in Figure 9. There are 50 trials and  $\delta = 0.01$ s. MTGAUE is the method described in (Tuleau-Malot et al., 2014) with  $q = 0.05$ . (TSC) is the trial-shuffling method with Monte-Carlo approximation ( $B = 10000$ ) and the selected windows are the ones whose  $p$ -value are less than 0.05. (TSC+BH) is the same method, except that the multiplicity of the tests is corrected by a Benjamini-Hochberg procedure ( $q = 0.05$ ). (P) corresponds to the Permutation UE method ( $B = 10000$ ,  $q = 0.05$ ).

experimental studies (Riehle et al., 2000; Grammont & Riehle, 2003; Riehle et al., 2006) and also used in (Tuleau-Malot et al., 2014). These data were collected on a 5-year-old male Rhesus monkey who was trained to perform a delayed multidirectional pointing task. The animal sat in a primate chair in front of a vertical panel on which seven touch-sensitive light-emitting diodes were mounted, one in the center and six placed

0898 equidistantly (60 degrees apart) on a circle around it. The monkey had to initiate a trial  
0899 by touching and then holding with the left hand the central target. After a fix delay of  
0900 500ms, the preparatory signal (PS) was presented by illuminating one of the six periph-  
0901 eral targets in green. After a delay of either 600ms or 1200ms, selected at random with  
0902 various probability, it turned red, serving as the response signal and pointing target.  
0903 During the first part of the delay, the probability  $p_{resp}$  for the response signal to occur at  
0904  $(500 + 600)\text{ms} = 1.1\text{s}$  was 0.3. Once this moment passed without signal occurrence, the  
0905 conditional probability for the signal to occur at  $(500 + 600 + 600)\text{ms} = 1.7\text{s}$  changed  
0906 to 1. The monkey was rewarded by a drop of juice after each correct trial. Reaction  
0907 time (RT) was defined as the release of the central target. Movement time (MT) was  
0908 defined as the touching of the correct peripheral target.

0909 **Recording technique.** Signals recorded from up to seven microelectrodes (quartz in-  
0910 sulated platinum-tungsten electrodes, impedance:  $2 - 5\text{M}\Omega$  at  $1000\text{Hz}$ ) were amplified  
0911 and band-pass filtered from  $300\text{Hz}$  to  $10\text{kHz}$ . Using a window discriminator, spikes  
0912 from only one single neuron per electrode were then isolated. Neuronal data along with  
0913 behavioral events (occurrences of signals and performance of the animal) were stored  
0914 on a PC for off-line analysis with a time resolution of  $10\text{kHz}$ .

0915 In the following study, only trials where the response signal (RS) occurs at  $1.7\text{s}$  are  
0916 considered. The expected signal (ES) corresponds to an eventually expected but not  
0917 confirmed signal, i.e. at  $1.2\text{s}$ . Pairs 13 and 40 of the data set are considered here, as they  
0918 were already treated in (Tuleau-Malot et al., 2014) and in (Riehle et al., 2000) with the  
0919 Multiple Shift method (MS) of (Grün et al., 1999). This last analysis is also displayed  
0920

0921 on Figures 10 and 11 together with the methods described in the present article.

0922 The (TSC+BH) method does not detect anything and is therefore not presented.  
0923 The Permutation UE (P) method detects less windows than the (MTGAUE), (TSC) and  
0924 (MS) methods. The above simulation study let us think that the extra detections of both  
0925 (MTGAUE) and (TSC) may be False Positives, since both methods do not control the  
0926 FDR as well as the Permutation UE method. However, the windows that are detected  
0927 by the Permutation UE (P) method are still in adequation with the experimental or  
0928 behavioral events. In particular, they still appear around the expected signal (ES) (blue  
0929 vertical bar), which is completely coherent with the analysis made in (Riehle et al.,  
0930 2000). Moreover (see Figure 11) the Permutation UE (P) method is able to detect also  
0931 significant lack of coincidences as the original (MS) method. In Figure 11, there are  
0932 also some windows that are detected by (P) but not by (TSC): this is also coherent with  
0933 the simulations of Figure 7 showing that (TSC) is too conservative and may have as  
0934 well too many false negatives.

## 0936 **5 Discussion**

0937 A UE method can be summarized in three steps:  
0938

- 0939 (i) choose a coincidence count,
- 0940 (ii) choose an approximation of the distribution of this count (or a function of this  
0941 count) under independence to find correct  $p$ -values (in the sense that the corre-  
0942 sponding tests control their False Positive (FP) rates),  
0943 (iii) combine the  $p$ -values for multiple testing on sliding windows.

0944 Our contribution to the steps (i) and (iii) is rather minor. As for the step (i), we  
0945 indeed choose to use the delayed coincidence count introduced in (Tuleau-Malot et al.,  
0946 2014), since it does not suffer from loss in synchrony detection. We here provide a fast  
0947 and efficient algorithm to compute it with an even better run time than a basic algorithm  
0948 for the binned coincidence count, using the sparsity of the signal. As for the step (iii),  
0949 we straightforwardly adapt what has been proposed in (Tuleau-Malot et al., 2014). Note  
0950 that up to an eventual logarithmic correction, this procedure mathematically guarantees  
0951 a control of the False Discovery Rate as soon as the  $p$ -values are correct.

0952 Our main contribution consists in a careful analysis of what has to be done to obtain  
0953 correct  $p$ -values, that is of the step (ii).

0954 **A distribution free procedure.** In this work, the only assumption that is made to  
0955 obtain correct  $p$ -values is that the trials are independent and identically distributed. In  
0956 particular no assumption is made on the underlying distribution of the spike trains.  
0957 They can of course be homogeneous or inhomogeneous Poisson processes as shown in  
0958 Figure 9 and Table 1. They can also be renewal processes in current or operational time  
0959 (Nawrot, 2010; Pipa et al., 2013), have a conditional intensity as, for instance, Hawkes  
0960 processes - see, for instance, Figure 9 and Table 1 or the simulations performed in  
0961 (Albert et al., 2015) - or Wold processes (Pouzat & Chaffiol, 2009), or they can have  
0962 even more complicated structure of dependence with respect to their history - see, for  
0963 instance, (Farkhooi et al., 2009, 2011). In fact they can be whatever one wants as long as  
0964 one assumes that the distribution of the point processes is the same across the trials and  
0965 that there is independence between the trials. All the mathematical material proving that  
0966

0967 they indeed can be "whatever" that is biologically reasonable, is contained in (Albert  
0968 et al., 2015), where it has been shown that all the neuroscience models that we know are  
0969 indeed satisfying the technical assumptions hidden behind those mathematical results.  
0970 In short and for a non mathematical reader, it amounts to assume that each individual  
0971 point process modeling a spike train cannot explode and cannot produce a gigantic  
0972 number of spikes per unit of time, assumption which is always satisfied in practice  
0973 thanks to biological constraints.

0974 **The centering issue.** Under this i.i.d. assumption, we have focused on two distinct  
0975 quantities: either  $C$ , the total number of coincidence, whose expectation  $c_0$  under inde-  
0976 pendence ( $H_0$ ) is not known, or  $U$ , a recentered count, which is obtained by subtracting  
0977 to  $C$  an estimate  $\hat{C}_0$  of the unknown expectation under ( $H_0$ ) and which is therefore of  
0978 zero mean under ( $H_0$ ). We have shown that, because we subtracted a random quantity,  
0979 namely  $\hat{C}_0$ , it is possible to obtain accurate approximation of the distribution of  $U$ , the  
0980 centered quantity, but the approximation does not hold for  $C$ , the non centered one:  
0981 this is the centering issue described in Section 3. In particular, the bootstrap principle,  
0982 which is at the root of several surrogate data methods based on a shuffling of the trials,  
0983 cannot be applied to non centered quantities It is therefore possible to see that the trial-  
0984 shuffling method introduced by (Pipa & Grün, 2003; Pipa et al., 2003) performs poorly  
0985 when directly applied to  $C$  (TSC) but that it very accurately approximates the desired  
0986 distribution once  $C$  is correctly centered (TSU) (see, for instance, Figure 7). The same  
0987 behavior is pointed out for the full bootstrap method, which is more classical from a  
0988 statistical point of view. The permutation method is slightly better in the sense that,  
0989

0990 on the one hand, it does not suffer from the centering issue since the test based on the  
0991 centered quantity  $\mathbf{U}$  and the test based on the non centered quantity  $\mathbf{C}$  are equivalent,  
0992 and on the other hand, it is possible to mathematically guarantee the level whatever  
0993 the number of trials (see the left part of Figure 7 where the corresponding  $p$ -values are  
0994 under the diagonal even for  $n = 20$  trials). This is why we chose the permutation to  
0995 complete the step (ii) of the UE method introduced in this article.

0996 **Practical implementation.** Note that we used a Monte-Carlo approximation of the  
0997 distribution in the provided complete algorithm, which has first been programmed and  
0998 parallelized in C++, and then interfaced with R. In (Pipa et al., 2003) is given an exact  
0999 algorithm when the trial-shuffling is applied to the coincidence count  $\mathbf{C}$  directly. We did  
1000 not follow this line of programming since this exact algorithm is quite long with respect  
1001 to the Monte-Carlo algorithm when the number of simulations is 10000 (as used in the  
1002 present work) and one can see on the bottom left of Figure 5 that the difference between  
1003 both results (Monte-Carlo and exact algorithms) is not detectable at first glance. Simu-  
1004 lations (in Figure 9 and Table 1) as well as a small real data set study show finally that  
1005 the Permutation UE method offers more guarantee in terms of FDR than the methods  
1006 of (Tuleau-Malot et al., 2014) and (Pipa & Grün, 2003; Pipa et al., 2003) applied to the  
1007 delayed coincidence count with a relatively comparable number of discoveries.

1008  
1009 **The i.i.d. assumption.** The main point that remains to be discussed is the i.i.d.  
1010 assumption in view of the classical sticking point in neuroscience: cross-trials non-  
1011 stationarity. As shown on experimental studies (Arieli et al., 1996; Churchland et al.,  
1012 2010; Avila-Akerberg & Chacron, 2011), there is evidence of fluctuating ongoing ac-

1013 tivity in real neuronal networks, which leads to great cross-trials firing rate variability.  
1014 Hence, and even if this variability seems sometimes to decrease with the stimulus, one  
1015 needs to take it into account. Thus, the main question from a statistical point of view is:  
1016 what does it mean for the distribution of the sample  $\mathbb{X}_n$ , that is the distribution of the  
1017 observed data set ?

1018 Several properties have been given in the literature as hints of cross-trials non-  
1019 stationarity. In (Avila-Akerberg & Chacron, 2011), a spike count having a positive  
1020 variance is a result of "trial to trial variability". Yet with such a definition, i.i.d. homoge-  
1021 nous Poisson processes, which actually have a spike count with positive variance, would  
1022 be considered as cross-trials non-stationary, as well as any possible random model for  
1023 spike trains. Other properties are expressed in terms of the Fano Factor ( $FF$ ), defined  
1024 as the quotient of the variance of the spike count by the expectation of the spike count.  
1025 In (Grün et al, 2003) and (Churchland et al., 2010) for instance, a  $FF$  strictly larger  
1026 than 1 is presented as a hint of cross-trials non-stationarity. But renewal processes with  
1027 Gamma interspike interval (ISI) distributions may satisfy  $FF > 1$ , which in fact only  
1028 indicates that the processes are simply not homogeneous Poisson processes. In (Nawrot  
1029 et al., 2008), a "measure" of non-stationarity across trials is given by the difference be-  
1030 tween the  $FF$  and the variation of the ISI ( $CV^2$ ), which is the quotient of the variance  
1031 of the ISI by the expectation of the ISI. Yet, in (Farkhooi et al., 2009; Nawrot, 2010),  
1032 other models are constructed, with correlated ISI's, that satisfy  $FF \neq CV^2$ , and that are  
1033 stationary across trial, this inequality only indicating that the processes are not renewal  
1034 processes.

1035 In view of all these studies, none of these properties, expected to be at least a

1036 hint of cross-trials non-stationarity, can be given as an exact definition of cross-trials  
1037 non-stationarity. In our opinion, the best way to understand what is cross-trials non-  
1038 stationarity is to carefully analyze the models that have been simulated to represent  
1039 such a cross-trials non-stationarity in the above articles. From the simple one of (Grün  
1040 et al., 2003) and simulated in Figure 8.C, to the very intricate one of (Farkhooi et al.,  
1041 2011) through the statistical models used in (Ventura et al., 2005), one can see that  
1042 they all share the principle of doubly stochastic processes. The article of (Churchland  
1043 et al., 2011) is the one that maybe best formalizes this observation, as the cross-trials  
1044 variability is explained from a "mixture of firing rate states", the firing rates changing  
1045 "gradually during decision formation". This is what we tried to catch in a very simple  
1046 way with the simulations of Figure 8.A and Figure 8.B. Following the description of  
1047 (Churchland et al., 2011), there is a hidden variable  $Y$ , called an "intensity command",  
1048 whose realization influences the parameters of the model for  $X$ : typically, the firing  
1049 rate of  $X$  is a function of  $Y$  whose value is fixed once  $Y$  is given. The variable  $Y$  may,  
1050 for instance, model either the variation of depth in anesthesia, the changes in the level  
1051 of attention of the animal or the degree of decision making. It can also be viewed as  
1052 the stimulus in experiments that are subject to stimulus variability (Ben-Shaul et al.,  
1053 2001) or as an oscillatory potential produced by a large non observed network of cells  
1054 influencing both neurons (Kass et al., 2011).

1055 From a probabilistic point of view, our interpretation is that cross-trials non-stationarity  
1056 means that the distribution of the couple  $X = (X^1, X^2)$  is not given intrinsically but  
1057 is given conditionally to a certain random variable  $Y$ , that we call *command variable*  
1058 hereafter in line with (Churchland et al., 2011). The question is then: can  $Y$  be decom-

posed in two independent "command" variables  $Y^1$  and  $Y^2$  that respectively govern the distributions of  $X^1$  and  $X^2$  or do we have a common command variable  $Y$ , that can be viewed as the "common source" of (Ben-Shaul et al., 2001)? In the first case, if  $(X^1, X^2)$  are independent conditionally to  $Y = (Y^1, Y^2)$ , if the distribution of  $X^1$  (respectively  $X^2$ ) is only governed by  $Y^1$  (respectively  $Y^2$ ) and if  $Y^1$  is independent on  $Y^2$ , then for all sets  $A, B$ ,

$$\begin{aligned} \mathbb{P}(X^1 \in A, X^2 \in B) &= \mathbb{E}[\mathbb{P}(X^1 \in A, X^2 \in B|Y)] \\ &= \mathbb{E}[\mathbb{P}(X^1 \in A|Y) \mathbb{P}(X^2 \in B|Y)] \text{ by independence of } (X^1, X^2) \text{ given } Y \\ &= \mathbb{E}[\mathbb{P}(X^1 \in A|Y^1) \mathbb{P}(X^2 \in B|Y^2)] \end{aligned} \quad (7)$$

since there is no common command variable

$$\begin{aligned} &= \mathbb{E}[\mathbb{P}(X^1 \in A|Y^1)] \mathbb{E}[\mathbb{P}(X^2 \in B|Y^2)] \text{ since } Y^1 \text{ is independent of } Y^2 \\ &= \mathbb{P}(X^1 \in A) \mathbb{P}(X^2 \in B). \end{aligned}$$

Hence in this case and despite the command variable  $Y$ , one is still under global independence between  $X^1$  and  $X^2$ , that is  $(H_0)$ . As long as the command variable is i.i.d. across the trials, the distribution of  $\mathbb{X}_n$  is therefore still the one of a  $n$  i.i.d. sample satisfying  $(H_0)$ . This is exactly what happens in Figure 8.C, where the simulation scheme of (Grün et al, 2003; Grün, 2009) exactly satisfies this.

Since bootstrap methods are distribution free, they can in particular handle the fact that the distribution of  $X$  is described via this doubly stochastic process. The only thing that matters is whether there is still global (unconditional) independence between  $X^1$  and  $X^2$ . We believe that the explanation for the bad behavior of the trial-shuffling described in (Grün et al, 2003; Grün, 2009) is not cross-trials non-stationarity but a

1082 centering defect, which can be seen via the behavior of (TSC) versus (TSU) on Figure  
1083 8.C. It is indeed possible that when using the binned coincidence count instead of the  
1084 delayed coincidence count, (TSC) goes from too conservative as on Figure 8.C to not  
1085 enough conservative as shown in the study of (Grün et al, 2003; Grün, 2009). In both  
1086 cases, (TSC) does not reproduce the right distribution under ( $H_0$ ) because the quantity  
1087 at hand is not correctly centered, but once this is corrected, (TSU) is perfectly able to  
1088 give correct  $p$ -values even in this cross-trials non-stationary case.

1089 The same explanation holds for Figure 8.B. In this case, the command variable is  
1090 the index of the trial but it influences only  $X^1$  and not  $X^2$ , so we are exactly in the same  
1091 set-up as without any common command variable: the  $p$ -values behave exactly as usual  
1092 under ( $H_0$ ). However in Figure 8.A, a common command variable (again the index of  
1093 the trial) governs both distributions: the  $p$ -values behave exactly as under ( $H_1$ ) in Figure  
1094 7. Note that it is actually reasonable to reject independence here: indeed (7) does not  
1095 hold and the variables  $X^1$  and  $X^2$  are globally dependent here, since there is definitely  
1096 a common command variable. A similar set-up of common command variable can be  
1097 viewed in the models of conditional dependence proposed by (Ventura et al., 2005) and  
1098 (Kass et al., 2011).

1099 To conclude, what the surrogate methods based on a shuffling of the trials can do  
1100 with respect to cross-trials non-stationarity is also to detect whether there is a common  
1101 command variable or not. In particular, if  $X^1$  and  $X^2$  are independent conditionally to  
1102 the common command variable  $Y$  and do not present any "fine temporal coordination of  
1103 spikes in neuronal preprocessing", as stated in (Grün et al, 2003), the test is still likely  
1104 to reject the independence assumption. Yet, this is not a False Positive with respect to

1105 the statistical meaning of the test. Indeed, in this situation the spike trains  $X^1$  and  $X^2$   
1106 are correlated since they are not globally independent. However, they are conditionally  
1107 independent once the command variable is fixed and in this sense they do not really  
1108 present any synchrony. This kind of distinction between correlation and synchrony was  
1109 already underlined and discussed on cross-correlograms by Brody (1999a,b).

1110 Finally, one could wonder what is really assumed by i.i.d. trials. The independence  
1111 between trials is, in our opinion, not really an issue since the trials are usually suffi-  
1112 ciently far apart in time. The main assumption is therefore the identical distribution. As  
1113 explained above, cross-trials non-stationarity interpreted as a command variable phe-  
1114 nomenon does not contradict this assumption. We can even not imagine how this as-  
1115 sumption can be defective in practice. Even in the extreme case where half of the trials  
1116 would be sampled from an anesthetized animal and the other half from a non anes-  
1117 thetized animal, considering the presence or not of anesthesia as a command variable  
1118 lead to i.i.d. trials from a mixture point of view.

1119 This naturally leads to the following completely open question. Is the global inde-  
1120 pendence property really the assumption that the neuroscientists want to test?

1121 On the one hand, in (Churchland et al., 2011), it is stated that "variance itself can  
1122 be diagnostic of neural computation". We interpret this in the present framework as  
1123 follows: if one is able to detect a common command variable (not known before hand),  
1124 then one is able to detect "neural computation". This line is totally in accordance with  
1125 the discussion of (Ben-Shaul et al., 2001) where global dependence can be viewed as  
1126 the presence of an "internal variable" when "the variability of all relevant stimuli or  
1127 actions has been accounted for" by the experimental design and therefore when the test

1128 has been applied only to trials that are homogeneous with respect to this experimental  
1129 design. Hence global independence may have a meaning in neuroscience.

1130 On the other hand, a more precise description of the dependence may be needed.  
1131 For instance, Hawkes processes allow to model local independence. For instance, if the  
1132 command variable is the spike train of a third observed neuron, it is eventually possible  
1133 via the methods of (Hansen et al., 2015; Reynaud-Bouret et al., 2013) and under a  
1134 Hawkes distribution assumption, to see whether this third neuron influences both  $X^1$   
1135 and  $X^2$  with or without direct (local) dependence between  $X^1$  and  $X^2$ . In the same  
1136 line, the works of (Ventura et al., 2005; Kass et al., 2011) give another precise model of  
1137 conditional independence that can be tested. Up to our knowledge, however, there is no  
1138 distribution free method that would be able to assess this, in particular if the command  
1139 variable is hidden.

1140 Therefore, contrary to what is currently believed, the present statistical study shows  
1141 that surrogate data methods based on a shuffling of the trials can behave properly under  
1142 cross-trials non-stationarity if they are applied to correctly centered quantities and if  
1143 one wishes to detect global dependence. The other popular surrogate method based on  
1144 dithering (Louis et al., 2010b) is much more difficult to study from a mathematical point  
1145 of view, principally because, unlike bootstrap methods, there is no general mathematical  
1146 theory explaining why moving individual spikes would mimic the overall distribution of  
1147 the coincidence count or any centered version of it under independence. One possible  
1148 guess, which is maybe far fetched, is that dithering, as a much more local surrogate data  
1149 procedure, may be able somehow to detect local and not global dependence, if one can  
1150 correctly tune it.

1151 Another open question, which seems much more achievable, is to adapt those boot-  
1152 strap procedures to more than two neurons. Indeed, delayed coincidence counts have  
1153 already been introduced in this case in (Chevallier & Laloë, 2015) and similar bootstrap  
1154 procedures have been developed for more than two real valued variables in the precur-  
1155 sor work of (Romano, 1989).

## 1156 1157 **Acknowledgments**

1159 We first of all want to thank A. Riehle, leader of the laboratory in which the data used  
1160 in this article were previously collected. We are very grateful to F. Grammont, who no-  
1161 tably recorded the data, for fruitful and stimulating discussions. We also warmly thank  
1162 Gilles Blanchard for all his relevant suggestions, and both referees for their comments,  
1163 which really helped us to improve the present manuscript. This work was granted ac-  
1164 cess to the HPC and visualization resources of "Centre de Calcul Interactif" hosted by  
1165 the University of Nice Sophia-Antipolis. It was partly supported by the french Agence  
1166 Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration), by the PEPS  
1167 BMI 2012-2013 *Estimation of dependence graphs for thalamo-cortical neurons and*  
1168 *multivariate Hawkes processes* and by the interdisciplinary axis MTC-NSC of the Uni-  
1169 versity of Nice Sophia-Antipolis. The PhD grant of M. Albert is funded by the PACA  
1170 french region.

## References

- 1174  
1175 Arieli, A., Sterkin, A., Grinvald, A., & Aertsen, A. (1996) Dynamics of Ongoing  
1176 activity: explanation of the large variability in evoked cortical responses. *Science*,  
1177 273, 1868–1871.
- 1178  
1179 Aertsen, A.M., Gerstein, G.L., Habib, M.K., & Palm, G. (1989) Dynamics of neuronal  
1180 firing correlation: modulation of "effective connectivity". *Journal of Neurophysiol-*  
1181 *ogy*, 61(5), 900–917.
- 1182  
1183 Albert, M., Bouret, Y., Fromont, M., & Reynaud-Bouret, P. (2015) Bootstrap and per-  
1184 mutation tests of independence for point processes. *to appear in Annals of Statistics*.
- 1185  
1186 Avila-Akerberg, O., & Chacron, M. J. Nonrenewal spike trains statistics: causes and  
1187 functional consequences on neural coding. *Exp. Brain Res.*, 210, 353–371.
- 1188  
1189 Benjamini, Y., & Hochberg, Y. (1995) Controlling the false discovery rate: a practical  
1190 and powerful approach to multiple testing. *Journal of the Royal Statistical Society*,  
1191 *Series B*, 57(1), 289–300.
- 1192  
1193 Benjamini, Y., & Yekutieli, D. (2001) The control of the false discovery rate in multiple  
1194 testing under dependency. *Annals of Statistics*, 29(4), 1165–1188.
- 1195  
1196 Ben-Shaul, Y., Bergman, H., Ritov, H., & Abeles, M. (2001) Trial to trial variability  
in either stimulus or action causes apparent correlation and synchrony in neuronal  
activity. *Journal of Neuroscience Methods*, 111, 99–110.
- Bickel, P. J., & Freedman, D.A. (1981) Some asymptotic theory for the bootstrap. *Ann.*  
*Statist.*, 9(6), 1196–1217.

- 1197 Brody, C. D. (1999a) Correlations without synchrony. *Neural Computation*, *11*, 1537–  
1198 1551.
- 1199 Brody, C. D. (1999b) Disambiguating different covariation types. *Neural Computation*,  
1200 *11*, 1527–1535.
- 1201  
1202 Chevallier, J., & Laloë, T. (2015) Detection of dependence patterns with delay. *to*  
1203 *appear in Biometrical Journal*.
- 1204 Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X.-J., Pouget, A., & Shadlen, M.  
1205 N. (2011) Variance as a signature of neural computations during decision making.  
1206 *Neuron*, *69*, 818–831.
- 1207 Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado,  
1208 G. S., Newsome, W. T., Clark, A. M., Hosseini, P., Scott, B. B., Bradley, D. C.,  
1209 Smith, M. A., Kohn, A., Movshon, J. A., Armstrong, K. M., Moore, T., Chang, S.  
1210 W., Snyder, L. H., Lisberger, S. G., Priebe, N. J., Finn, I. M., Ferster, D., Ryu, S.  
1211 I., Santhanam, G., Sahani, M., & Shenoy, K. V. (2010) Stimulus onset quenches  
1212 neural variability: a widespread cortical phenomenon. *Nature Neuroscience*, *13*(3),  
1213 369–378.
- 1214  
1215 Daley, D.J., & Vere-Jones, D. (2005) An introduction to the theory of point processes.  
1216 *Springer series in statistics* Volume I.
- 1217 Denker, M., Wiebelt, B., Fliegner, D., Diesmann, M., & Morrison, A. (2010) Practically  
1218 Trivial Parallel Data Processing in a Neuroscience Laboratory. *In Analysis of Parallel*  
1219 *Spike Trains*, Grün, S., & Rotter, S., Springer Series in Computational Neuroscience.

- 1220 Farkhooi, F., Froese, A., Muller, E., Menzel, R., & Nawrot, M. P. (2013) Cellular  
1221 adaptation facilitates sparse and reliable coding in sensory pathways. *PLOS Comp.*  
1222 *Biol.*, 9(10), e1003251.
- 1223 Farkhooi, F., Muller, E., & Nawrot, M. P. (2011) Adaptation reduces variability of  
1224 neuronal population code. *Physical Review E*, 83, 050905.
- 1225 Farkhooi, F., Strube-Bloss, M. F., & Nawrot, M. P. (2009) Serial correlation in neural  
1226 spike trains: Experimental evidence, stochastic modeling, and single neuron variabil-  
1227 ity. *Physical Review E*, 79, 021905.
- 1228
- 1229 Giné, E. (1997) Lectures on some aspects of the bootstrap. *Lecture Notes in Math*  
1230 1665, 37–152. Ecole d’été de Probabilités de Saint-Flour, XXVI-1996.
- 1231 Grammont, F., & Riehle, A. (2003) Spike synchronisation and firing rate in a popula-  
1232 tion of motor cortical neurons in relation to movement direction and reaction time.  
1233 *Biological Cybernetics*, 88, 360–373.
- 1234
- 1235 Grün, S. (1996) Unitary joint-events in multiple-neuron spiking activity: Detection,  
1236 significance and interpretation. *Thun: Verlag Harri Deutsch*.
- 1237 Grün, S. (2009) Data-driven Significance estimation for precise spike correlation. *J.*  
1238 *Neurophysiol.*, 101, 1126–1140.
- 1239 Grün, S., Diesmann, M., & Aertsen, A.M. (2002) Unitary events in multiple single-  
1240 neuron activity: I. Detection and significance *Neural Computation*, 14(1), 43–80.
- 1241 Grün, S., Diesmann, M., & Aertsen, A.M. (2002) Unitary events in multiple single-  
1242 neuron activity: II. Nonstationary data *Neural Computation*, 14(1), 81–119.

- 1243 Grün, S., Diesmann, M., & Aertsen, A.M. (2010) Unitary Events Analysis. *In Analysis*  
1244 *of Parallel Spike Trains*, Grün, S., & Rotter, S., Springer Series in Computational  
1245 Neuroscience.
- 1246 Grün, S., Diesmann, M., Grammont, F., Riehle, A., & Aertsen, A. (1999) Detecting  
1247 unitary events without discretization in time. *Journal of Neuroscience methods* 94.
- 1248  
1249 Grün, S., Riehle, A., & Diesmann, M. (2003) Effect of cross-trial nonstationarity on  
1250 joint-spike events. *Biol. Cybern.*, 88, 335–351.
- 1251 Gütig, R., Aertsen, A.M., & Rotter, S. (2001). Statistical Significance of Coincident  
1252 Spikes: Count-Based Versus Rate-Based Statistics. *Neural Computation*, 14, 121–  
1253 153.
- 1254 Hansen, N.R., Reynaud-Bouret, P., & Rivoirard, V. (2015). Lasso and probabilistic  
1255 inequalities for multivariate point processes. *Bernoulli*, 21(1), 83–143.
- 1256  
1257 Hoeffding, V. (1952) The large sample power of tests based on permutation of the  
1258 observations. *The Annals of Mathematical Statistics*, 3(2), 169–192.
- 1259 Kass, R.E., Kelly, R. C., & Loh, W.-L. (2011) Assessment of synchrony in multiple  
1260 neural spike trains using loglinear point process models. *Annals of Applied Statistics*,  
1261 5(2B), 1262–1292.
- 1262 Kilavik, B.E., Roux, S., Ponce-Alvarez, A. Confais, J. Grün, S., & Riehle, A.(2009)  
1263 Long-term modifications in motor cortical dynamics induced by intensive practice.  
1264 *Journal of Neuroscience*, 29(40), 12653–12663.
- 1265

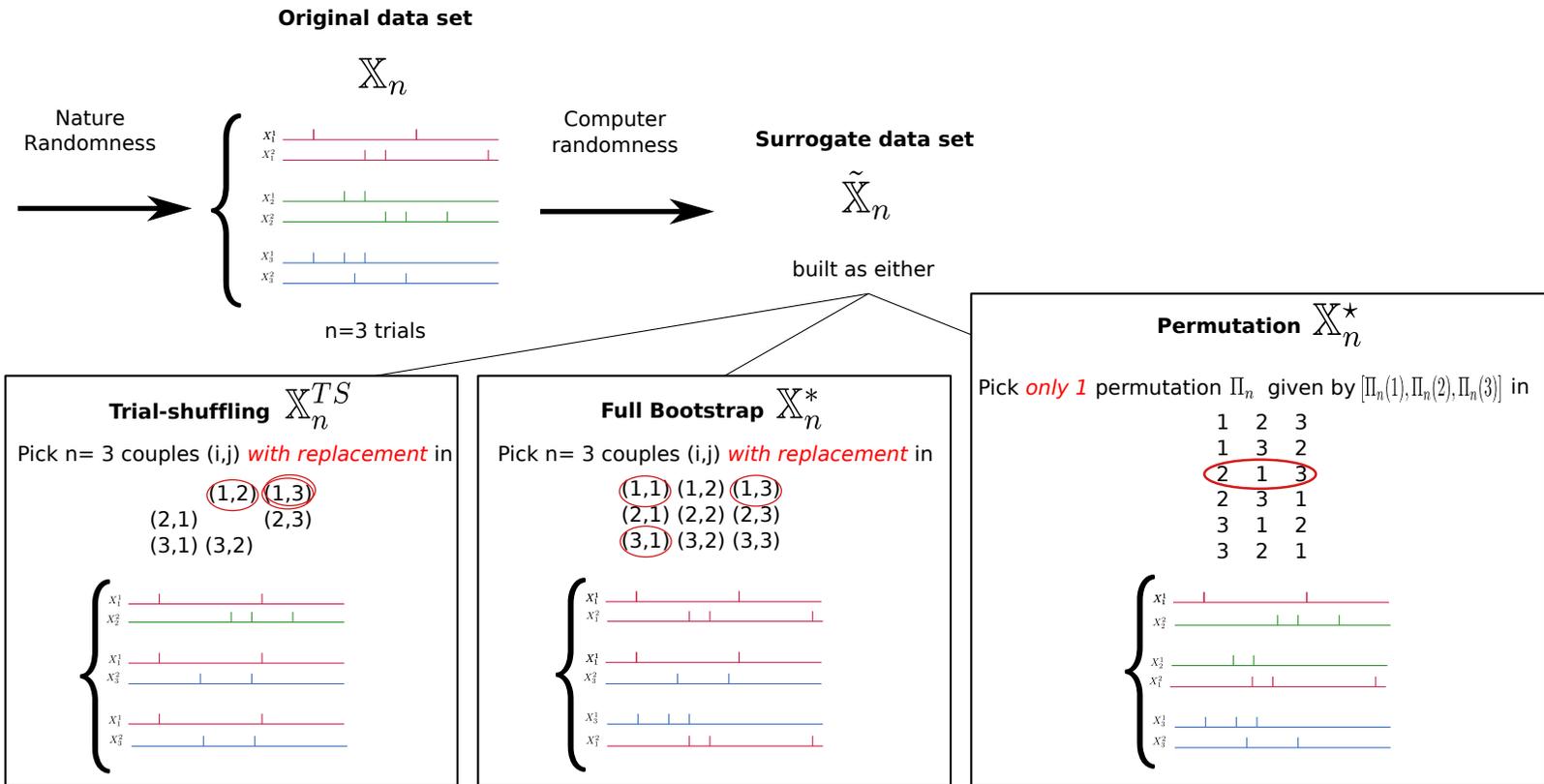
- 1266 Litwin-Kumar, A., & Doiron, B. (2012) Slow dynamics and high variability in balanced  
1267 cortical networks with clustered connections. *Nature Neuroscience*, 15(11), 1498–  
1268 1505.
- 1269 Louis, S., Borgelt, C., & Grün, S. (2010) Generation and Selection of Surrogate Meth-  
1270 ods for Correlation Analysis. *In Analysis of Parallel Spike Trains*, Grün, S., & Rotter,  
1271 S., Springer Series in Computational Neuroscience.
- 1272 Louis, S., Gerstein, G. L., & Diesmann, M. (2010) Surrogate spike train generation  
1273 through dithering in operational time. *Frontiers in Computational Neuroscience*, 4  
1274 (article 127).
- 1275
- 1276 Nawrot, M. P. (2010) Analysis and interpretation of interval and count variability in  
1277 neural spike trains. *In Analysis of Parallel Spike Trains*, Grün, S., & Rotter, S.,  
1278 Springer Series in Computational Neuroscience.
- 1279 Nawrot, M. P., Boucsein, C., Molina, V. R., Riehle, A., Aertsen, A., & Rotter, S. (2008)  
1280 Measurement of variability dynamics in cortical spike trains. *Journal of Neuroscience*  
1281 *Methods*, 169, 374–390.
- 1282
- 1283 Perkel, D.H., Gernstein, G.L., & Moore, G.P. (1967) Neuronal spike trains and stochas-  
1284 tic point processes. *Biophysical Journal*, 7, 419–440.
- 1285 Pipa, G., Diesmann, M., & Grün, S. (2003) Significance of joint-spike events based on  
1286 trial-shuffling by efficient combinatorial methods. *Complexity*, 8(4), 1–8.
- 1287 Pipa, G., & Grün, S. (2003) Non-parametric significance estimation of joint-spike  
1288 events by shuffling and resampling. *Neurocomputing*, 52-54, 31–37.

- 1289 Pipa, G., Grün, S., & van Vreeswijk, C. (2013). Impact of Spike Train Autostructure on  
1290 Probability Distribution of Joint Spike Events. *Neural Computation*, 25, 1123–1163.
- 1291 Pouzat, C., & Chaffiol, A. (2009) Automatic spike train analysis and report generation.  
1292 An implementation with R, R2HTML and STAR. *Journal of Neuroscience Methods*,  
1293 181, 119–144.
- 1294  
1295 Reynaud-Bouret, P., Rivoirard, V., & Tuleau-Malot, C. (2013) Inference of functional  
1296 connectivity in Neurosciences via Hawkes processes. *1st IEEE Global Conference*  
1297 *on Signal and Information Processing, Austin, Texas.*
- 1298 Riehle, A., Grammont, F., Diesmann, M., & Grün, S. (2000) Dynamical changes and  
1299 temporal precision of synchronised spiking activity in monkey motor cortex during  
1300 movement preparation. *Journal of Physiology*, 94, 569–582.
- 1301 Riehle, A., Grammont, F., & MacKay, A. (2006) Cancellation of a planned movement  
1302 in monkey motor cortex. *Neuroreport*, 17(3), 281–285.
- 1303  
1304 Romano, J.P. (1989) Bootstrap and Randomization Tests of some Nonparametric Hy-  
1305 potheses. *Ann. Statist.* 17(1), 141–159.
- 1306 Romano, J.P., & Wolf, M. (2005) Exact and approximate step-down methods for mul-  
1307 tiple hypothesis testing. *J. Amer. Statist. Assoc.*, 100, 94–108.
- 1308  
1309 Singer, W. (1993) Synchronization of cortical activity and its putative role in informa-  
1310 tion processing and learning. *Annual Review of Physiology*, 55, 349–374.
- 1311 Tuleau-Malot, C., Rouis, A., Grammont, F., & Reynaud-Bouret, P. Multiple Tests

based on a Gaussian Approximation of the Unitary Events method with delayed coincidence count. *Neural Computation*, 26(7), 1408–1454.

Ventura, V. (2010) Bootstrap tests of hypotheses. *In Analysis of Parallel Spike Trains*, Grün, S., & Rotter, S., Springer Series in Computational Neuroscience.

Ventura, V., Cai, C., & Kass, R.E. (2010) Trial-to-trial variability and its effect on time-varying dependency between two neurons. *Journal of Neurophysiology*, 94, 2928–2939.



**Unconditional distribution:** all possible choices of both Nature and Computer randomness

**Conditional distribution:** 1 fixed original data set (Nature randomness), all possible choices of Computer randomness

Figure 4: Schematic view of the three bootstrap procedures. Note in particular that  $n$  draws with replacement are necessary for the trial-shuffling and the full bootstrap approach, whereas only one draw of one permutation is necessary for the permutation approach. Note also that it is perfectly possible that a surrogate data set done by trial-shuffling or full bootstrap approaches may perfectly pick twice the same trial and at the same time leave out one or more of the original trials, whereas the permutation is always exhaustive in this sense. Such typical draws are given by the red circles, leading to the given surrogate data set for each method. Finally note that the unconditional distribution let both randomness (Nature and Computer) vary and that this is not realistic since in practice we have only one original data set. This is why the conditional distribution is the one that can be simulated via a computer for a given observation.

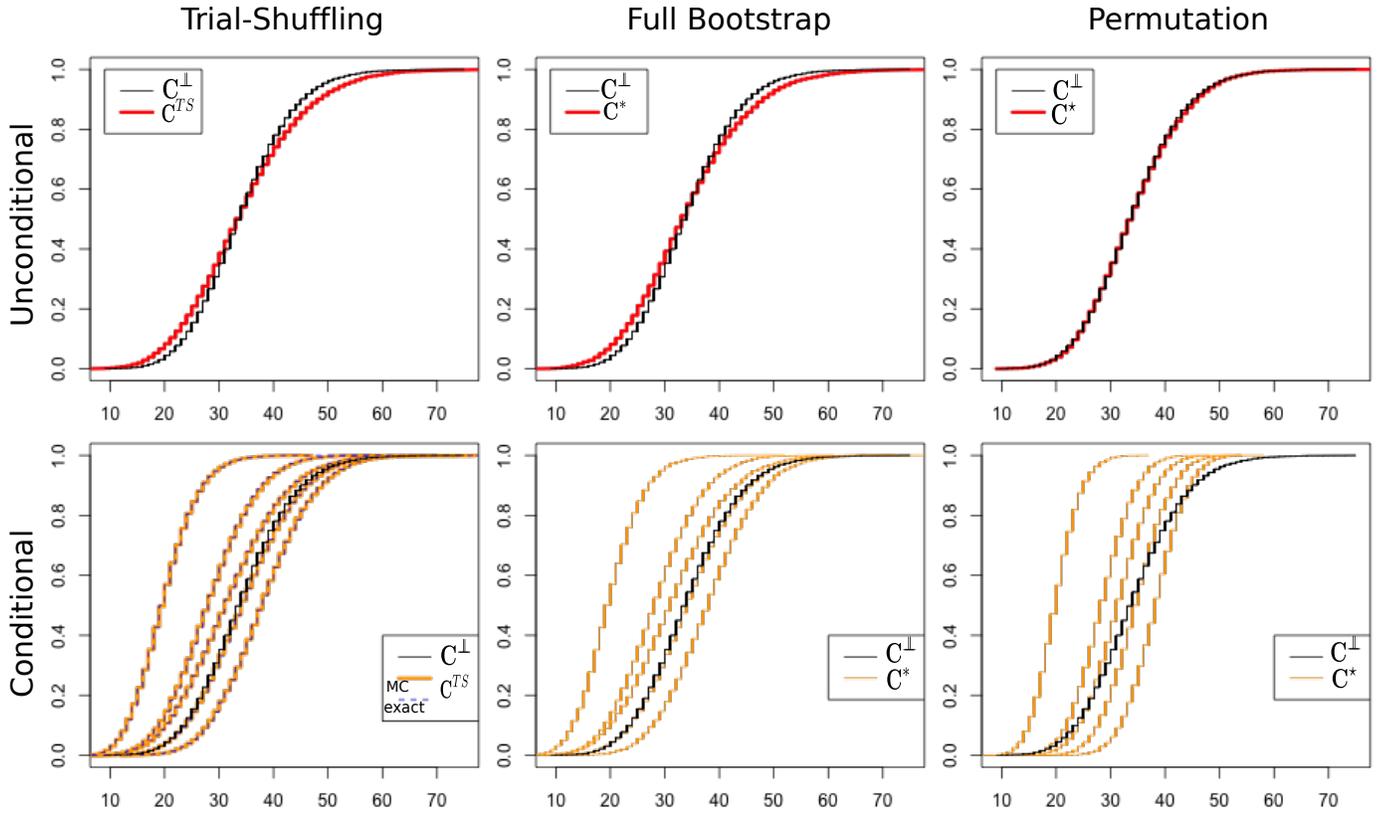


Figure 5: The unconditional distribution and conditional distributions of  $C$  under  $(H_0)$ . C.d.f. of  $C(\mathbb{X}_n^\perp)$  and (for the first line) of  $C^{TS} = C(\mathbb{X}_n^{TS})$ , of  $C^* = C(\mathbb{X}_n^*)$  and of  $C^* = C(\mathbb{X}_n^{\Pi_n})$  obtained from 10000 simulations of  $n = 20$  trials of two independent Poisson processes of firing rate 30Hz on a window of length 0.1s with  $\delta = 0.01$ s. On the second line, in addition to the c.d.f. of  $C(\mathbb{X}_n^\perp)$ , five observations of  $\mathbb{X}_n = \mathbb{X}_n^\perp$  have been simulated in the same set-up and given these observations, the conditional c.d.f. have been approximated by simulating 10000 times the extra-randomness corresponding to  $\tilde{\mathbb{X}}_n$ . For the trial-shuffling, in addition to this approximate Monte-Carlo method (MC), the exact conditional c.d.f. has been obtained thanks to the algorithm of (Pipa et al., 2003).

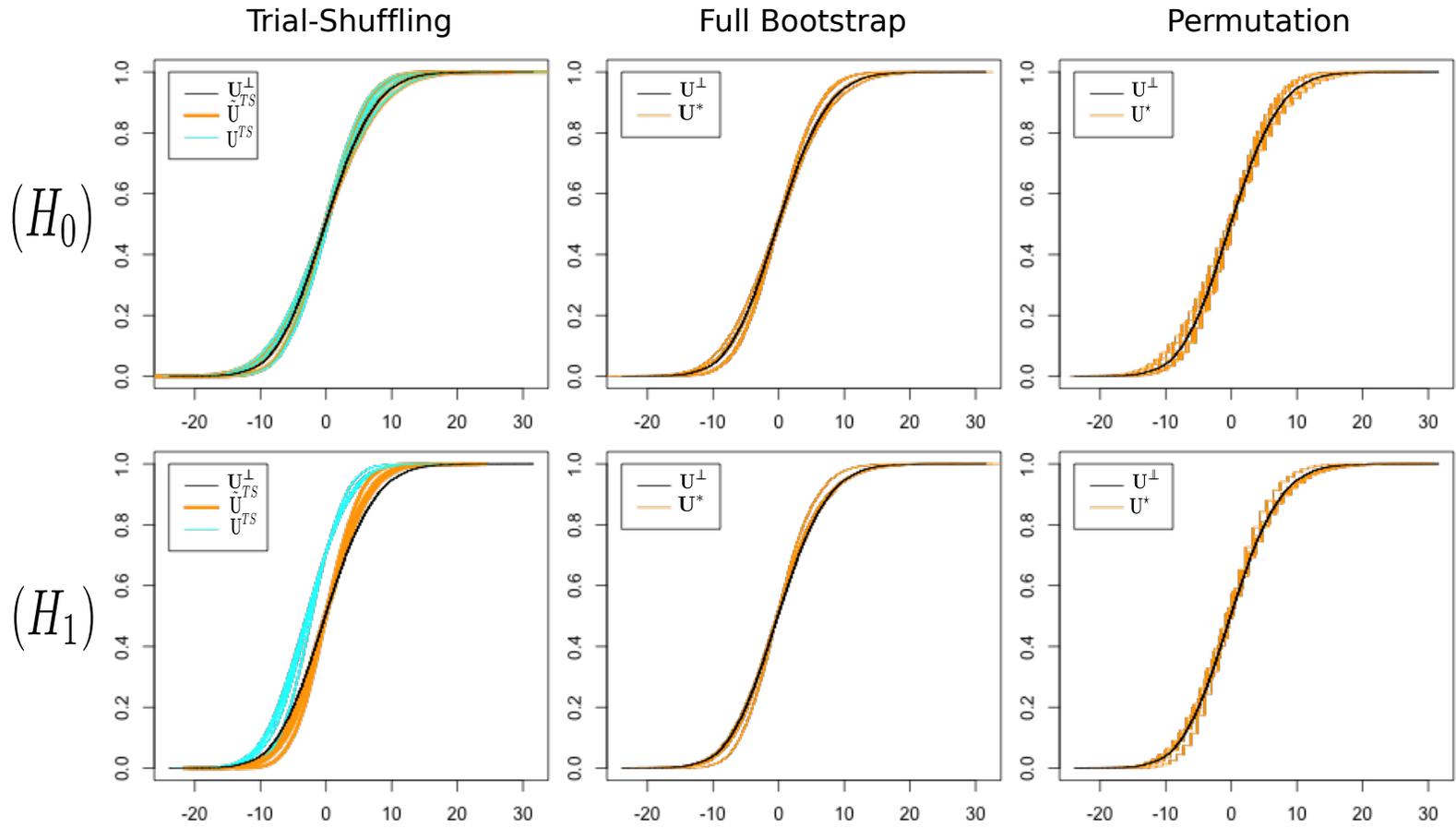


Figure 6: Conditional distribution of  $U(\tilde{\mathbb{X}}_n)$  (or its recentered version  $\tilde{U}^{TS}$  for the Trial-Shuffling) given  $\mathbb{X}_n$ . Cumulative distribution functions of  $U^\perp = U(\mathbb{X}_n^\perp)$  in black, obtained by simulation as in Figure 5. For the first line, under  $(H_0)$ , five observations of  $\mathbb{X}_n^\perp$  in the same set-up have been fixed and given these observations, the conditional c.d.f. of  $U^{TS} = U(\mathbb{X}_n^{TS})$ , of  $\tilde{U}^{TS} = U^{TS} + U^{obs}/n$ , of  $U^* = U(\mathbb{X}_n^*)$  and of  $U^* = U(\mathbb{X}_n^{\Pi n})$  have been obtained as in Figure 5. For the second line, five observations of  $\mathbb{X}_n$ , simulated under  $(H_1)$  with marginals equal to the ones of the first line but satisfying  $X^1 = X^2$ , have been simulated and conditional c.d.f. are obtained in the same way as above.

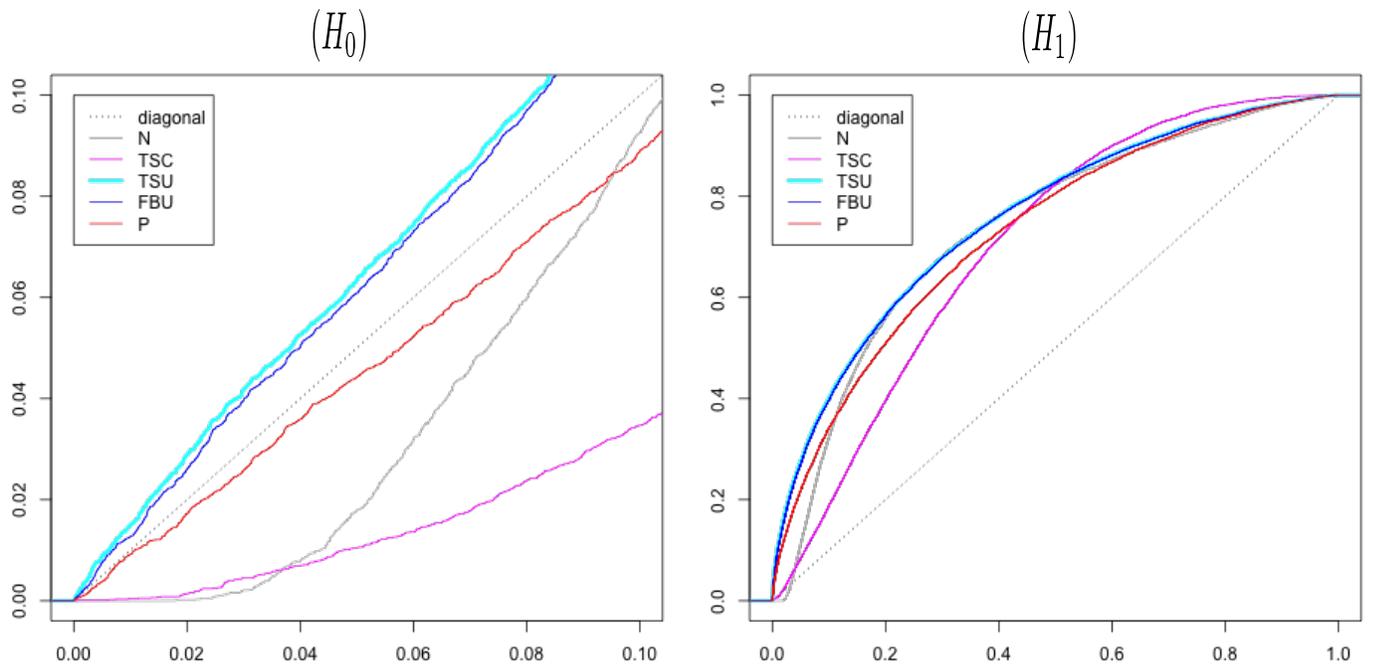


Figure 7: Distribution of the  $p$ -values for the different tests. C.d.f. under both  $(H_0)$  and  $(H_1)$  of the  $p$ -values for the five tests: naive (N), Trial-Shuffling version **C** (TSC), Trial-Shuffling version **U** (TSU), Full Bootstrap version **U** (FBU), and Permutation (P). Under  $(H_0)$ , the c.d.f. are obtained by simulations done as in Figure 5; the c.d.f. are then plotted only for small  $p$ -values ( $\leq 0.1$ ). Under  $(H_1)$ , the couple  $(X^1, X^2)$  is constructed by injection (Grün et al., 2010; Tuleau-Malot et al., 2014), i.e. as  $(N^1 \cup N^{inj}, N^2 \cup N^{inj})$  where  $(N^1, N^2)$  are two independent Poisson processes of firing rate 27 Hz on a window of length 0.1s and where  $N^{inj}$  is a common Poisson process of firing rate 3Hz; once again, 20 i.i.d. trials are simulated 10000 times to obtain the corresponding c.d.f. with  $\delta = 0.01$ s.

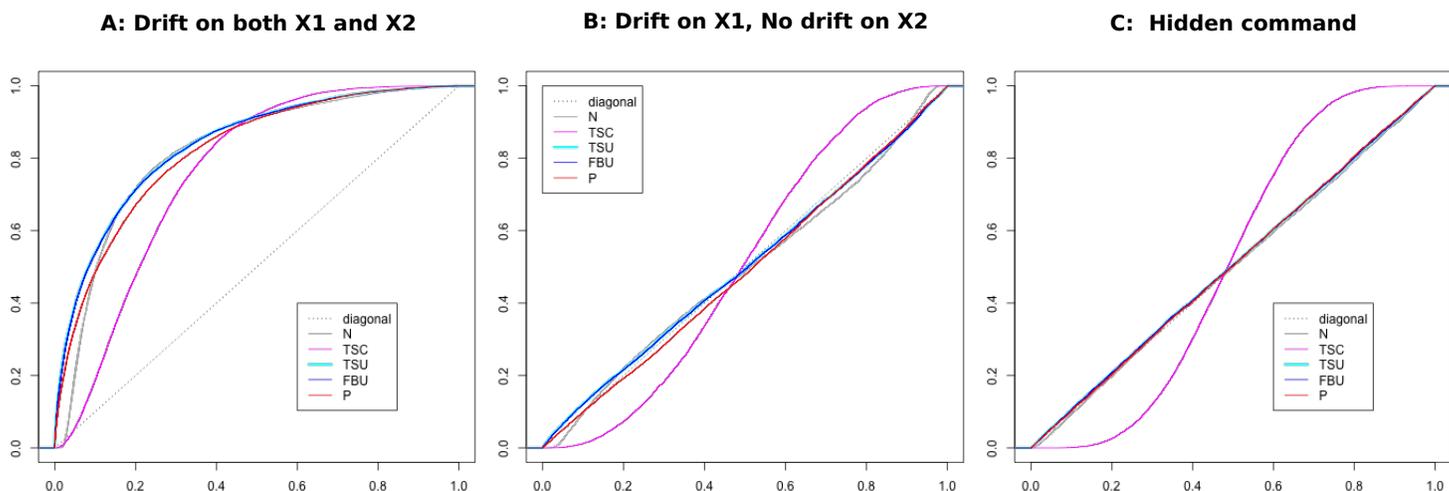


Figure 8: Distribution of the  $p$ -values for the different tests under varying firing rates. C.d.f. of the  $p$ -values for the five tests: naive (N), Trial-Shuffling version C (TSC), Trial-Shuffling version U (TSU), Full Bootstrap version U (FBU), and Permutation (P) computed over 10 000 simulations. In **A**,  $n = 20$  trials are drawn, the firing rates are constant on each trial and they are regularly increasing, from trial 1 to trial 20, from respectively 10Hz to 100Hz for  $X^1$  and from 10Hz to 50Hz for  $X^2$ . Once the rates fixed, both spike trains in each trial are independent homogeneous Poisson processes of the prescribed rates. The length of the interval  $[a, b]$  is 0.1s and  $\delta = 0.01$ s. In **B**, the same set-up is taken except that the firing rate of  $X^2$  is fixed equal to 50 Hz. In **C**, the simulation set up of (Grün, 2009) is taken: 100 trials of 1s duration and in each trial  $i$  and each component  $j$  is independently drawn (i) first, a hidden command variable,  $Y_i^j$ , which is here a Bernoulli variable of parameter 0.7 (ii) the spike train  $X_i^j$  is then simulated as a homogeneous Poisson process of firing rate 30Hz, if  $Y_i^j=1$  and of firing rate 90Hz, if  $Y_i^j = 0$ .

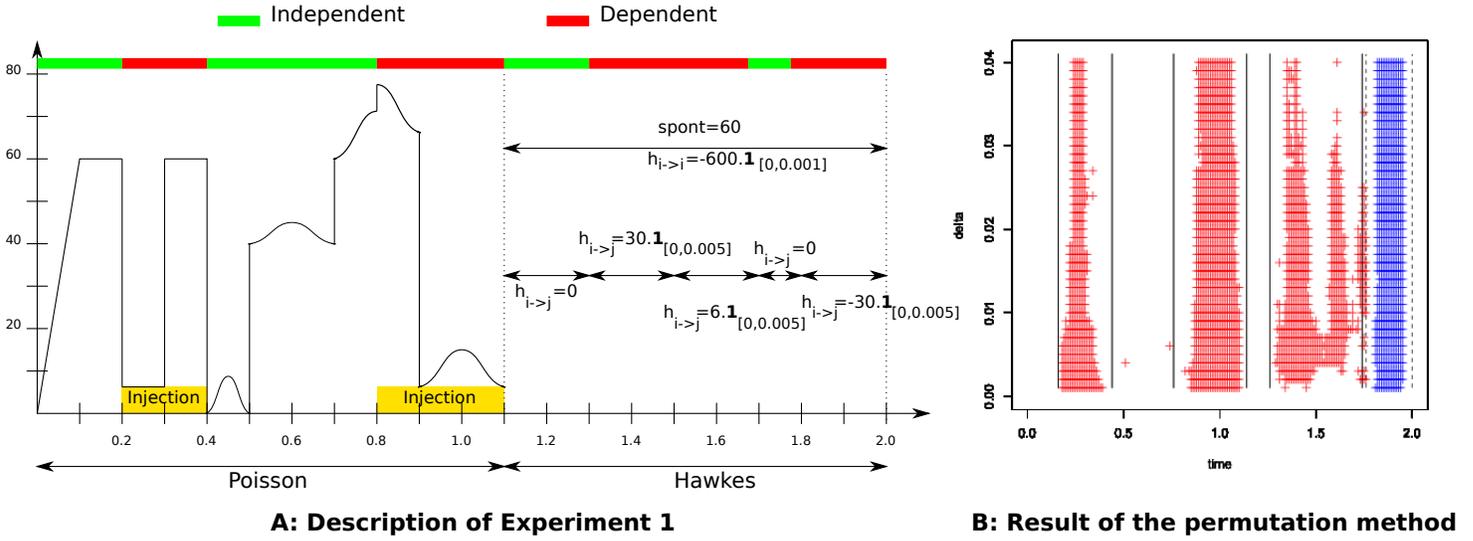


Figure 9: Multiple tests. 9.A: description of Experiment 1. In the Poisson part, the intensity of both Poisson processes is plotted. The injection component corresponds to the part of a shared Poisson process which is injected in both processes corresponding to  $X^1$  and  $X^2$ , as explained in Figure 7. In the Hawkes part (see (Tuleau-Malot et al., 2014) for a complete description), formulas for the spontaneous parameters and both self interaction  $h_{i \rightarrow i}$  and cross interaction  $h_{i \rightarrow j}$  functions are given. 9.B: results of the Permutation UE method ( $B = 10000$ ,  $q = 0.05$ ) performed on 191 overlapping windows of the form  $[a, a + 0.1]$  for  $a$  in  $\{0, 0.01, \dots, 1.9\}$  on one run of simulation for 50 trials of Experiment 1. A red (resp. blue) cross is represented at the center of the window when it is detected by a  $p_W^+$  (resp.  $p_W^-$ ). Each horizontal line corresponds to a different  $\delta$  in  $\{0.001, 0.002, \dots, 0.04\}$ . The black vertical lines delimit the regions where the independence hypothesis is not satisfied: plain for positive dependence (i.e. where  $C^{obs}$  should be too large), and dashed for negative dependence (i.e. where  $C^{obs}$  should be too small).

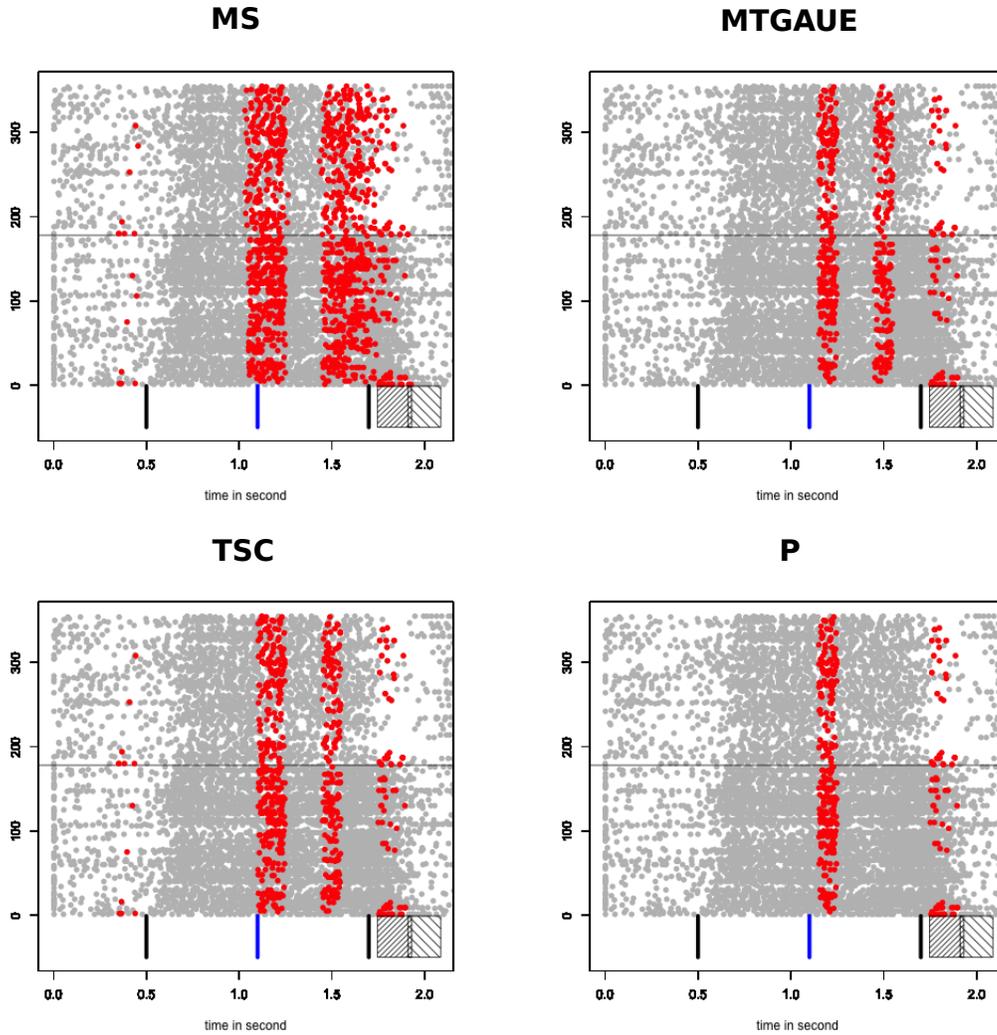


Figure 10: Raster plots of the pair of neurons 13. In red the Unitary Events where the coincidence count is significantly too large for the three methods (MTGAUE, TSC and P) presented in Table 1 and for the Multiple Shift method (MS), with  $\delta = 0.02s$  and  $B = 10000$ , on overlapping windows of the form  $[a, a + 0.1]$  for  $a$  in  $\{0, 0.05, \dots, 1.95\}$ . No interval was detected for a significantly too small coincidence count. Signs on bottom corresponds to behavioral events. The first black vertical bar corresponds to the preparatory signal (PS), the blue vertical bar to the expected signal (ES), the second black vertical bar to the response signal (RS). The first hatched box corresponds to the interval [mean reaction time (RT) minus its standard deviation, mean reaction time (RT) plus its standard deviation], the second hatched box corresponds to the same thing but for the movement time (MT).

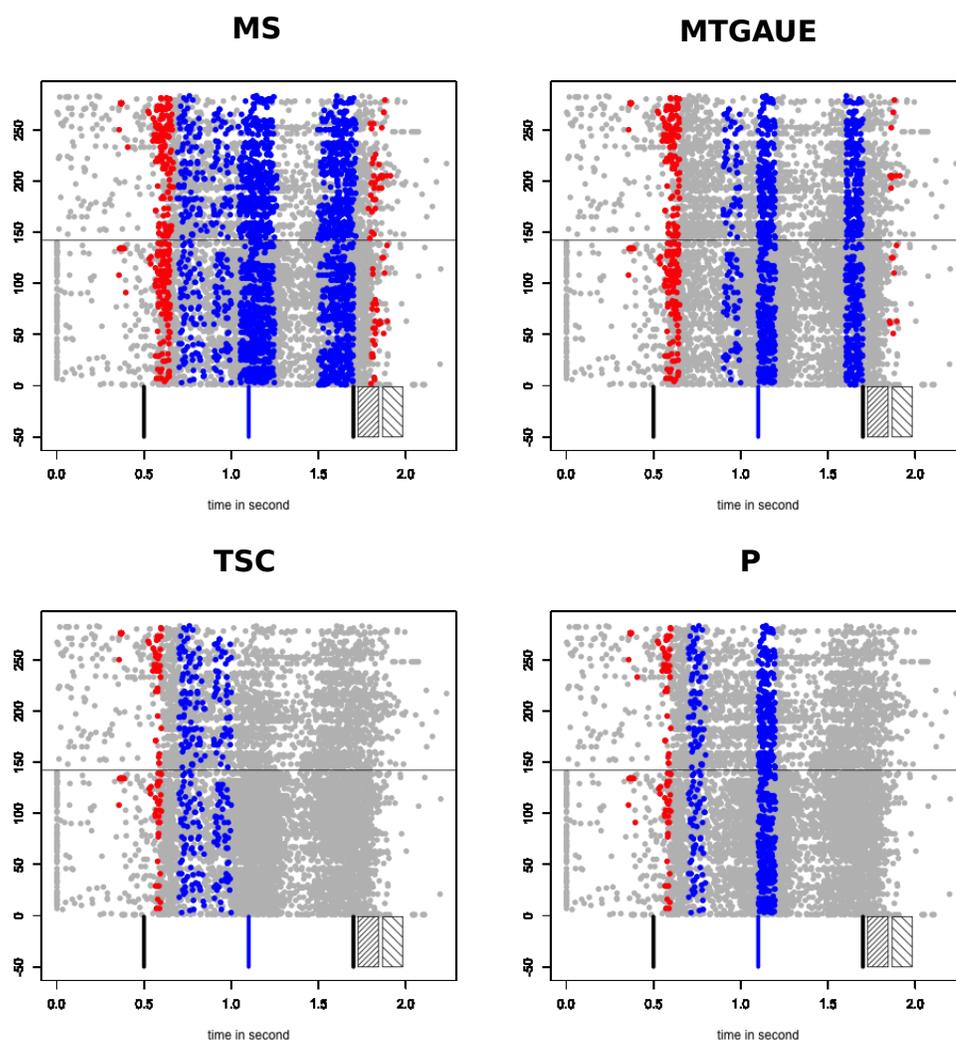


Figure 11: Raster plots of the pair of neurons 40. In red the Unitary Events where the coincidence count is significantly too large for the three methods (MTGAUE, TSC and P) presented in Table 1 and for the Multiple Shift method (MS), with  $\delta = 0.02s$  and  $B = 10000$ , on overlapping windows of the form  $[a, a + 0.1]$  for  $a$  in  $\{0, 0.05, \dots, 2.1\}$ . In blue the Unitary Events where the coincidence count is significantly too small with the same convention. Signs on bottom corresponds to behavioral events as described in Figure 10.