CrossMark

Multi-scale detection of rate changes in spike trains with weak dependencies

Michael Messer[1] · Kauê M. Costa[2] · Jochen Roeper[2] · Gaby Schneider[1]

Abstract The statistical analysis of neuronal spike trains by models of point processes often relies on the assumption of constant process parameters. However, it is a well-known problem that the parameters of empirical spike trains can be highly variable, such as for example the firing rate. In order to test the null hypothesis of a constant rate and to estimate the change points, a Multiple Filter Test (MFT) and a corresponding algorithm (MFA) have been proposed that can be applied under the assumption of independent inter spike intervals (ISIs). As empirical spike trains often show weak dependencies in the correlation structure of ISIs, we extend the MFT here to point processes associated with short range dependencies. By specifically estimating serial dependencies in the test statistic, we show that the new MFT can be applied to a variety of empirical firing patterns, including positive and negative serial correlations as well as tonic and bursty firing. The new MFT is applied to a data set of empirical spike trains with serial correlations, and simulations show improved performance against methods that assume independence. In case of positive correlations, our new MFT is necessary to reduce the number of false positives, which can be highly enhanced when falsely assuming independence. For the frequent case of negative correlations, the new MFT shows an improved detection probability of change points and thus, also a higher potential of signal extraction from noisy spike trains.

Keywords Spike train analysis · Non-stationarity · Serial correlation · Change point detection · Point processes · Multi scale

Action Editor: Liam Paninski

✉ Gaby Schneider
schneider@math.uni-frankfurt.de

Michael Messer
messer@math.uni-frankfurt.de

Kauê M. Costa
kaue.m.costa@gmail.com

Jochen Roeper
roeper@em.uni-frankfurt.de

[1] Institute of Mathematics, Johann Wolfgang Goethe University Frankfurt, Frankfurt, Germany

[2] Institute of Neurophysiology, Johann Wolfgang Goethe University Frankfurt, Frankfurt, Germany

1 Introduction

One fundamental property of neurons is that they can code information by varying their firing rate. In many neurons, information is encoded in transient spike rate changes against a variable baseline. The statistical structure of baseline firing in spike trains is an important determinant of neuronal information encoding (Luczak et al. 2013; Hartmann et al. 2015). Likewise, state-related changes in baseline firing rate alone affect signal to noise ratio and the quality of information encoding (Lee and Dan 2012). In this context, statistical estimation of rate change points is an important tool for extracting relevant features from neuronal signals, especially during so called spontaneous firing neuronal activity during sleep, periods of quiet wakefulness, under anesthesia, or whenever there is no direct behavioral or sensorial trigger for the recorded neural signal, but which

does carry important information about the structure of neuronal networks and the biophysics of individual neurons (Schiemann et al. 2012; Luczak et al. 2013; Hartmann et al. 2015).

General point process models have been proposed for the description of varying firing rates and their dependence on past spiking activity, external stimuli or behavioral events (Brown et al. 2004; Koyama et al. 2010; Pillow et al. 2008; Paninski 2004; Truccolo et al. 2004). The present article is motivated by the observation that in addition to representing potentially important neuronal signals, changes in the firing rate can often have a crucial impact on a large number of standard statistical spike train analyses that require the assumption of a constant firing rate (e.g., Brody 1999; Grün et al. 2002; Schneider 2008). Therefore, the main aim is to present a statistical test of the null hypothesis of constant rate and a method that can estimate change points in the firing rate in order to divide a spike train into sections of approximately constant rate.

A statistical method that aims at detecting rate change points in neuronal spike trains should take into account several phenomena and challenges observed in empirical data (see also Fig. 1). First, distributions of inter spike intervals (ISIs) can be highly diverse, and rate changes can occur on different time scales. Second, other process parameters such as the variance are not known in practice. And finally, as one of the main issues in the present paper, neuronal spike trains have often been reported to show serial dependencies of low orders (Lowen and Teich 1991; Ratnam and Nelson 2000; Chacron et al. 2001; Nawrot et al. 2007; Farkhooi et al. 2009), implying that independence of ISIs can not necessarily be assumed in practice.

Serial correlations themselves have been proposed to be a crucial aspect of information transmission in neuronal spike trains, for example, by reducing variability of spike count through negative correlations, thus increasing signal detection efficiency by a post-synaptic neuron (Chacron et al. 2001; Ratnam and Nelson 2000; Chacron et al. 2004; Nawrot et al. 2007). Several models of neuronal information coding have been proposed that incorporate mechanisms for positive and/or negative serial ISI correlations (Avila-Akerberg and Chacron 2011; Schwalger and Lindner 2013; Shiau et al. 2015). The concept that serial correlations shape the way in which informative spike changes are detected by neuronal systems inspired us to develop a statistical analysis method that can detect rate changes in spike trains while assuming and incorporating serial ISI correlations. Specifically, our novel method can detect rate changes in spike trains with short range dependencies in which the covariance structure of life times is unknown and rate changes may occur at different time scales.

As a proof of principle, we applied our analysis to spike train recordings obtained from spontaneous activity of DA neurons in anaesthetized mice. These trains include more or less regular single spike or bursty patterns, (e.g., Bingmer et al. 2011; Schiemann et al. 2012, see Fig. 1c, d for examples), and the activity of this class of neurons has been previously described with spike train models with serial dependencies, such as stochastic cluster processes (Bingmer et al. 2011) or Hidden Markov Models (Camproux et al. 1996). The dataset presented here shows serial correlations between successive ISIs, which can be strong for small lags, but decay fast towards zero, in accordance with the literature on serial ISI correlations (Fig. 1).

For the detection of rate changes in point processes, methods were developed e.g., by Kendall and Kendall (1980), Csörgő and Horváth (1987), Steinebach and Zhang (1993), Gut and Steinebach (2002), Gut and Steinebach

Fig. 1 Spike trains with serial correlations of ISIs from the sample data set. **a** Positive correlations, (**c**, **d**) negative serial correlation of order one. **b** Serial correlations of ISIs, derived from mean in disjoint windows of 50 ISIs each, with 95 % confidence limits
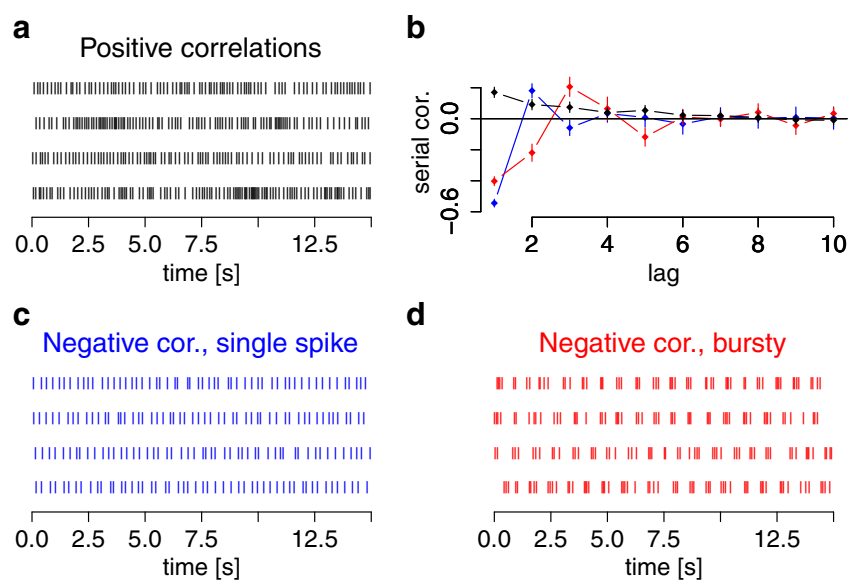
(2009), and Messer et al. (2014). In the context of stochastic time series, change point detection techniques that allow for dependencies have been developed by, e.g., Tang and MacNeill (1993), Lavielle (1999), Ray and Tsay (2002), Berkes et al. (2006), and Dehling et al. (2013). Further, several interesting methods that focus on the aspect of multi-scale detection have been proposed recently by Frick et al. (2014), Fryzlewicz (2014), and Messer et al. (2014).

Our novel method extends the multiple filter test (MFT) proposed in (Messer et al. 2014) to respecting weak dependencies in the ISIs. Regarding the point process model the ISIs are often referred to as the *life times* of the point events. The MFT was designed specifically for spike trains with a wide range of ISI distributions and multi scale rate changes. The idea in the corresponding filtered derivative approach is to study for every time point the difference between the number of spikes in the left and right window, scaled by an estimate of its standard deviation. Under specific assumptions, one obtains a limit process that is independent of all parameters of the underlying spike train. This limit process can be used to define rejection thresholds of the null hypothesis of constant rate and to estimate the rate change points. By simultaneous application of multiple moving windows, change points at multiple time scales can be detected. A corresponding algorithm (MFA, see Messer et al. 2014) can then be used to estimate the change points.

The idea behind extending the MFT to weak dependencies is based on the fact that under independence, the variance of the life times $\{\xi_i\}_{i\geq 1}$, $\mathbb{V}ar(\xi_1)$, is used as a scaling factor of the test statistic. If independence does not hold, this term needs to be replaced by

$$\rho^2 := \mathbb{V}ar(\xi_1) + 2\sum_{\ell=1}^{\infty} \mathrm{Cov}(\xi_1, \xi_{1+\ell}).$$

We then require a consistent estimate of ρ^2 in practical application. Here we focus on the practically important case of m-dependence, i.e., when $\mathrm{Cov}(\xi_1, \xi_{1+\ell}) = 0$ for all $\ell > m$, with some $m \in \mathbb{N}$, which yields consistency of the standard estimators for the summands of ρ^2.

The paper is organized as follows. We first review the ideas of the MFT assuming independence and the corresponding MFA for change point detection in Section 2.1. In Section 2.2 we derive a modification that can be applied to spike trains with weak dependencies. Section 2.3.1 gives examples of such theoretical processes to illustrate their correspondence to neuronal spike trains, particularly including also tonic and oscillatory bursty processes. Section 3 uses simulations to discuss estimation principles of ρ^2 and m and practical performance of the proposed method including also a recommendation for the choice of the window size. Particularly, we show that disrespecting serial correlations or globally estimating ρ^2 or m can yield erroneous

results, and illustrate improved performance of the modified MFT and MFA with regard to the number and location of change points. In Section 4, we apply the derived statistical method and algorithms to a data set of spike train recordings obtained from spontaneous activity of DA neurons in anesthetized mice.

2 Extension of the multiple filter test to weak dependencies

We consider a finite spike train of length $T > 0$ on the time interval $[0, T]$ as a sequence of spikes $0 < S_1 < S_2 < \cdots < S_{N_T}$, where N_t denotes the number of spikes up to time t. The ISIs are denoted by $\{\xi_i\}_{i\geq 1}$, with $\xi_1 = S_1$ and

$$\xi_i = S_i - S_{i-1} \quad \text{for} \quad i = 2, 3, \ldots, N_T.$$

The ISIs are considered realizations of random variables, and the aim is to construct a statistical test for the null hypothesis that the (positive) mean of all ISIs, i.e., the firing rate, is constant,

$$H_0: \ \mathbb{E}[\xi_i] = \mathbb{E}[\xi_1] =: \mu > 0 \quad \text{for all } i = 1, \ldots, N_T. \ (1)$$

For the alternative of k change points $c_1, \ldots, c_k \in [0, T]$, we assume $k+1$ (independent) processes with constant rates $\mu_1^{-1}, \ldots, \mu_{k+1}^{-1}$, while $\mu_j \neq \mu_{j+1}$ for all j. At time zero start in the first process with rate μ_1^{-1}, at the first change point c_1 jump into the second process of rate μ_2^{-1} etc. Then the resulting process is a piecewise combination of sections with different rates. If the null hypothesis is rejected, we are interested in estimating the change points $c_1, \ldots, c_k$ in order to segment the spike train into sections of constant rate.

2.1 The MFT for rate changes in renewal processes

Here we describe the main idea of the MFT (for more details see Messer et al. 2014). The MFT is based on a filtered derivative approach that compares the numbers of events, $N_{\mathrm{le}} := N_t - N_{t-h}$ and $N_{\mathrm{ri}} := N_{t+h} - N_t$ in the left and right window of size $h \in (0, T/2)$ for every time $t \in [h, T-h]$. By standardizing with a consistent estimator of the standard deviation of this difference, $\hat{s}_{h,t}$, one obtains a filtered derivative process

$$G_{h,t} := \frac{N_{\mathrm{ri}} - N_{\mathrm{le}}}{\hat{s}_{h,t}}. \quad (2)$$

Large differences between the numbers of events in the left and right window, i.e., large deviations of G from zero indicate deviations from the null hypothesis of constant rate. In order to test statistical significance of these deviations, the maximal deviation $\max_t |G_{h,t}|$ from zero could serve as a test statistic for one window, and the rejection threshold at

level α can be derived from the limit process of G as follows. Using an extension $G^{(n)} := G_{nh,nt}$ in an asymptotic setting in which the window size nh and the time nT (or alternatively, the firing rate) grow linearly in n, $G^{(n)}$ can be shown to converge weakly to a functional L of a standard Brownian motion W,

$$L_{h,t} := \left((W_{t+h} - W_t) - (W_t - W_{t-h}) \right) / \sqrt{2h}, \tag{3}$$

under the null hypothesis of a constant rate. Note that $L_{h,t} \sim N(0, 1)$ for all h and t, i.e., $\hat{s}_{h,t}$ standardizes the difference of the number of events in both windows. As L does not depend on parameters of the underlying process, the distribution of $\max_t |L_{h,t}|$ can be easily simulated to obtain a rejection threshold Q for a statistical test at level α.

In order to allow detection of change points at multiple time scales, the MFT combines multiple windows from a finite set H and the corresponding processes $G_{h,t}$. As the distribution of $\max_t |G_{h,t}|$ depends on the window size h, the process G is rescaled to give about the same weight to every window size, resulting in a rescaled process

$$R_{h,t} := \frac{|G_{h,t}| - \hat{\mu}_{M_h^*}}{\hat{\sigma}_{M_h^*}},$$

where $\hat{\mu}_{M_h^*}$ and $\hat{\sigma}_{M_h^*}$ denote the estimated mean and standard deviation of $M_h^* := \max_t |L_{h,t}|$ obtained in simulations by simulating W and deriving L from W as in Eq. (3). The maximum of all R-processes,

$$M := \max_{h,t} R_{h,t},$$

is used as a test statistic. The rejection threshold Q can then be derived from the corresponding distribution of $\max_h (M_h^* - \hat{\mu}_{M_h^*})/\hat{\sigma}_{M_h^*}$, which can be obtained in simulations.

This approach has three practical advantages: First, it does not require previous knowledge of process parameters because G is scaled such that the limit process does not depend on the parameters of the underlying process. Second, it can be applied to a wide range of processes, i.e., Poisson or Gamma processes or processes with complex or unknown ISI distributions as long as ISIs are independent and identically distributed (Steinebach and Eastwood 1995). It even holds for processes with independent but not necessarily identically distributed ISIs in the sense that the variance of ISIs may show a certain degree of variation between regular and irregular phases (Messer et al. 2014) as can sometimes be observed in empirical spike trains. Third, this approach allows the simultaneous use of multiple windows in a finite set H and thus, analysis of change points at multiple time scales. Due to the asymptotic nature of the method, the smallest window should contain at least about $100 - 200$ spikes in order to approximately keep the significance level.

The MFT is applied to a simulated spike train with three rate changes in Fig. 2. The upper panel indicates the rescaled processes for a window set $H = \{50, 100, 200\}$. The maximum M exceeds the rejection threshold Q, and the null hypothesis of constant rate is rejected. Then, the MFA successively estimates the change points. For every window h, change point candidates $\hat{c}_j$ are identified by successively locating the maxima of $(R_{h,t})_t$ and then deleting their h-neighborhood $[\hat{c}_j - h, \hat{c}_j + h]$. Change point candidates are then successively combined (see also the articles by Fryzlewicz (2014) and Frick et al. (2014) for similar approaches), preferring candidates of smaller windows and adding only those whose h-neighborhood does not overlap accepted change points. This is motivated by the idea that large windows tend to be affected by multiple change points, which may reduce their estimation precision. In Fig. 2, change points with fast, strong changes are estimated with small windows, while change points with slow and weak changes are estimated with larger windows.

2.2 The MFT for weak dependencies

The main purpose of this paper is to study the MFT in case of weak dependence of ISIs. We will show here that this requires two assumptions: First, a generalized class of point processes that also include weak dependencies (Definition 2.1), and second, consistent estimation of process parameters (Proposition 2.2).

Under independence, a consistent estimator of the standard deviation of $(N_{\mathrm{ri}} - N_{\mathrm{le}})$ in Eq. (2) is given by

$$\left(2nh\hat{\sigma}^2/\hat{\mu}^3 \right)^{1/2} \tag{4}$$

if $\hat{\mu} > 0$, and zero otherwise, where $\hat{\mu}$ and $\hat{\sigma}^2$ denote the empirical mean and variance of the ISI lengths in the
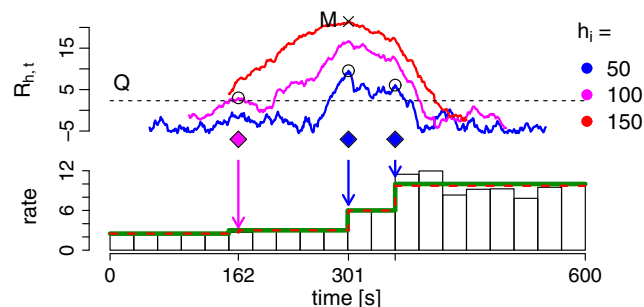

Fig. 2 Application of the MFT and MFA to a piecewise renewal process with Gamma-distributed intervals ξ_i with variance $\mathrm{Var}(\xi_i) = 0.2^2$ and rates 2.5, 3, 6, 10 Hz with change points at 150, 300, 360 seconds. Upper panel indicates rescaled processes $R_{h,t}$ for window sizes $h \in \{50, 100, 150\}$ and rejection threshold Q (*dashed*) at 5 % level, and lower panel indicates rate histogram, true rate (*thick, solid*) and estimated rate (*thin, dashed*) with estimated change points (*diamonds and arrows*). The color indicates the window with which the change point was detected

analysis window. In case of non-zero covariances between successive ISIs, the estimate of the variance of ISIs $\sigma^2 := \mathbb{V}ar(\xi_1)$ needs to be replaced by an estimator of

$$\rho^2 = \sigma^2 + 2 \sum_{\ell=1}^{\infty} \rho_\ell, \tag{5}$$

where $\rho_\ell := \text{Cov}(\xi_1, \xi_{1+\ell})$, yielding

$$\hat{s} := \hat{s}_{nh,nt} := \left(2nh\hat{\rho}^2/\hat{\mu}^3\right)^{1/2}, \tag{6}$$

where details on $\hat{\rho}^2$ and $\hat{\mu}$ can be found in Section 2.3.2. Here we show under which assumptions on the point processes and modifications of the MFT one obtains the same convergence and thus, applicability to spike trains with weakly dependent ISIs. To that end, we require a class of point processes $\mathscr{P}$ for which the ISIs fulfill a functional central limit theorem (FCLT) and for which consistency of $\hat{s}$ can be concluded.

Definition 2.1 The class of point processes $\mathscr{P}$ is given by all point processes on the positive line whose life times $\{\xi_i\}_{i \geq 1}$ are stationary, ergodic, almost surely positive and square-integrable and further they fulfill $\rho^2 > 0$ (see Eq. (5)) as well as

$$\sum_{i=2}^{\infty} \|\mathbb{E}[\xi_1 - \mathbb{E}[\xi_1] | \{\xi_k | k \geq i\}]\| < \infty. \tag{7}$$

Here, $\|\cdot\|$ denotes the L^2-norm (as the conditional expectation in Eq. (7) is a random variable). Stationarity means that the distribution of any subset of life times is invariant under a time shift of their indices. The assumptions on $\{\xi_i\}_{i \geq 1}$ particularly imply a FCLT as well as ergodic theorems. The FCLT will be used to derive the convergence of the filtered derivative process (Proposition 2.2) and the ergodic theorems will be used for consistent parameter estimation (Lemma A.1 and A.2). See (Billingsley 1999) for details on the notions of stationarity and ergodicity. Further note that the summation condition (7) implies absolute convergence of the series (5) (see Billingsley 1999, Thm. 19.1). This condition particularly holds true for the special case of m-dependent sequences.

Throughout this article, $\xrightarrow{d}$ denotes convergence in distribution, and $(D[0, \infty), d_{SK})$ denotes the space of càdlàg-functions on $[0, \infty)$ endowed with Skorokhod-topology, and analogous for $(D[h, T - h], d_{SK})$.

The following proposition ensures that the MFT can be applied to point processes $\Phi \in \mathscr{P}$ when their parameters are consistently estimated.

Proposition 2.2 *Let $\Phi \in \mathscr{P}$ with ISIs $\{\xi_i\}_{i \geq 1}$ and let $\hat{s}^2$ be an estimator for $s^2 = 2nh\rho^2/\mu^3$ that satisfies in $(D[h, T - h], d_{SK})$ as $n \to \infty$ that $(\hat{s}/s)_t \to (1)_t$ in probability.*

Then it holds for the filtered derivative process $G^{(n)} = (G_{nh,nt})_t$ as given in Eq. (2) *in $(D[h, T - h], d_{SK})$ as $n \to \infty$*

$$G^{(n)} \xrightarrow{d} L.$$

Proof From the conditions on $\mathscr{P}$ it follows that in $(D[0, \infty), d_{SK})$ as $n \to \infty$

$$\left(\frac{1}{\rho \sqrt{n}} \sum_{i=1}^{[nt]} (\xi_i - \mu)\right)_t \xrightarrow{d} W, \tag{8}$$

where W denotes a standard Brownian motion (see Billingsley 1999, Thm. 19.1). For $t \geq 0$ let

$$Z_t^{(n)} := (N_{nt} - nt/\mu)/(n\rho^2/\mu^3)^{1/2}$$

denote the rescaled counting process. According to Vervaat (1972) it follows from Eq. (8) that in $(D[0, \infty), d_{SK})$ as $n \to \infty$ it holds $Z^{(n)} \xrightarrow{d} W$. We then define a continuous map $\varphi_h : (D[0, \infty), d_{SK}) \to (D[h, T - h], d_{SK})$ via $f(t) \xmapsto{\varphi_h} ((f(t + h) - f(t)) - (f(t) - f(t - h)))/(2h)^{1/2}$. By continuous mapping theorem it follows in $(D[h, T - h], d_{SK})$ for $n \to \infty$ that

$$((N_{\text{ri}}^{(n)} - N_{\text{le}}^{(n)})/(2nh\rho^2/\mu^3)^{1/2})_t \xrightarrow{d} L,$$

where $N_{\text{ri}}^{(n)} := N_{n(t+h)} - N_{nt}$ and $N_{\text{le}}^{(n)} := N_{nt} - N_{n(t-h)}$. Due to the consistency assumption of the estimator $\hat{s}$, we can exchange $(2nh\rho^2/\mu^3)^{1/2}$ with $\hat{s}$ by Slutsky's theorem. $\qquad \square$

2.3 Examples for practical application

Proposition 2.2 states that the MFT is applicable to processes in the class $\mathscr{P}$ if one uses the modified filtered derivative process

$$G_{h,t} := \frac{N_{\text{ri}} - N_{\text{le}}}{\hat{s}_{h,t}}, \tag{9}$$

with $\hat{s}^2$ a consistent estimator of $s^2 = 2h\rho^2/\mu^3$ and $\rho^2 = \sigma^2 + 2 \sum_{\ell=1}^{\infty} \rho_\ell$ with the convention $G_{h,t} := 0$ if $\hat{s}_{h,t} = 0$. In order to illustrate practical applicability specifically to spike trains with weakly dependent life times, we give examples of processes in $\mathscr{P}$ that resemble empirical spike trains (Section 2.3.1) and examples of consistent estimators of s (Section 2.3.2).

2.3.1 Processes in $\mathscr{P}$

The assumptions of processes in $\mathscr{P}$ are fulfilled for example in renewal processes with independent and identically distributed ISIs. Here, we focus on dependencies in the ISI structure, i.e., processes with stationary and ergodic ISIs as stated in Definition 2.1. In a simple but practically important case, the ISIs are m-dependent for an $m \in \mathbb{N}$, i.e., $\rho_\ell = 0$ for all $\ell > m$.

Here we give three examples of m-dependent processes in Fig. 3 that resemble the neuronal spike trains shown in Fig. 1. Panel a shows a process with $m = 3$ and positive serial correlations given by life times

$$\xi_i := a_0 X_i + a_1 X_{i-1} + \ldots + a_m X_{i-m}, \qquad (10)$$

with $X_1, X_2, \ldots$ independent with expectation μ_X and variance $\sigma_X^2 > 0$. This implies

$$\sigma^2 = \mathrm{Var}(\xi_i) = \sigma_X^2 \sum_{j=0}^{m} a_j^2$$

and $\rho_\ell = \sigma_X^2 \sum_{j=0}^{m-\ell} a_j a_{j+\ell}$ for $\ell \leq m$, and $\rho_\ell = 0$ for $\ell > m$ i.e.,

$$\rho^2 = \sigma_X^2 \left(\sum_{j=0}^{m} a_j^2 + 2 \sum_{\ell=1}^{m} \sum_{j=0}^{m-\ell} a_j a_{j+\ell} \right).$$

Appropriate conditions on the a_i and the distribution of X_i ensure almost surely positive ISIs.

Panel c shows an example of a single spike process similar to Fig. 1c. It is described by ISIs

$$\xi_i = U_i + Z_i - Z_{i-1}, \qquad (11)$$

where U_i, Z_i are independent and uniformly distributed with $U_i \sim U[\nu - \sigma_1, \nu + \sigma_1]$ and $Z_i \sim U[-\sigma_2, \sigma_2]$, with $\nu, \sigma_1, \sigma_2 > 0$ and $\sigma_1 + 2\sigma_2 \leq \mu$, which assures $\xi_i > 0$. In this process, all ISIs ξ_i are identically distributed with mean ν and variance $\sigma^2 = (1/3)(\sigma_1^2 + 2\sigma_2^2)$. The process is 1-dependent with negative covariance of lag one given by $\rho_1 = -\mathrm{Var}(Z_i) = -(1/3)\sigma_2^2$ (panel b). The spikes of this process can be regarded as jittered uniformly with jitter Z_i around the unobservable beats of a background rhythm with period ν which is a renewal process with independent and

uniformly distributed intervals U_i. Related doubly stochastic Cox processes have been used earlier for the description of single spike processes (Bingmer et al. 2011). Similar to Hidden Markov Models (Camproux et al. 1996), they can also be used for the description of oscillatory bursty activity as in Fig. 1d.

In order to illustrate applicability of Proposition 2.2 also to oscillatory bursty spike trains, Fig. 3d shows an example of a 2-dependent oscillatory bursty process similar to the spike train in Fig. 1d. Every ISI ξ_i is described by

$$\begin{aligned} \xi_i = &\ I_i(1 - I_{i-1})X_i + I_{i-1}J_i Y_i + I_{i-2}J_i' Y_i' \qquad (12) \\ &+ (1 - \max(I_i(1 - I_{i-1}), I_{i-1}, I_{i-2}J_i))Y_i'', \end{aligned}$$

where $(I_i)_{i\geq1}$, $(J_i)_{i\geq1}$, $(J_i')_{i\geq1}$ are independent sequences of independent $\{0, 1\}$−valued random variables with success probabilities p_I and $p_J = p_{J'}$, and $(X_i)_{i\geq1}$, $(Y_i)_{i\geq1}$, $(Y_i')_{i\geq1}$ and $(Y_i'')_{i\geq1}$ are independent sequences of independent and almost surely positive random variables and Y_i, Y_i', Y_i'' are identically distributed for all i. Obviously, all ISIs are identically distributed and the process is 2-dependent. The idea is that X_i takes large values to generate large ISIs, while Y_i, Y_i', Y_i'' take small values. Then, an ISI ξ_i takes a large value if $I_i = 1$ and $I_{i-1} = 0$, such that in this example, a long ISI is typically followed by at least one short ISI, leading to negative serial correlation (panel b). The last summand in Eq. (12) only ensures that $\xi_i > 0$.

2.3.2 Consistent estimators

In addition to requiring a process in $\mathscr{P}$, the second ingredient of Proposition 2.2 is a consistent estimator $\hat{s}$. Common approaches in the setting of dependencies

Fig. 3 Examples of processes from $\mathscr{P}$ (panels **a,c,d**) that resemble spike train patterns from Fig. 1, and corresponding covariance structure (**b**). a. A 3-dependent process according to Eq. (10), with X_i Gamma distributed and $\mu = 0.2$, $\sigma = 0.1$, $a_k = c^k$, $c = 0.25$. c. A 1-dependent process similar to Fig. 1c, simulated according to Eq. (11), with $\mu = 0.3$, $\sigma_1 = 0.06$, $\sigma_2 = 0.12$. d. A 2-dependent process similar to Fig. 1d, simulated according to model (12), with $p_I = .5$, $p_J = .4$, $X_i \sim U[0.45, 0.73]$, $Y_i \sim U[.01, .12]$
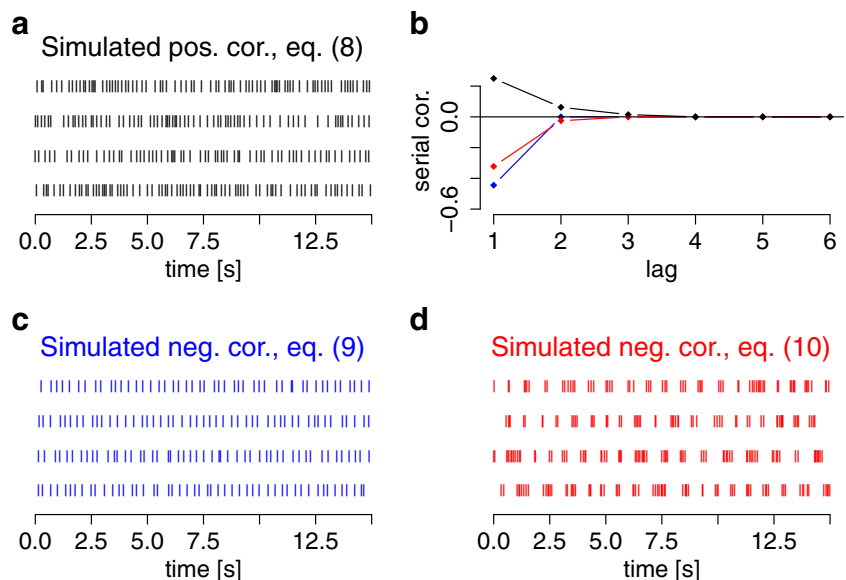
include methods based on covariance kernel estimation (De Jong and Davidson 2000, e.g.,Wied et al. 2012) or the Bartlett-estimator (Berkes et al. 2005; Xiao and Wu 2012; Kirch and Muhsal 2014). Here we focus on two simple estimators - a *global* and a *local* estimator - that are particularly useful in practical application. Under m-dependence, we show consistency under the null hypothesis of constant rate, even for the local estimator. The local estimator is particularly useful in the presence of change points because the global estimator is sensitive to rate changes and therefore tends to be biased in these cases. In contrast, the local estimator does not tend to be biased on most time sections (see Section 3).

In case of m-dependence ρ^2 equals a finite sum

$$\rho^2 = \sigma^2 + 2\sum_{\ell=1}^{m} \rho_\ell. \tag{13}$$

The global estimator uses global estimates of the variance and covariances in Eq. (13) from the entire spike train using standard estimators

$$\hat{\rho}_\ell := \left(\frac{1}{N_{nT} - (\ell+1)} \sum_{i=1}^{N_{nT}-(\ell+1)} \xi_i \xi_{i+\ell} \right) - \hat{\mu}^2, \tag{14}$$

$$\hat{\rho}^2 := \hat{\sigma}^2 + 2\sum_{\ell=1}^{m} \hat{\rho}_\ell, \tag{15}$$

where $\hat{\mu}$ denotes the empirical mean of all ISIs. Lemma A.1 in the Appendix shows that this yields a consistent estimator

$$\hat{s}^2 := 2hn\hat{\rho}^2/\hat{\mu}^3 \tag{16}$$

under the null hypothesis.

As mentioned above, one main disadvantage of global parameter estimation is that it tends to be biased under the
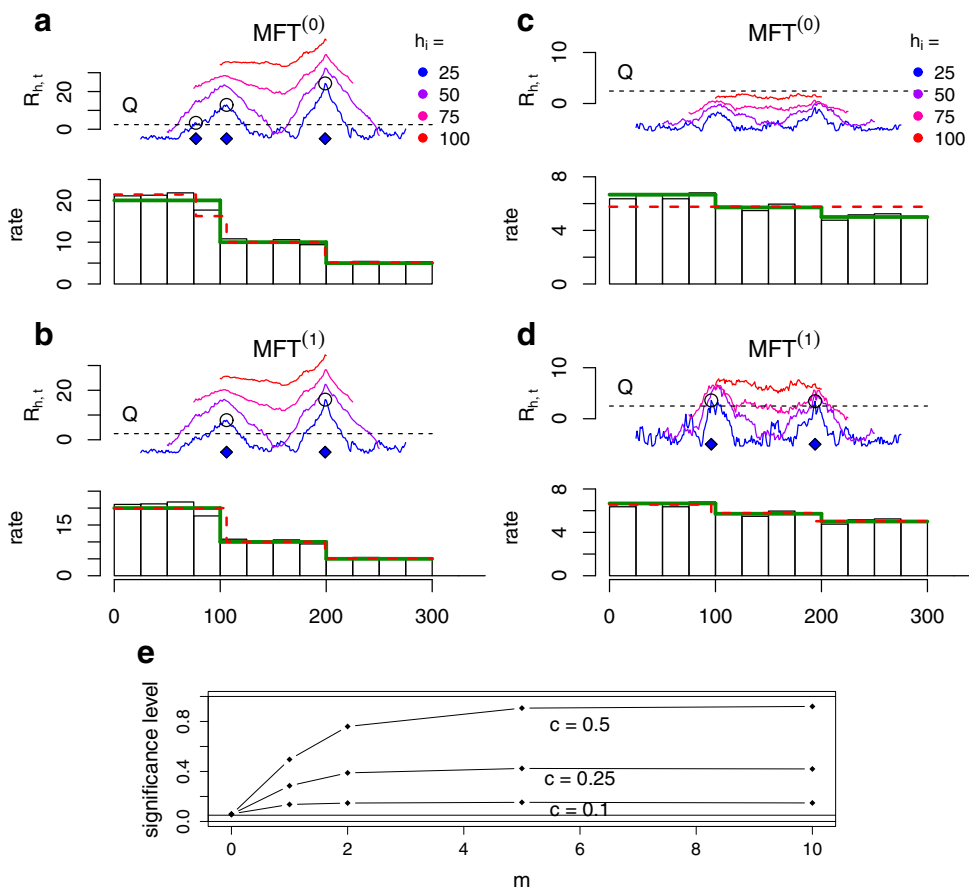

Fig. 4 The classical MFT$^{(0)}$ can estimate too many change points when applied to processes with positive correlations (**a**,**e**) and may show reduced test power when applied to negative correlations (**c**). True rate profiles indicated in thick solid, estimated rate profiles in thin, dashed. Here, the MFT$^{(1)}$ (**b**,**d**) detects all true change points and no false positives. e. Significance level of classical MFT$^{(0)}$ for positive serial correlations obtained in 10000 simulations, where the threshold was chosen such that under independence, the MFT$^{(0)}$ would yield an asymptotic significance level of 5 % (indicated by horizontal line). All simulations were performed using model (10) with $T = 300$, $H = \{25, 50, 75, 100\}$ and $m = 1$ in (**a-d**) and varying m in e. The coefficients a_i were $a_0 = 1$ throughout and $a_1 = 0.5$ in a, b, $a_1 = -0.5$ in c,d and $a_k = c^k, c \in \{0.1, 0.25, 0.5\}$ in e. The X_i were Gamma-distributed in a,b,e and, in order to ensure a.s. positive ISIs, uniformly distributed in c,d. The parameters μ_X, σ_X were chosen such that the ISIs ξ_i had standard deviation 0.15 and the given rate profiles (**a-d**) or $\mu = 0.1$ (**e**)

alternative hypothesis (see Section 3 and Fig. 6d). There-fore, we suggest to use an analogous local estimator, which for every t uses only the ISIs in the window $(n(t-h), n(t+h)]$. More precisely, for every time t, we estimate ρ^2 and μ analogously, but only from the life times that lie within the windows (separate estimation for the left and the right window) and let the local estimator be

$$\hat{s}^2 := \left(\frac{\hat{\rho}_{\mathrm{ri}}^2}{\hat{\mu}_{\mathrm{ri}}^3} + \frac{\hat{\rho}_{\mathrm{le}}^2}{\hat{\mu}_{\mathrm{le}}^3} \right) nh. \tag{17}$$

For the case of independent life times, i.e., $m = 0$, consistency of this estimator was shown in Messer et al. (2014). In Lemma A.2 in the Appendix we show consistency of this estimator for m-dependent processes.

3 Practical application of the MFT for weak dependencies

Section 2 presented theoretically a class of processes, estimators and statistics that allow to apply the MFT and MFA for the estimation of rate change points in spike trains with weakly dependent ISIs. Here we use simulations to illustrate the difference between the proposed method and the classical MFT that assumes independence. In addition, we discuss the important practical issue of estimating serial dependencies and of choosing the set of windows H, particularly the smallest window. Simulations are performed using models (10) and (11), which yield flexible and simple formulas for serial correlations.

For ease of notation we denote the MFT and MFA that assume m-dependence by MFT$^{(m)}$ and MFA$^{(m)}$. The classical MFT assuming independence will therefore be denoted by MFT$^{(0)}$. All procedures use the statistic described in Eq. (9). Under m-dependence, ρ^2 is estimated up to the m-th summand in the MFT$^{(m)}$. The corresponding estimator of s is denoted by $\hat{s}^{(m)}$.

First, we show that falsely applying the classical MFT$^{(0)}$ yields too many false positives in cases of positive correlations and reduced test power for negative correlations. This is because MFT$^{(0)}$ uses $\hat{\rho}^2 := \hat{\sigma}^2$, disrespecting potential serial correlations. Positive correlations yield $\rho^2 > \sigma^2$ and thus increase the number of false positives in the MFT$^{(0)}$ when the scaling $\hat{s}^{(0)}$ is spuriously low (Figs. 4a, e and 5b). Vice versa, negative correlations yield conservative results for the MFT$^{(0)}$ and a reduced test power (Fig. 4c), while in the given example, the MFT$^{(m)}$ can detect the given change points with high precision (Fig. 4d).

Second, we illustrate the performance of the MFA$^{(m)}$ when m is known using the standard estimators $\hat{s}^{(m)}$ from Section 2.2 and emphasize that s should be estimated locally. In particular, we propose to use the local estimator
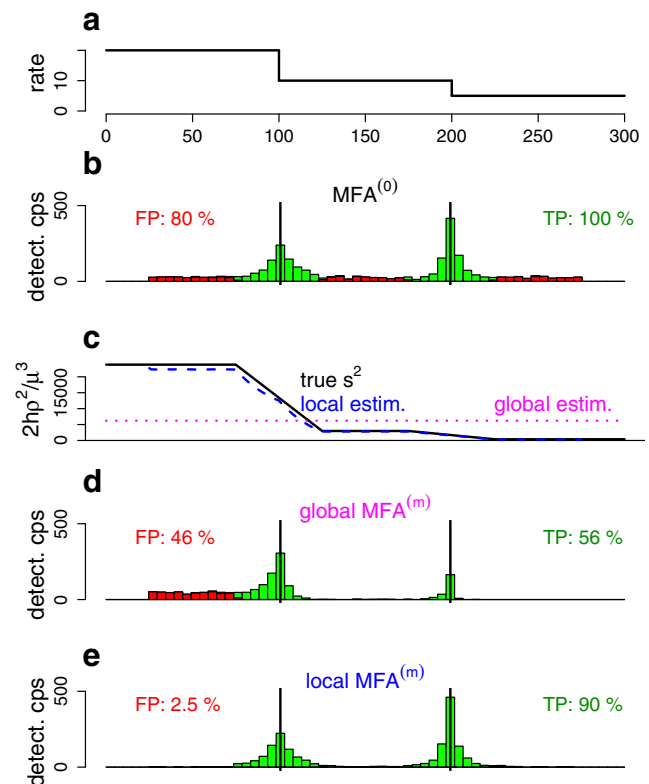

Fig. 5 Number of detected change points in 1000 simulations under the alternative with rate profile given in a. b. MFA$^{(0)}$ without accounting for serial correlations. c. True value of local s^2 (*solid*) compared with local estimation (*dashed*) and global estimation (*dotted*). d. MFA$^{(m)}$ with global estimate of s. e. MFA$^{(m)}$ with estimates of s derived separately in every analysis window. Simulations according to model (10), with X_i Gamma distributed, $a_k = 0.5^k$, $T = 300$, $H = \{25, 50, 75, 100\}$, $\sigma = 0.15$, $m = 3$, $c = 0.5$

from Eq. (17) because a global estimator (Eq. (16)) would be biased in case of rate changes and thus, reduce test power and/or increase the number of false positives. This effect is illustrated in Fig. 5. Spike trains with positive serial correlations are simulated with a rate profile with two change points (panel a). As described above, the classical MFA$^{(0)}$ assumes independence and therefore shows many false positives (panel b). Using the MFA$^{(m)}$ with global estimation of s is also unsatisfactory as is shows increased false positive rate on the left and decreased detection rate on the right (panel d). This is because the rate changes cause the true value of s^2 to change across time (panel c). The global estimate (dotted) falsely uses a global μ and therefore a biased global estimate of ρ^2 (see also Fig. 6d) and thus underestimates s^2 on the left and overestimates on the right. In contrast, the estimates from local windows (blue) correspond closely to the true value of s^2, and accordingly, the corresponding MFA$^{(m)}$ using local estimators detects the change points with high precision without showing an increased false positive rate (panel e). For individual
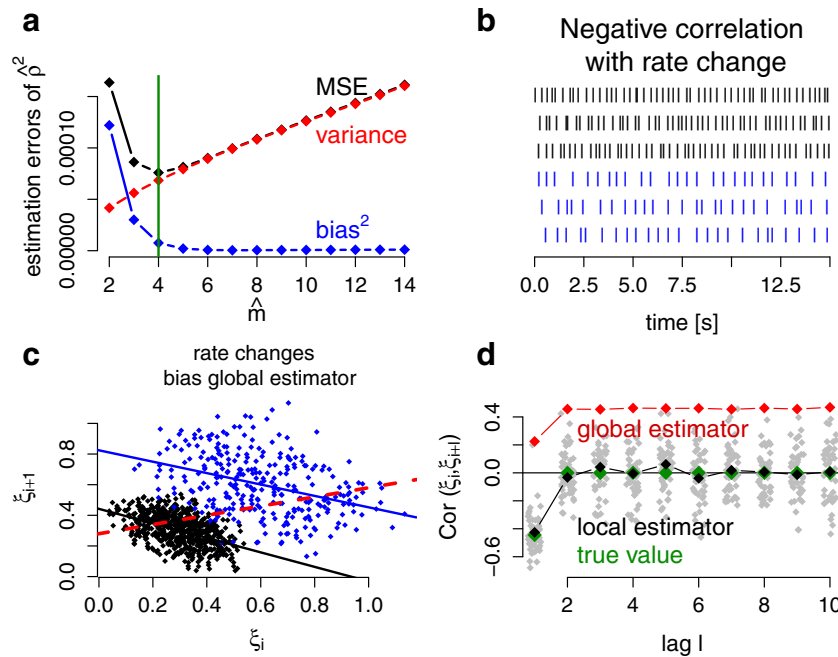
Fig. 6 **a**. Estimation errors of $\hat{\rho}^2$ for different values of m. Simulations according to model (10), with X_i Gamma distributed, $a_k = 0.5^k$, $T = 300$, $\mu = 0.15$, $\sigma = 0.15$, $m = 7$. Serial correlation decreases exponentially with the lag, such that the summands with large lags show only small contributions to ρ^2. Neglecting these in the estimation of ρ yields a small bias but a highly reduced variance. **b**. A spike train with negative first order serial correlation but a rate change. **c**. First order correlation is negative locally within sections of constant rate (*dots and solid regression lines*), but positive globally across sections (*dashed regression line*). **d**. Estimates of serial correlations of lags $\ell \in \{1, 2, \ldots, 10\}$ for the spiketrain shown partly in *b*. Global estimates (*red*), local estimates (*grey*) derived in disjoint windows of length 50 ISIs, their medians (*black*) and true serial correlations (*green*)

examples with positively or negatively correlated life times in the case of 1-dependence see also Fig. 4b and d).

Third, we discuss the estimation of m, which is typically unknown in practice. If a spike train was arbitrarily long, we could simply use all serial correlations up to an arbitrarily large order as $\rho_\ell = 0$ for $\ell > m$, which does not bias the estimation of ρ (this is the idea behind approaches for consistent estimation under long-range dependence, see De Jong and Davidson 2000; Berkes et al. 2005; Wied et al. 2012; Xiao and Wu 2012). However, in practice, this approach is not applicable because for finite spike trains it highly increases the variance of $\hat{\rho}$ and thus, the probability of over- or underestimating ρ, whereas the former decreases the test power and the latter increases the number of false positives. Therefore, it is important to include only the largest summands into the estimation of ρ^2, while summands with smaller contributions can be neglected. This effect reduces the mean squared error (MSE) of $\hat{\rho}^2$ by introducing small bias but reducing variance as shown in Fig. 6a where for $m = 7$, $\hat{m} = 4$ yields the smallest MSE.

Therefore, we consider here only the practically important case in which serial correlations decrease with the lag, and propose to search the smallest lag ℓ^* for which the serial correlation is not significantly different from zero (e.g. on

the 5 % level) and to use $\hat{m} = \ell^* - 1$ as an estimate of m. As before, the evaluation of statistically significant deviations from zero must be based on local estimates because potential rate changes can bias the estimates of serial correlations as illustrated in Fig. 6b–d. Panel b shows a simulated spike train according to model (11) with negative first order serial correlations, i.e., $\rho_1 < 0$, and a rate change point in the middle. The corresponding successive ISIs ξ_i, ξ_{i+1} on which the estimation of ρ_1 is based are shown in panel c. The global estimate of ρ_1 is not even negative but positive (dashed line in c), whereas the true correlation is indicated by the blue and black lines with negative slope; a phenomenon known as Simpson's paradox.

We therefore propose to estimate m by splitting up the process into disjoint sections. In each section, serial correlations up to a maximal lag are calculated, and systematic deviations from zero are investigated for each lag. These sections should be long enough to provide good estimates of serial correlations, and small enough so that most windows remain unaffected by potential change points. In Fig. 6d, the estimates derived from the local estimators in small windows (black and grey dots) agree well with the true correlation structure (green) of the spike train shown in panel b, whereas the global estimators (red) are highly biased.

Finally, we investigate the practical applicability of the proposed procedure to finite windows as it relies on asymptotic thresholds. As mentioned earlier (see also Fig. 9 in Messer et al. 2014), simulations suggest that the MFT$^{(0)}$ keeps the asymptotic significance level if the smallest window contains about $100 - 200$ spikes for spike trains with medium regularity, i.e., if σ^2/μ^3 is not too small. If we assume additional covariance structure, we need to consider the term ρ^2/μ^3 instead, which basically determines the asymptotic value of the denominator of $G_{h,t}$. If it takes values close to zero, estimation error may lead to negative estimates of ρ^2/μ^3, in which case $\hat{s}$ and $G_{h,t}$ would be not defined. In addition, estimates of ρ^2/μ^3 in the neighborhood may be positive, but extremely small, causing sharp peaks in $G_{h,t}$ and therefore, false positives, particularly when using smaller windows (Fig. 7b, red curve with estimated change point). This needs to be taken into account in practice because negative serial correlations may yield very small ρ^2/μ^3. We therefore suggest to slightly modify the MFA by excluding the h-neighborhood of points in which the denominator of $G_{h,t}$ is not defined by setting $\hat{s} := 0$ in this neighborhood (such that G is also set to zero in this case, Fig. 7b, green curve). As this has asymptotically no effect, $\hat{s}$ remains consistent. The empirical significance level of this modified MFT$^{(m)}$ is investigated by application to the three simulated spike trains from Fig. 3 by varying the minimal window size. Figure 7a shows that in these simulations, again about $150 - 200$ spikes are required to approximately reach the asymptotic significance level.
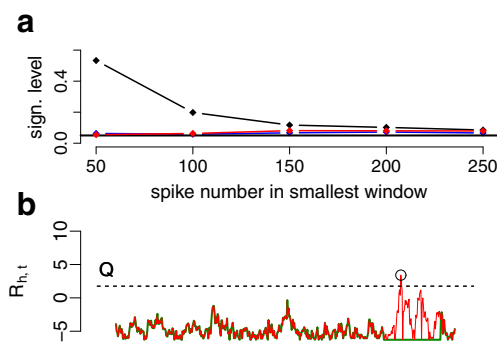

Fig. 7 Practical applicability to finite data sets and window choice. **a**. Empirical significance level of MFT$^{(m)}$ applied to simulations of spike trains from Fig. 3. For positively correlated ISIs (*black*), the significance level approaches the asymptotic 0.05 (*horizontal line*) when increasing the smallest window and thus, the spike number. For negatively correlated ISIs (*red, blue*), cutting out the h-neighborhood of points with undefined $\hat{s}$, which occurs particularly in small windows, reduces the number of false positives. **b**. Illustration of test modification. By cutting out the h-neighborhood (*green curve*) around falsely detected change points (*red curve*) caused in the neighborhood of points with undefined $\hat{s}$, false positives are reduced for small analysis windows, particularly when ρ^2/μ^3 is close to zero

4 Application to spike train recordings

We apply the proposed methods, principles and algorithms to an experimental data set of spike trains obtained from spontaneous activity recordings of dopaminergic neurons in the substantia nigra and ventral tegmental area of anaesthetized mice, as described previously (Schiemann et al. 2012; Subramaniam et al. 2014). The data set contains 44 spike trains of length 600 seconds, with a mean rate of about 4 spikes per second. The set of analysis windows was therefore chosen as $H = \{50, 75, 100\}$ seconds, yielding an expected number of about 200 spikes in the smallest window.

We estimated the maximal lag $\hat{m}$ for every spike train separately as described in Section 3 (Fig. 6d). To that end, we used disjoint windows of 50 ISIs to estimate serial correlations, and estimated $m + 1$ as the first lag for which deviations from zero were not significant on the 5 %-level using a Wilcoxon test. Figure 8a shows a typical example for one spike train. The serial correlation of lag one shows considerable deviation from zero, the correlation of lag two is small but still deviating from zero, and all other correlations do not strongly deviate from zero, leading to $\hat{m} = 2$ for this spike train. The corresponding estimates of serial correlations up to $\hat{m}_i$ are shown in panel b for all spike trains. The values of $\hat{m}$ were $\hat{m} \leq 3$ in about 90 % of all cases, ranging up to a maximum of 7, and the estimated serial correlation tended to be negative in the majority of spike trains.

In this more frequent case of negative serial correlations, the MFA$^{(\hat{m})}$ typically detected more change points than the MFA$^{(0)}$, leading also to rate profiles matching better with visual inspection (D-F). In order to measure this effect as a function of the degree of serial correlations, we plotted the difference between the number of change points estimated by the MFA$^{(0)}$ and by the MFA$^{(\hat{m})}$ in panel c as a function of an estimate of the term

$$2 \sum_{\ell=1}^{m} \mathrm{Cor}(\xi_i, \xi_{i+\ell}) = \frac{\rho^2 - \sigma^2}{\sigma^2}, \qquad (18)$$

which measures the contribution of serial correlations to ρ^2. As expected, when this term is negative, the MFA$^{(0)}$ typically estimated much fewer change points, often none at all. In the rare cases where this term was positive, the MFA$^{(0)}$ typically estimated more change points than the MFA$^{(\hat{m})}$.

5 Discussion

We have presented a multiple filter test (MFT) that can test the null hypothesis of constant firing rate and estimate change points in the rate of spike trains especially if these show dependencies in their ISI structure as is often observed

experimentally. Detection of subtle rate changes can be used for extracting meaningful signals from neuronal spike trains and, more generally, it can be an important preprocessing step for statistical analyses that are sensitive to rate changes.

Our procedure incorporates multiple features that are particularly important for practical application in spike train analysis: (1) un unknown number of rate changes can occur (2) on multiple time scales, (3) other process parameters such as the variance of inter spike intervals can be unknown, and (4) processes can show a high variety of patterns and distributions, including particularly serial dependencies.

The initial version of the MFT for rate change detection introduced in Messer et al. (2014) was developed for renewal processes with a wide range of life time distributions but assumed independence of ISIs, which does often not hold in empirical neuronal spike trains. The MFT uses a filtered derivative process with multiple filters that converges weakly to a parameter free limit process that can be used to obtain the rejection threshold for the test. By specifically estimating serial dependencies in the test statistic, we show that the new MFT can be applied to a variety of empirical firing patterns, including positive and negative serial correlations as well as tonic and bursty firing. Note that the conditions for the present new MFT include models where the life times are independent or where the life times are dependent but show no serial correlations. In these cases the results of the present MFT would be identical to the results

of the original MFT (Messer et al. 2014). This is because zero serial correlation implies that $\rho^2 = \sigma^2$ i.e., the terms that are responsible for the difference in the methods are identical.

For practical application, it is necessary to estimate the denominator of the test statistic, s, consistently. We have therefore proposed a consistent local parameter estimator under m-dependence. Although more complex theoretical approaches for consistent estimation are available for the more general case of ergodicity (Berkes et al. 2005; Wu and Pourahmadi 2009; Xiao and Wu 2012; Kirch and Muhsal 2014), we focus on m-dependence because it is technically simple and suitable for empirical data analysis with finite spike numbers. Especially under the alternative of rate changes, global estimators of s are affected by rate changes and yield erroneous results. Therefore, our simulations argue strongly for local estimates of s within small windows as these are less affected by potential change points. Even these local estimators require that m is small relative to the window size used for estimation. This implies that even under m-dependence the performance can be suboptimal if m is large and change points occur frequently. This is because large m requires large windows with constant firing rate for the estimation of s. If change points occur frequently, such windows cannot be found, and consequently, $\hat{s}$ will be affected by change points within the used estimation windows. Therefore, in practice, only cases with a moderate

Fig. 8 Application of the MFA to a data set of spike trains with weak serial correlations. **a.** Serial correlations estimated in disjoint windows of 50 life times each (*grey*), and medians (*black*). Vertical line indicates cutoff value for $\hat{m}$ for the respective spike train. **b.** Serial correlations are short and typically negative. Median serial correlations derived as in A for all spike trains, plotted up to the respective estimate m_i for every spike train i. **c.** Difference between the number of change points estimated by the MFA$^{(0)}$ and the MFA$^{(\hat{m})}$, as a function of the contribution of the serial correlations to ρ^2 (Eq. (18)). **d.** Application of the MFA$^{(0)}$ and **e.** the MFA$^{(\hat{m})}$ to one spike train with correlation profile similar to **a. f.** The rate profile of the sample spike train, and the rate profiles estimated by the MFA$^{(0)}$ (red, one estimated change point) and the MFA$^{(\hat{m})}$ (blue, six estimated change points)
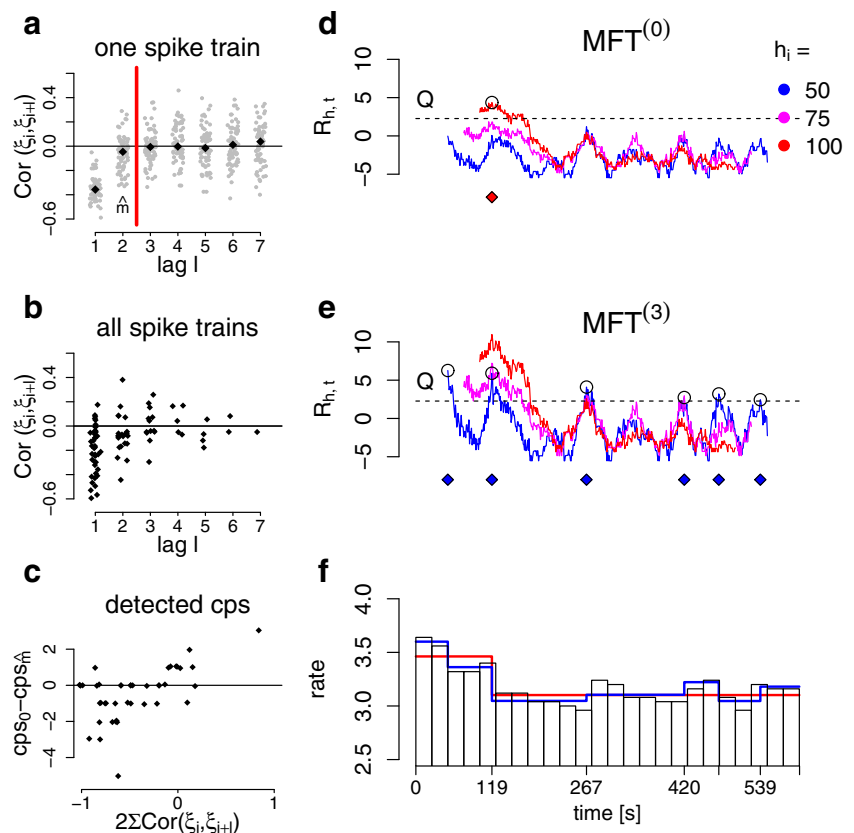
number of change points and short range dependencies can be considered, i.e., when m is small or serial correlations decay fast with the lag. According to practical examples such as the data set used here (e.g., Fig. 1 and Ratnam and Nelson 2000; Chacron et al. 2001; Nawrot et al. 2007; Farkhooi et al. 2009), this is a typical case for empirical spike trains.

One practical limitation of the presented method is its asymptotic nature, which requires a sufficient number of spikes, i.e., about $100 - 200$ events in the smallest window, which prevents change point detection in shorter time scales. Therefore, it can be considered particularly useful for spontaneous activity, rather than for short trials with many external stimuli or behavioral events. For these cases, different methods such as for example point process adaptive filter methods (e.g., Eden et al. 2004) may be useful. The main problem with using smaller window sizes is that the asymptotic threshold, Q, is too low when the smallest window does not contain sufficiently many spikes. One possibility to deal with this issue could be to replace Q by a threshold Q_b derived from a block bootstrap procedure (Singh 1981; Gonçalves and Politis 2011; Kreiss and Lahiri 2012), where the block size needs to be chosen such that serial correlations can be treated appropriately. In our simulations of the spike trains used in Fig. 7a, a block bootstrap procedure kept the asymptotic significance level by increasing the rejection threshold Q_b (data not shown). However, while Q always depends only on the window set H and the time T, under the alternative hypothesis of change points, Q_b largely depends on the properties of the spike train. This can render interpretation difficult in case of change points. Bootstrap can be advantageous when ρ^2 is close to zero, for example due to strong negative correlations, such that large amounts of data would need to be excluded from the analysis due to negative estimates of $\hat{s}$, potentially also including the change points themselves. In such cases, bootstrap procedures can enhance detection probability by avoiding this exclusion. In other cases, detection probability can be reduced, which often makes the use of small windows equally unsatisfactory for bootstrap procedures. In addition, the derivation of Q_b takes considerably longer than the derivation of Q. We therefore recommend to use the asymptotic threshold and a minimal spike number of about $100-200$ events in the smallest window, but bootstrap options are also made available in the provided code.

As a second limitation, the present method assumes the rate to be a step function with clear change points. As a consequence, other forms of the rate function, such as ramps or rhythmic behavior, will be described by corresponding step functions.

Our simulations illustrate the necessity of incorporating serial correlation in the MFT. For positive correlations, our new MFT is necessary to reduce the number of

false positives, which can be highly enhanced when falsely assuming independence. For the frequent case of negative correlations, these reduce the variability of the spike count and therefore enhance the detection probability of change points, yielding a higher potential of signal extraction from noisy spike trains. Indeed, it has been suggested that sensorial neural systems, such as the electroreceptive organs of weakly electric fish (Chacron et al. 2001) and primary somatosensory cortical neurons in rats (Nawrot et al. 2007) use this feature to increase their information transfer capacity. In this, our method takes into account a feature of information transfer in point processes with a direct correlate in the actual function of neuronal circuits.

In order to illustrate the performance of the method, we have applied the new MFA$^{(\hat{m})}$ to a data set of empirical spike trains and compared its performance to the classical MFA$^{(0)}$ that falsely assumes independence of ISIs. For all spike trains, serial correlations of small orders were estimated by using small windows to account for potential bias caused by rate changes. In the rare case of positive correlations, the classical MFT$^{(0)}$ that falsely assumes independence detected up to twice as many change points as the new MFT$^{(\hat{m})}$. In the more typical case of negative serial correlations, the new MFT$^{(\hat{m})}$ detected many more change points than the MFT$^{(0)}$. The new MFT$^{(\hat{m})}$ then yielded rate profiles matching better with visual inspection, indicating a higher detection power of potential neuronal rate signals. Potential applications of our novel algorithm include the extraction of information-rich signals from noisy spike trains, especially when there are no clear behavioral or sensorial triggers, e.g. spontaneous activity recordings. It can also potentially be used as a pre-processing step for other statistical analyses, and for detecting long-term but subtle rate changes, which may reflect transitions of neuromodulatory states (Lee and Dan 2012).

Compliance with Ethical Standards

Appendix: A Proofs

Here we show consistency of the estimators $\hat{s}^2$ of s^2 in Eqs. (16) and (17). Recall that these were

global estimator: $2h\hat{\rho}^2/\hat{\mu}^3$ (see Lemma A.1)

local estimator: $\left(\dfrac{\hat{\rho}_{ri}^2}{\hat{\mu}_{ri}^3} + \dfrac{\hat{\rho}_{le}^2}{\hat{\mu}_{le}^3}\right)h$ (see Lemma A.2)

The used estimators $\hat{\rho}, \hat{\mu}, \hat{\rho}_{\text{le}}, \hat{\rho}_{\text{ri}}, \hat{\mu}_{\text{le}}, \hat{\mu}_{\text{ri}}$ are the empirical means and estimates of ρ given in Eq. (14), derived from the whole process in the global estimator and from the local right and left windows at time t in the local estimator.

Lemma A.1 *Let $\{\xi_i\}_{i \geq 1}$ be an m-dependent process in $\mathscr{P}$ and $(\hat{s}_{nh,nt}^2)_t$ the global estimator as in Eq. (16). Then it holds in $(D[h, T - h], d_{\|\cdot\|})$ almost surely as $n \to \infty$ that*

$$(\hat{s}_{nh,nt}^2/n)_t \longrightarrow (2h\rho^2/\mu^3)_t,$$

where $d_{\|\cdot\|}$ denotes the supremum norm.

Proof Note that the global estimator $\hat{s}$ does not depend on h and t, i.e., the formulation of $\hat{s}$ as a process is artificial. We show that $\hat{\mu} \to \mu$ a.s. and $\hat{\rho}_\ell \to \rho_\ell$ a.s. as $n \to \infty$ for $\ell = 0, 1, 2, \ldots$ where $\rho_0 = \sigma^2$. Since $\{\xi_i\}_{i \geq 1}$ is m-dependent and square-integrable, the sequence $\{\xi_i \xi_{i+\ell}\}_{i \geq 1}$ is integrable and $(m + \ell)$-dependent, thus ergodic. Then, the ergodic theorem, see e.g., Klenke (2008), states almost surely as $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i \longrightarrow \mathbb{E}[\xi_1] = \mu \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\xi_i\xi_{i+\ell} \longrightarrow \mathbb{E}[\xi_1\xi_{1+\ell}]. \tag{19}$$

Since the life times are a.s. positive and integrable, it follows $N_{nT} \to \infty$ a.s. as $n \to \infty$ (cmp. the proof to Lemma A.1. in Messer et al. (2014)). Thus, in Eq. (19), the value n can be exchanged with the random number of observations N_{nT} (respectively $N_{nT} - (\ell - 1)$). Hence, for $n \to \infty$, we find $\hat{\mu} \to \mu$ a.s. and $\hat{\rho}_\ell \to \rho_\ell$ a.s., so that the finite sum $\hat{\rho}^2 \to \rho^2$ a.s. By construction of $\hat{s}^2$ the statement holds. $\square$

Lemma A.2 *Let $\{\xi_i\}_{i \geq 1}$ be an m-dependent process in $\mathscr{P}$ and for all $T > 0$ and $h \in (0, T/2]$ let $((\hat{s}_{nh,nt})^2)_t$ be the local estimator as in Eq. (17). Then it holds in $(D[h, T - h], d_{\|\cdot\|})$ almost surely as $n \to \infty$ that $((\hat{s}_{nh,nt})^2/n)_t \to (2h\rho^2/\mu^3)_t$.*

Proof We show the uniform a.s. convergence of $(\hat{\mu}_{\text{le}})_t$ and $(\hat{\mu}_{\text{ri}})_t$ to the constant μ in Lemma A.4, and the uniform a.s. convergence of the summands $(\hat{\rho}_{\text{le},\ell})_t$ and $(\hat{\rho}_{\text{ri},\ell})_t$ of $\hat{\rho}_{\text{le}}^2$ and $\hat{\rho}_{\text{ri}}^2$ to the constant ρ_ℓ in Lemma A.5. This implies the statement, since uniform almost sure convergence interchanges with finite sums in general and with products if the limits are constant. $\square$

We start with a uniform a.s. result for the scaled counting process $(N_t)_{t \geq 0}$. Throughout, we use the following approach: First, we state an almost sure convergence result for the finite dimensional marginals of the processes. This essentially results from the ergodic theorem.

Then, by a discretization argument, we show uniform a.s. convergence.

Lemma A.3 *Let $\{\xi_i\}_{i \geq 1}$ be a process in $\mathscr{P}$ with $\mathbb{E}[\xi_1] = \mu$. Then we have in $(D[h, T - h], d_{\|\cdot\|})$ almost surely as $n \to \infty$ that*

$$\left(\frac{N_{nt} - N_{n(t-h)}}{nh/\mu}\right)_t \longrightarrow (1)_t, \tag{20}$$

$$\left(\frac{N_{n(t+h)} - N_{nt}}{nh/\mu}\right)_t \longrightarrow (1)_t. \tag{21}$$

Proof We show Eqs. (21); (20) follows analogously.

For $S_n := \sum_{i=1}^{n}\xi_i$ for $n \geq 1$, the ergodic theorem implies $S_n/n \to \mu$ a.s. for $n \to \infty$. As we have $N_t \to \infty$ a.s. as $t \to \infty$, $S_{N_t}/N_t \to \mu$ a.s. as $t \to \infty$. Now, for all $t \geq 0$ we find $S_{N_t} \leq t \leq S_{N_t+1}$, so that (for all t sufficiently large such that $N_t \geq 1$)

$$\frac{S_{N_t}}{N_t} \leq \frac{t}{N_t} \leq \frac{S_{N_t+1}}{N_t + 1}\frac{N_t + 1}{N_t}.$$

Since the left hand side and the right hand side tend to μ almost surely we obtain $N_t/t \to 1/\mu$ a.s. as $t \to \infty$. For $0 \leq s < t$, this implies, as $n \to \infty$, almost surely

$$\frac{N_{nt} - N_{ns}}{n(t-s)} = \frac{t}{t-s}\frac{N_{nt}}{nt} - \frac{s}{t-s}\frac{N_{ns}}{ns}$$

$$\longrightarrow \frac{t}{t-s}\frac{1}{\mu} - \frac{s}{t-s}\frac{1}{\mu} = \frac{1}{\mu}. \tag{22}$$

This implies the convergence of the finite dimensional marginal of Eq. (21). The uniform convergence follows by a discretization argument analogously to the proof of Lemma A.14 in Messer et al. (2014). $\square$

Next, we show the uniform a.s. convergence of the estimators $(\hat{\mu}_{\text{ri}})_t$, $(\hat{\mu}_{\text{le}})_t$, $(\hat{\sigma}_{\text{ri}}^2)_t$ and $(\hat{\sigma}_{\text{le}}^2)_t$.

Lemma A.4 *Let $\{\xi_i\}_{i \geq 1} \in \mathscr{P}$ with $\mu := \mathbb{E}[\xi_1]$. Then it holds in $(D[h, T - h], d_{\|\cdot\|})$ almost surely as $n \to \infty$ that*

$$(\hat{\mu}_{le})_t \longrightarrow (\mu)_t \qquad \text{and} \qquad (\hat{\mu}_{ri})_t \longrightarrow (\mu)_t.$$

Proof Again we prove the statement only for the right window. We find $(1/n)\sum_{i=1}^{n}\xi_i \to \mu$ a.s., such that $(1/N_t)\sum_{i=1}^{N_t}\xi_i \to \mu$ a.s. as $n \to \infty$. Then we conclude for all $0 < s < t$ (the case $s = 0$ being similar) as $n \to \infty$ almost surely

$$\frac{1}{N_{nt} - N_{ns}}\sum_{i=N_{ns}+1}^{N_{nt}}\xi_i$$

$$= \frac{N_{nt}}{N_{nt} - N_{ns}}\left(\frac{1}{N_{nt}}\sum_{i=1}^{N_{nt}}\xi_i - \frac{N_{ns}}{N_{nt}}\frac{1}{N_{ns}}\sum_{i=1}^{N_{ns}}\xi_i\right)$$

$$\longrightarrow \frac{t}{t-s}\left(\mu - \frac{s}{t}\mu\right) = \mu, \tag{23}$$

making use of Lemma A.3. Thus, for every fixed t we obtain almost surely as $n \to \infty$

$$\hat{\mu}_{\mathrm{ri}} = \frac{1}{N_{n(t+h)} - N_{nt} - 1} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i \longrightarrow \mu. \qquad (24)$$

The a.s. convergence holds for finitely many t simultaneously. As above, the uniform convergence follows by a discretization argument analogously to the proof of Lemma A.15 in Messer et al. (2014). □

Now we show the uniform a.s. convergence of covariance estimators.

Lemma A.5 *Let $\{\xi_i\}_{i \geq 1} \in \mathscr{P}$, and let $\hat{\rho}_{le,\ell}$ and $\hat{\rho}_{ri,\ell}$ be the local estimators of ρ_ℓ in the left and right window, see Eqs. (14),(17), for $\ell = 0, 1, 2 \ldots$, where $\rho_0 = \sigma^2$. Then in $(D[h, T-h], d_{\|\cdot\|})$ a.s. as $n \to \infty$ we have*

$$\left(\hat{\rho}_{le,\ell} \right)_t \longrightarrow (\rho_\ell)_t \qquad and \qquad (\hat{\rho}_{ri,\ell})_t \longrightarrow (\rho_\ell)_t.$$

Proof Again we conclude $(1/n) \sum_{i=1}^{n} \xi_i \xi_{i+\ell} \to \mathbb{E}[\xi_1 \xi_{1+\ell}]$ a.s. as $n \to \infty$. Using $N_{nT} \to \infty$, we find $(1/N_{nt}) \sum_{i=1}^{N_{nt}} \xi_i \xi_{i+\ell} \to \mathbb{E}[\xi_1 \xi_{1+\ell}]$ a.s. as $n \to \infty$. With a similar argument as in Eq. (23), we find for all $0 \leq s < t$ almost surely as $n \to \infty$

$$\frac{1}{N_{nt} - N_{ns} - (\ell + 1)} \sum_{i=N_{ns}+2}^{N_{nt}-\ell} \xi_i \xi_{i+\ell} \to \mathbb{E}[\xi_1 \xi_{1+\ell}]. \qquad (25)$$

Together with the previous Lemma A.4 this implies the almost sure convergence $\hat{\rho}_{ri,\ell} \to \mathbb{E}[\xi_1 \xi_{1+\ell}] - \mathbb{E}[\xi_1]^2 = \rho_\ell$ for every fixed t and thus for the finite dimensional marginals.

In order to obtain the convergence in $(D[h, T-h], d_{\|\cdot\|})$, we show a.s. as $n \to \infty$ that

$$\left(\frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i \xi_{i+\ell} \right)_t \longrightarrow (\mathbb{E}[\xi_1 \xi_{1+\ell}])_t. \qquad (26)$$

The convergence of the finite dimensional marginals follows from Eq. (25) together with Lemma A.3 and Slutsky's theorem. We show the uniform convergence (26) even for $t \in [0, T-h]$. It suffices to show almost surely that

$$\lim_{n \to \infty} \sup_{t \in [0, T-h]} \frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i \xi_{i+\ell} \leq \mathbb{E}[\xi_1 \xi_{1+\ell}], \qquad (27)$$

$$\lim_{n \to \infty} \inf_{t \in [0, T-h]} \frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i \xi_{i+\ell} \geq \mathbb{E}[\xi_1 \xi_{1+\ell}].$$

Again, we make use of a discretization argument as in Messer et al. (2014). We make it explicit here, since the mixing terms $\xi_i \xi_{i+\ell}$ were not explicitly considered in the latter article. For an $\varepsilon > 0$ with $T/\varepsilon \in \mathbb{N}$, we decompose the time

interval $[0, nT]$ into equidistant sections of length $n\varepsilon$. Using the notation $|\lceil x \rceil| := \lceil x \rceil + 1, x \in \mathbb{R}$, we bound

$$\sup_{t \in [0, T-h]} \frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i \xi_{i+\ell}$$

$$\leq \max_{j=0,1,\ldots,T/\varepsilon-|\lceil h/\varepsilon \rceil|} \frac{\mu}{nh} \sum_{i=N_{jn\varepsilon}}^{N_{jn\varepsilon+n|\lceil h/\varepsilon \rceil|\varepsilon}} \xi_i \xi_{i+\ell}$$

$$\leq \max_{j=0,1,\ldots,T/\varepsilon-|\lceil h/\varepsilon \rceil|} \frac{\mu}{nh} \sum_{i=N_{jn\varepsilon+nh}}^{N_{jn\varepsilon+n|\lceil h/\varepsilon \rceil|\varepsilon}} \xi_i \xi_{i+\ell}$$

$$+ \max_{j=0,1,\ldots,T/\varepsilon-|\lceil h/\varepsilon \rceil|} \frac{\mu}{nh} \sum_{i=N_{jn\varepsilon}}^{N_{jn\varepsilon+nh}} \xi_i \xi_{i+\ell}.$$

For any $\delta > 0$ we can choose $\varepsilon > 0$ so that $\max_{j=0,\ldots,T/\varepsilon-|\lceil h/\varepsilon \rceil|} (N_{jn\varepsilon+n|\lceil h/\varepsilon \rceil|\varepsilon} - N_{jn\varepsilon+nh})/(\delta n/\mu) \to 1$ a.s. for $n \to \infty$. Then, for $n \to \infty$, the first summand in the latter display converges to $(\delta/h)\mathbb{E}[\xi_1 \xi_{1+\ell}]$ a.s. and the second summand to $\mathbb{E}[\xi_1 \xi_{1+\ell}]$ a.s., since convergence (26) holds for finitely many t. Since δ can be chosen arbitrarily small, we find the first inequality of Eq. (27). The second one follows analogously. Thus, the convergence in Eq. (26) follows. We then exchange the normalization according to Lemma A.3. Omitting $\ell+1$ summands does not change the limit such that the uniform a.s. convergence of $(\hat{\rho}_{ri,\ell})_t$ is shown. Analogously, the uniform a.s. convergence of $(\hat{\rho}_{le,\ell})_t$ is shown. □

References

Avila-Akerberg, O., & Chacron, M.J. (2011). Nonrenewal spike train statistics: causes and functional consequences on neural coding. *Exp. Brain Res.*, *210*(0), 353–71.

Berkes, I., Horváth, L., Kokoszka, P., & Shao, Q.-M. (2005). Almost sure convergence of the Bartlett estimator. *Period. Math. Hungar.*, *51*(1), 11–25.

Berkes, I., Horváth, L., Kokoszka, P., & Shao, Q.-M. (2006). On discriminating between long-range dependence and changes in mean. *Ann. Statist.*, *34*(3), 1140–1165.

Billingsley, P. (1999). Convergence of probability measures. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication.

Bingmer, M., Schiemann, J., Roeper, J., & Schneider, G. (2011). Measuring burstiness and regularity in oscillatory spike trains. *J. Neurosci. Methods*, *201*, 426–37.

Brody, C.D. (1999). Correlations without synchrony. *Neural Comput.*, *11*(7), 1537–1551.

Brown, E.N., Kass, R.E., & Mitra, P.P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.*, *7*(5), 456–461.

Camproux, A.C., Saunier, F., Chovet, G., Thalabard, J.C., & Thomas, G. (1996). A hidden markov model approach to neuron firing patterns. *Biophys. J.*, *71*(5), 2404–12.

Chacron, M.J., Lindner, B., & Longtin, A. (2004). Noise shaping by interval correlations increases information transfer. *Phys. Rev. Lett.*, *92*(8), 080601.

Chacron, M.J., Longtin, A., & Maler, L. (2001). Negative interspike interval correlations increase the neuronal capacity for encoding time-dependent stimuli. *J. Neurosci.*, *21*(14), 5328–5343.

Csörgő, M., & Horváth, L. (1987). Asymptotic distributions of pontograms. *Math. Proc. Cambridge Philos. Soc.*, *101*(1), 131–139.

De Jong, R., & Davidson, J. (2000). Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica*, *68*(2), 407–424.

Dehling, H., Rooch, A., & Taqqu, M.S. (2013). Non-parametric change-point tests for long-range dependent data. *Scand. J. Stat.*, *40*(1), 153–173.

Eden, U.T., Frank, L.M., Barbieri, R., Solo, V., & Brown, E.N. (2004). Dynamic analysis of neural encoding by point process adaptive filtering. *Neural Comput.*, *16*(5), 971–98.

Farkhooi, F., Strube-Bloss, M., & Nawrot, M. (2009). Serial correlation in neural spike trains: Experimental evidence, stochastic modeling, and single neuron variability. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, *79*(2 Pt 1), 021905.

Frick, K., Munk, A., & Sieling, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society*, *76*(3), 495–580.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, *42*(6), 2243–2281.

Gonçalves, S., & Politis, D. (2011). Discussion: Bootstrap methods for dependent data: A review. *J. Kor. Stat. Soc.*, *40*, 383–6.

Grün, S., Diesmann, M., & Aertsen, A. (2002). 'Unitary events' in multiple single-neuron activity. II. Non-stationary data. *Neural Comput.*, *14*(1), 81–119.

Gut, A., & Steinebach, J. (2002). Truncated sequential change-point detection based on renewal counting processes. *Scand. J. Statist.*, *29*(4), 693–719.

Gut, A., & Steinebach, J. (2009). Truncated sequential change-point detection based on renewal counting processes. II. *J. Statist. Plann. Inference*, *139*(6), 1921–1936.

Hartmann, C., Lazar, A., Nessler, B., & Triesch, J. (2015). Where's the noise? key features of spontaneous activity and neural variability arise through learning in a deterministic network. *PLoS Computational Biology*, *11*(12).

Kendall, D.G., & Kendall, W.S. (1980). Alignments in two-dimensional random sets of points. *Adv. in Appl Probab.*, *12*(2), 380–424.

Kirch, C., & Muhsal, B. (2014). *A MOSUM procedure for the estimation of multiple random change points*: Preprint.

Klenke, A. (2008). Probability theory. Universitext. Springer-Verlag London Ltd., London. A comprehensive course, Translated from the 2006 German original.

Koyama, S., Eden, U.T., Brown, E.N., & Kass, R.E. (2010). Bayesian decoding of neural spike trains. *Annals of the Institute of Statistical Mathematics*, *62*(1), 37–59.

Kreiss, J.P., & Lahiri, S.N. (2012). Bootstrap methods for time series. In *Time Series Analysis: Methods and Applications, 30:Ch. 1. Elsevier*.

Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Process. Appl.*, *83*(1), 79–102.

Lee, S.H., & Dan, Y. (2012). Neuromodulation of brain states. *Neuron*, *76*(1), 209–222.

Lowen, S.B., & Teich, M.C. (1991). Auditory-nerve action potentials form a nonrenewal point process over short as well as long time scales. *J. Acoust. Soc. Am.*, *92*, 803–6.

Luczak, A., Bartho, P., & Harris, K.D. (2013). Gating of sensory input by spontaneous cortical activity. *Journal of Neuroscience*, *33*(4), 1684–1695.

Messer, M., Kirchner, M., Schiemann, J., Roeper, J., Neininger, R., & Schneider, G. (2014). A multiple filter test for the detection of rate changes in renewal processes with varying variance. *Ann. Appl. Stat.*, *8*(4), 2027–2067.

Nawrot, M.P., Boucsein, C., Rodriguez-Molina, V., Aertsen, A., Grün, S., & Rotter, S. (2007). Serial interval statistics of spontaneous activity in cortical neurons *in vivo* and *in vitro*. *Neurocomputing*, *70*(10), 1717–1722.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Comp. Neur. Sys.*, *15*, 243–62.

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilinsky, E.J., & Simoncelli, E.P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, *454*(7202), 995–9.

Ratnam, R., & Nelson, M.E. (2000). Nonrenewal statistics of electrosensory afferent spike trains: implications for the detection of weak sensory signals. *J. Neurosci.*, *20*(17), 6672–6683.

Ray, B.K., & Tsay, R.S. (2002). Bayesian methods for change-point detection in long-range dependent processes. *J. Time Ser. Anal.*, *23*(6), 687–705.

Schiemann, J., Klose, V., Schlaudraff, F., Bingmer, M., Seino, S., Magill, P.J., Schneider, G., Liss, B., & Roeper, J. (2012). K-atp channels control in vivo burst firing of dopamine neurons in the medial substantia nigra and novelty-induced behavior. *Nat. Neurosci.*, *15*(9), 1272–1280.

Schneider, G. (2008). Messages of oscillatory correlograms - a spike-train model. *Neural Comput.*, *20*(5), 1211–1238.

Schwalger, T., & Lindner, B. (2013). Patterns of interval correlations in neural oscillators with adaptation. *Front. Comp. Neurosci*, *7*, 164.

Shiau, L., Schwalger, T., & Lindner, B. (2015). Interspike interval correlation in a stochastic exponential integrate-and-fire model with subthreshold and spike- triggered adaptation. *J. Comp. Neurosci.*, *38*, 589.

Singh, K. (1981). On asymptotic accuracy of Efron's bootstrap. *Ann. Stat.*, *9*, 1187–95.

Steinebach, J., & Eastwood, V.R. (1995). On extreme value asymptotics for increments of renewal processes. *J. Statist. Plann. Inference*, *45*(1-2), 301–312.

Steinebach, J., & Zhang, H.Q. (1993). On a weighted embedding for pontograms. *Stochastic Process. Appl.*, *47*(2), 183–195.

Subramaniam, M., Althof, D., Gispert, S., Schwenk, J., Auburger, G., Kulik, A., Fakler, B., & Roeper, J. (2014). Mutant α-synuclein enhances firing frequencies in dopamine substantia nigra neurons by oxidative impairment of a-type potassium channels. *The Journal of Neuroscience*, *34*(41), 13586–99.

Tang, S.M., & MacNeill, I.B. (1993). The effect of serial correlation on tests for parameter change at unknown time. *The Annals of Statistics*, *21*(1), 552–75.

Trucculo, W., Eden, U.T., Fellows, M.R., Donoghue, J.P., & Brown, E.N. (2004). A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *J. Neurophysiol.*, *93*, 1074–89.

Vervaat, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrsch. Verw. Geb.*, *23*(4), 245–253.

Wied, D., Krämer, W., & Dehling, H. (2012). Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory*, *28*(3), 570–589.

Wu, W.B., & Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica*, *19*(4), 1755–1768.

Xiao, H., & Wu, W.B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.*, *40*(1), 466–493.