A MULTIPLE FILTER TEST FOR THE DETECTION OF RATE CHANGES IN RENEWAL PROCESSES WITH VARYING VARIANCE

By Michael Messer[*,1], Marietta Kirchner[†], Julia Schiemann[*,‡],
Jochen Roeper[*,1,2], Ralph Neininger[*] and Gaby Schneider[1,2*]

*Goethe University Frankfurt[*], Heidelberg University[†]
and University of Edinburgh[‡]*

Nonstationarity of the event rate is a persistent problem in modeling time series of events, such as neuronal spike trains. Motivated by a variety of patterns in neurophysiological spike train recordings, we define a general class of renewal processes. This class is used to test the null hypothesis of stationary rate versus a wide alternative of renewal processes with finitely many rate changes (change points). Our test extends ideas from the filtered derivative approach by using multiple moving windows simultaneously. To adjust the rejection threshold of the test, we use a Gaussian process, which emerges as the limit of the filtered derivative process. We also develop a multiple filter algorithm, which can be used when the null hypothesis is rejected in order to estimate the number and location of change points. We analyze the benefits of multiple filtering and its increased detection probability as compared to a single window approach. Application to spike trains recorded from dopamine midbrain neurons in anesthetized mice illustrates the relevance of the proposed techniques as preprocessing steps for methods that assume rate stationarity. In over 70% of all analyzed spike trains classified as rate nonstationary, different change points were detected by different window sizes.

1. Introduction. In neurophysiology, spike trains are often analyzed with statistical models based on point processes, for example, renewal processes Johnson (1996), Kass, Ventura and Brown (2005), Nawrot et al. (2008), Perkel, Gerstein and Moore (1967a), Rieke et al. (1999). A large field of statistical neuroscience focuses on the coordination between parallel point processes [Brown, Kass and Mitra (2004), Grün and Rotter (2010), Perkel, Gerstein and Moore (1967b)]. In many models used for such analyses, rate stationarity is a crucial assumption, and variations of the underlying firing rate can affect the results of the applied techniques [e.g., Brody (1999), Grün, Riehle and Diesmann (2003)]. In order to avoid such problems, several authors have suggested local techniques, which involve the separate treatment of sections with approximately stationary rate [see, e.g., Grün, Diesmann and Aertsen (2002), Schneider (2008), Staude, Rotter and Grün

(2008)] when spike trains show nonstationary properties. Therefore, it is important to capture these nonstationary properties, that is, to detect the violation of rate stationarity and to locate the changes in the firing rate of neurons.

In this paper we contribute to the change point analysis of point processes. Motivated by the modeling of empirical data from neurophysiology, we define a general class of renewal processes. In this class, we test the null hypothesis of rate stationarity versus a wide alternative of renewal processes with finitely many rate changes. Our test extends ideas from the filtered derivative approach [Bertrand (2000), Steinebach and Eastwood (1995)] by using multiple moving windows simultaneously instead of just one moving window. To adjust the rejection threshold of the test, we use a Gaussian process, which emerges as the limit of the filtered derivative process. Additionally, we develop a multiple filter algorithm, which can be used when the null hypothesis is rejected in order to estimate the number and location of change points. We analyze the benefits of our multiple filter algorithm and study the increase in detection probability against single window techniques. This procedure can serve as a preprocessing step, splitting up the time series into sections, in which the analyses of interest can be performed separately. As an example, Figure 1 illustrates a point process with nonstationary rate, in which we aim to estimate the number and location of change points.

For identifying the number and positions of change points in time series, many techniques are available in mathematical statistics. For an overview see, for example, Basseville and Nikiforov (1993), Brodsky and Darkhovsky (1993), Csörgő and Horváth (1997). Typically, these techniques are derived in the context of time series models with independent and identically distributed (i.i.d.) random variables. The classical parametric test uses a maximized likelihood quotient in order to analyze the entire process, which leads to so-called pontograms in point/renewal process theory [see Csörgő and Horváth (1987, 1997), Kendall and Kendall (1980), Steinebach and Zhang (1993)]. The resulting test statistics have extreme-value type limits [Hušková and Slabý (2001)]. As a second approach, moving window analyses in the context of renewal processes have been studied by Steinebach and Eastwood (1995). These local concepts successively investigate the life times of the point process instead of referring to the entire process.

Motivated by applications, we present two extensions of existing methods: first, the high variability of point processes observed empirically requires a sufficiently general class of point process models. Accordingly, we first introduce in Section 2 a new class of renewal processes with varying variance (RPVV), which allow a certain variability in the variance of the life time distributions. This generalization

FIG. 1. *A time series of events for which visual inspection suggests a nonstationary rate. For a general class of point processes, we present a statistical test and an algorithm based on multiple windows in order to identify the number and location of change points in the rate.*

has the additional advantage that rate changes can be investigated irrespective of variance changes and that the latter could then be analyzed in a subsequent, separate analysis which respects the identified rate changes. As a second extension to existing methods, we take into account that rate changes can occur on fast and slow time scales within the same time series. We propose a multiple filter technique that applies multiple windows simultaneously. This technique consists of a statistical multiple filter test (MFT) for the null hypothesis of rate stationarity and a multiple filter algorithm (MFA) for change point detection.

In Section 3 we first extend techniques introduced by Steinebach and Eastwood (1995) to our class of RPVVs. In particular, we prove asymptotic results for a moving average approach called filtered derivative, which is based on comparing the number of events in adjacent windows. We then introduce a statistical test that is based on a set of filtered derivative processes, each process corresponding to one window size. The maximum over all processes serves as a test statistic, indicating deviations from rate stationarity if this maximum exceeds a threshold Q. By scaling each process, we attempt to give every window a similar impact on the maximum distribution.

For practical application, we provide in Section 4 a multiple filter algorithm for change point detection, in which the results obtained by multiple window sizes are combined. For each individual window, the algorithm successively searches for extreme values of the filtered derivative, similar to the techniques proposed by Bertrand (2000), Bertrand, Fhima and Guillin (2011).

In Section 5 we evaluate the MFT, discuss the significance level in finite data sets and compare it to bootstrap methods. Most importantly, we show by exemplary simulations that the MFA can have an increased detection probability over single window techniques even when a best window size is known. Thus, by using multiple window sizes, one can detect rate changes in fast and slow time scales simultaneously, increase the detection probability and avoid the problem of choosing one near-optimal bandwidth [cf., e.g., Basseville and Nikiforov (1993), Csörgő and Horváth (1997), Jones, Marron and Sheather (1996), Nawrot, Aertsen and Rotter (1999), Shimazaki and Shinomoto (2007)].

Finally, we apply the MFT to a sample data set of spike train recordings obtained as spontaneous single-unit activity from identified dopamine neurons in the substantia nigra of anesthetized mice (Section 6). In the sample data set, the detected change points agree closely with visual inspection. In over 70% of all spike trains, which are classified to have a nonstationary rate, different change points are detected by different window sizes.

2. The point process model. In this section we extend the assumptions of classical renewal processes by introducing a class of renewal processes with varying variance (RPVV) (Section 2.1). These processes are assumed rate stationary, but the variance of life times may show a certain degree of variability. Examples of such processes are given in Section 2.2. For the alternative hypothesis (Section 2.3)

we combine several null elements, resulting in processes with a piecewise station-ary rate. In this model we aim to detect rate changes irrespective of other point process properties, such as the variability of the life times or even changes in the variability of life times.

2.1. *Renewal processes with varying variance (RPVV).*   We write a point pro-cess Φ as an increasing sequence of events

$$0 < S_1 < S_2 < S_3 < \cdots,$$

where S_i denotes the occurrence time of the ith event, for $i = 1, 2, \ldots$. Alterna-tively, Φ is determined by its life times $(\xi_i)_{i \geq 1}$, where

$$\xi_1 = S_1 \quad \text{and} \quad \xi_i = S_i - S_{i-1} \quad \text{for } i = 2, 3, \ldots,$$

or by the counting process $(N_t)_{t \geq 0}$, where

$$(1) \qquad\qquad N_t = \max\{i \geq 1 | S_i \leq t\}, \qquad t \geq 0,$$

with the convention $\max \varnothing := 0$.

Under the null hypothesis, we assume that a spike train can be described as an element Φ of the following family of rate stationary processes, which we term renewal processes with varying variance (RPVV).

DEFINITION 2.1 [Renewal process with varying variance (RPVV)].   Let $T > 0$, and let Φ be a renewal process restricted on $(0, T]$ whose life times, $\xi_1, \xi_2, \ldots$, are assumed to be independent, positive and square-integrable random variables with positive variances, such that for some $\mu, \sigma, c > 0$ and all $\varepsilon > 0$, with asymptotics as $n \to \infty$, we have

$$(2) \qquad \text{rate stationarity:} \qquad\qquad \mathbb{E}[\xi_i] = \mu \qquad \text{for all } i \in \mathbb{N},$$

$$(3) \qquad \text{variance regularity:} \qquad\qquad \frac{1}{n} \sum_{i=1}^{n} \text{Var}(\xi_i) \to \sigma^2,$$

$$(4) \qquad \text{Lindeberg condition:}$$
$$\frac{\sum_{i=1}^{n} \mathbb{E}[(\xi_i - \mu)^2 \mathbb{1}_{\{(\xi_i - \mu)^2 > \varepsilon^2 \sum_{i=1}^{n} \text{Var}(\xi_i)\}}]}{\sum_{i=1}^{n} \text{Var}(\xi_i)} \to 0,$$

$$(5) \qquad \text{uniform variance bound:} \qquad\qquad \sup_{i \in \mathbb{N}} \text{Var}(\xi_i) < c,$$

$$(6) \qquad \text{SLLN for squared life times:} \qquad \frac{1}{n} \sum_{i=1}^{n} (\xi_i^2 - \mathbb{E}[\xi_i^2]) \to 0 \qquad \text{a.s.}$$

Thus, an RPVV can be a renewal process with i.i.d. life times, and thus constant variance of life times. This applies, for example, to Poisson processes or to processes with independent and $\Gamma(p, \lambda)$-distributed life times, called here Gamma-processes. In addition, the variance of life times can also show a certain variability as specified in (3) and (5). Assumptions (2)–(6) are technically sufficient for the asymptotic results that support our methods: condition (3) imposes a regularity of the life times' variances over time. The Lindeberg condition (4) is later used for process convergence to Brownian motion that allows to deduce asymptotics for the related counting process. There, condition (5) will be used additionally. Assumption (6) is the strong law of large numbers (SLLN) for the squares of the life times, which will be needed for strong consistency of an estimation of σ^2 below. Note that by Kolmogorov's conditions [Petrov (1995), Theorem 6.8] $(\xi_i^2)_{i \geq 1}$ satisfying the SLLN is equivalent to

$$(7) \qquad \sum_{i=1}^{\infty} P(|\xi_i^2 - \mathbb{E}[\xi_i^2]| \geq i) < \infty,$$

$$(8) \qquad \sum_{i=1}^{\infty} \frac{1}{i^2} \mathbb{E}[(\xi_i^2 - \mathbb{E}[\xi_i^2])^2 \mathbb{1}_{\{|\xi_i^2 - \mathbb{E}[\xi_i^2]| < i\}}] < \infty,$$

$$(9) \qquad \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(\xi_i^2 - \mathbb{E}[\xi_i^2]) \mathbb{1}_{\{|\xi_i^2 - \mathbb{E}[\xi_i^2]| < n\}}] \to 0 \qquad \text{as } n \to \infty.$$

The most important assumption (2) states that in an RPVV, the mean rate $1/\mu$ is constant across time. We therefore also use the short notation $\Phi(\mu)$.

2.2. *Examples of RPVVs.* Here, we give examples of point processes that satisfy the assumptions of an RPVV from Definition 2.1. We assume rate stationarity [condition (2)]. Figure 2 shows examples of such processes. Panels A–D indicate the evolution of variances of life times, and panels E–F illustrate point processes with the corresponding variances and Gamma-distributed life times. Because Gamma-processes have been used frequently in order to describe neuronal spiking activity [cf., and the references therein, Nawrot et al. (2008)], we also use Gamma processes for all simulations in the present article, choosing suitable combinations of rate and regularity parameters for each simulation.

The most simple example of an RPVV is a process with i.i.d. life times (Figure 2A and E). As a second example (Figure 2B and F), an RPVV can be a process in which the variances of life times converge to a constant. Third, the variance of life times can alter regularly between two different values (Figure 2C). The corresponding point process (panel G) shows regular and irregular sections. This example can be extended such that the mean variance of life times is constant at equidistant grid points $g, 2g, \ldots$ (Figure 2D and H).
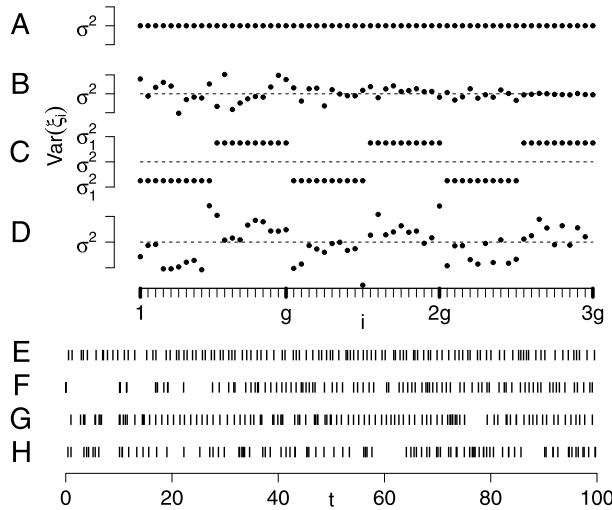
FIG. 2. *Examples of RPVVs according to Definition* 2.1. *(A–D) The variances* $\mathrm{Var}(\xi_i)$ *of life times* ξ_i *is indicated by points.* $\mathrm{Var}(\xi_i)$ *can be constant* (σ^2) *(A), can converge to a constant* σ^2 *(B), or can be a step function alternating between different, fixed values in a regular manner (C). In (D), the mean variance of the g life times* $(\xi_1, \ldots, \xi_g)$, $(\xi_{g+1}, \ldots, \xi_{2g})$, *etc., is a constant* σ^2. *(E–H) Realizations* $(T = 100)$ *of point processes with* $\Gamma(p_i, \lambda_i)$-*distributed life times* ξ_i *with constant expectation* $\mathbb{E}[\xi_i] = p_i/\lambda_i = 1$, *that is,* $p_i = \lambda_i$. *The variances* $\mathrm{Var}(\xi_i) = p_i/\lambda_i^2 = 1/\lambda_i$ *are given in (A–D), respectively. (E) Independent and* $\Gamma(5, 5)$-*distributed life times with constant variance* $\mathrm{Var}(\xi_i) = \sigma^2 = 1/5$. *(F)* $\mathrm{Var}(\xi_i) = 1/\lambda_i \to \sigma^2 = 0.1$. *(G) The variance alternates in a regular manner, changing after* $g/2 = 20$ *life times between* $p_i = \lambda_i = 1$ *(Poisson process) and* $p_i = \lambda_i = 20$ *(a regular Gamma process). (H) For* $g = 40$, *the mean variance of the g life times* $(\xi_1, \ldots, \xi_g)$, $(\xi_{g+1}, \ldots, \xi_{2g})$, *etc., equals unity.*

2.3. *The full model.* In contrast to the null assumption, the alternative hypothesis assumes that Φ is piecewise an RPVV, where the mean rate can change between the different sections. Formally, we assume that under the alternative hypothesis, a spike train is an element of the class constructed in Construction 2.2.

CONSTRUCTION 2.2. *Let* $T > 0$, *and let* $\mathbf{C}$ *denote the set of all finite subsets of* $(0, T]$. *Assume* $C := \{c_1, \ldots, c_k\} \in \mathbf{C}$, *with* $c_1 < \cdots < c_k$.

At time 0 start $k + 1$ *independent RPVVs* $\Phi_1(\mu_1), \ldots, \Phi_{k+1}(\mu_{k+1})$ *with*

$$\mu_i \neq \mu_{i+1} \qquad for\ i = 1, \ldots, k.$$

Let $c_0 := 0$, $c_{k+1} := T$ *and define*

$$(10) \qquad \Phi := \bigcup_{i=1}^{k+1} \Phi_i|_{(c_{i-1}, c_i]},$$

where $\Phi_i|_{(c_{i-1}, c_i]}$ *denotes the restriction of* Φ_i *to the interval* $(c_{i-1}, c_i]$.
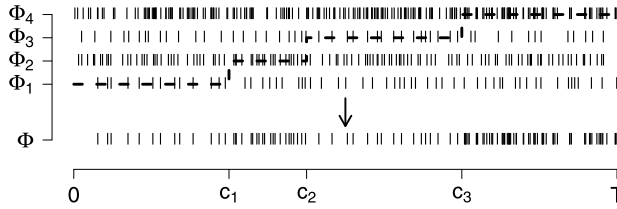
FIG. 3. *The change point model combines a set of RPVVs. A realization of a process Φ on $(0, T]$ that results from Construction 2.2. Φ has three change points c_1, c_2, c_3 and originates from the four RPVVs $\Phi_1, \ldots, \Phi_4$, jumping from process Φ_i to Φ_{i+1} at change point c_i.*

The times $c_1, \ldots, c_k$ are called change points. An example of a point process generated according to this construction is shown in Figure 3. The resulting rate of Φ is a step function with change points $c_1, \ldots, c_k$.

We now define a model set $\mathcal{M} := \mathcal{M}(T)$ to be the family of processes that derive from Construction 2.2 and test the null hypothesis:

H_0: $\Phi \in \mathcal{M}$ with $C = \varnothing$, that is, Φ is an RPVV, in particular rate stationary, against the alternative.

H_A: $\Phi \in \mathcal{M}$ and $C \neq \varnothing$, that is, there is at least one change point.

3. The multiple filter test (MFT). In order to test the above null hypothesis of rate stationarity in the model set $\mathcal{M}$, we derive here a multiple filter test (MFT). Section 3.1 summarizes the construction of the test. Details on parameter estimation and limit results are given in Sections 3.2 and 3.3.

3.1. *Derivation of the MFT.* The main idea of the MFT is to extend a filtered derivative technique [see the contributions Basseville and Nikiforov (1993), Brodsky and Darkhovsky (1993), Csörgő and Horváth (1997)], which slides two adjacent windows of size h and compares the number of events in the left and right window. Formally, let $T > 0$ and Φ be an element of the model set $\mathcal{M}$. For $h \in (0, T/2)$ we define an analysis region $\tau_h := (h, T - h]$. Let $N_{(a,b]}(\Phi)$ denote the number of elements of Φ in the interval $(a, b] \subset (0, T]$. For each point $t \in \tau_h$ we compare the number of events

$$N_{\mathrm{le}} := N_{(t-h,t]}(\Phi) \quad \text{and} \quad N_{\mathrm{ri}} := N_{(t,t+h]}(\Phi)$$

in the left and right window (Figure 4A).

A large difference $N_{\mathrm{ri}} - N_{\mathrm{le}}$ can indicate deviations from the null hypothesis of rate stationarity. But because the variance of the difference depends on process parameters, the difference $N_{\mathrm{ri}} - N_{\mathrm{le}}$ will be normed as follows:

$$(11) \qquad G_{h,t} := G_{h,t}(\Phi) := \frac{N_{\mathrm{ri}} - N_{\mathrm{le}}}{\hat{s}} \qquad \text{if } \hat{s} > 0,$$

and $G_{h,t} := 0$ if $\hat{s} = 0$ for all $t \in \tau_h$ (Figure 4B). The term $\hat{s}$ denotes an estimator of $\sqrt{\mathrm{Var}[N_{\mathrm{ri}} - N_{\mathrm{le}}]}$, which is defined in (20). We will show in Section 3.3 that the
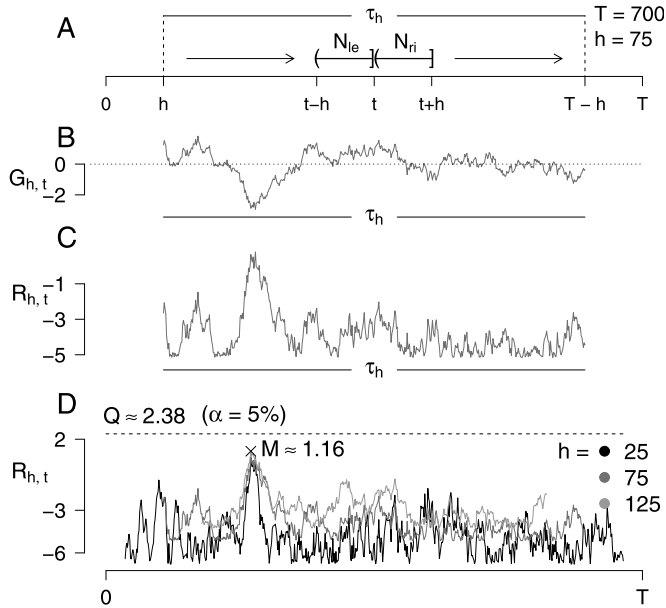
FIG. 4. *Illustration of the computational steps and processes involved in the MFT. The MFT is applied here to a stationary process on $(0, 700]$ with independent and $\Gamma(0.25, 5)$-distributed life times. (A) For one window size $h = 75$, the number of events in the left and right window, $N_{\text{le}}, N_{\text{ri}}$, are derived for every $t \in \tau_h$. (B) The process $(G_{h,t})_{t \in \tau_h}$ for one window $h = 75$. (C) The scaled process $(R_{h,t})_{t \in \tau_h}$ for one window $h = 75$. (D) All scaled processes $(R_{h,t})_{t \in \tau_h}$ for $h \in H = \{25, 75, 125\}$. Different gray shades indicate different window sizes, the asymptotic threshold Q is represented by a dashed line. Here, the test statistic $M = \max_{h,t} R_{h,t} < Q$ and, thus, the null hypothesis of rate stationarity is not rejected.*

process $(G_{h,t})_{t \in \tau_h}$ converges to a $2h$-dependent Gaussian process $(L_{h,t})_{t \in \tau_h}$. The limit process $(L_{h,t})_{t \in \tau_h}$ is a continuous functional of a standard Brownian motion and depends only on T and h. In particular, it is independent from the parameters of Φ such as, for example, the rate or regularity.

Large absolute values of $G_{h,t}$ indicate potential deviations from rate stationarity. Therefore, the maximum

$$M_h := \max_{t \in \tau_h} |G_{h,t}|$$

can serve as a test statistic for a single window.

In order to combine multiple window sizes of a finite set $H \subset (0, T/2]$, we consider a set of stochastic processes $\{(G_{h,t})_{t \in \tau_h} | h \in H\}$, which are all derived from the same underlying point process Φ. Each process $(G_{h,t})_{t \in \tau_h}$ results in one maximum M_h. Instead of using the raw maxima M_h, we suggest to standardize M_h because the distribution of M_h depends on h. As mentioned above, the process $(G_{h,t})_{t \in \tau_h}$ is $2h$-dependent, and a smaller h results in weaker temporal de-

pendencies of the process. This leads to higher chance fluctuations in $(G_{h,t})_t$ for smaller h, and thus a higher rejection threshold.

If the expectation and variance of M_h were known, we could use the term

$$(12) \qquad \frac{M_h - \mathbb{E}[M_h]}{\sqrt{\mathrm{Var}(M_h)}}$$

in order to give every window a similar impact on the global maximum of all processes. Here, we approximate the expectation and variance using simulations of the set of limit processes $\{(L_{h,t})_{t \in \tau_h} | h \in H\}$. Defining $M_h^* := \sup_{t \in \tau_h} |L_{h,t}|$, we approximate the expectation $\mathbb{E}[M_h]$ by the empirical mean $\overline{M}_h^*$ and the variance $\mathrm{Var}(M_h)$ by the empirical variance $v(M_h^*)$. The resulting test statistic M across all windows is defined as the global maximum

$$(13) \qquad M := \max_{h \in H} \left(\frac{M_h - \overline{M}_h^*}{\sqrt{v(M_h^*)}} \right).$$

Finally, we reject the null hypothesis at level α if $M > Q := Q(\alpha, T, H)$. The threshold Q is defined such that under the null hypothesis, $M > Q$ with probability α. In order to derive Q, one can again use the limit processes $\{(L_{h,t})_{t \in \tau_h} | h \in H\}$ and approximate Q by the empirical quantile of

$$(14) \qquad M^* := \sup_{h \in H} \left(\frac{M_h^* - \overline{M}_h^*}{\sqrt{v(M_h^*)}} \right).$$

Note that all limit processes $(L_{h,t})_t$ are derived from the same Brownian motion in order to ensure comparability with the processes $(G_{h,t})_t$, which result from the same point process Φ.

For change point detection explained later in Section 4 and for graphical illustration, we use the scaled process

$$(15) \qquad R_{h,t} := \left(\frac{|G_{h,t}| - \overline{M}_h^*}{\sqrt{v(M_h^*)}} \right) \qquad \text{(Figure 4C)},$$

which scales $(G_{h,t})_{t \in \tau_h}$ and accounts for the scaling of the maxima. Because the maximum of all processes $(R_{h,t})$,

$$(16) \qquad M = \max_{h \in H} \max_{t \in \tau_h} R_{h,t}$$

is identical to the above global test statistic, it can be read directly from the graph. The processes $(R_{h,t})_{t \in \tau_h}$ and their comparison with the threshold Q are illustrated in Figure 4D.

3.2. *Variance estimation.* By definition of our auxiliary variables $G_{h,t}$ [see (11)], we need to specify an estimator $\hat{s}^2$ for the variance of $N_{ri} - N_{le}$. The idea is to estimate the variance from the life times of the elements in the left and right windows of $G_{h,t}$.

Let $\xi_1, \xi_2, \ldots$ be the life times of an RPVV with constant μ and σ^2 as in (2) and (3). Given T and h, for every $t \in \tau_h$ we define

$$(17) \qquad \gamma_{le}(t, h) := \{\xi_i : S_i, S_{i-1} \in (t - h, t], i = 1, 2, \ldots\},$$

the set of all life times that correspond to events in the left window. We relabel this set of life times $\xi_1^{le}, \xi_2^{le}, \ldots$. Analogously for the right window, we obtain $\gamma_{ri}(t, h) = \{\xi_1^{ri}, \xi_2^{ri}, \ldots\}$.

The empirical mean of the life times in the left window is denoted by

$$(18) \qquad \hat{\mu}_{le} := \hat{\mu}_{le}(t, h) := \overline{\gamma_{le}(t, h)} \qquad \text{if } |\gamma_{le}| > 0,$$

and $\hat{\mu}_{le} := 0$ if $|\gamma_{le}| = 0$. The empirical variance of the life times is

$$(19) \qquad \hat{\sigma}_{le}^2 := \hat{\sigma}_{le}^2(t, h) := v(\gamma_{le}(t, h)) \qquad \text{if } |\gamma_{le}| > 1,$$

and $\hat{\sigma}_{le}^2 := 0$ if $|\gamma_{le}| \leq 1$. The bar denotes the empirical mean, $v(\cdot)$ denotes the corrected sample variance of $\gamma_{le}(t, h)$, and $|\cdot|$ denotes the number of elements. Analogously, we define $\hat{\mu}_{ri}$ and $\hat{\sigma}_{ri}^2$ for the right window.

As an estimator for the variance of $N_{ri} - N_{le}$ we propose

$$(20) \qquad \hat{s}^2 := \hat{s}^2(t, h) := \left(\frac{\hat{\sigma}_{ri}^2}{\hat{\mu}_{ri}^3} + \frac{\hat{\sigma}_{le}^2}{\hat{\mu}_{le}^3} \right) h \qquad \text{if } \hat{\mu}_{le} \wedge \hat{\mu}_{ri} > 0,$$

and $\hat{s}^2 := 0$ otherwise, where $\wedge$ denotes the minimum. Note that $\hat{s}^2$ is zero by definition if the number of events is less than two in any window. We prove strong consistency of these estimators in an appropriate asymptotic setting in Appendix A.3. Heuristically, this estimator is suggested by the fact that under our conditions on the life times of the RPVV we obtain for the number N_t of events up to time t that, as $t \to \infty$, we have

$$(21) \qquad \frac{N_t - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \xrightarrow{d} N(0, 1) \quad \text{and} \quad \text{Var}[N_t] \sim \sigma^2 t / \mu^3,$$

where $\xrightarrow{d}$ denotes convergence in distribution. Hence, we obtain

$$\text{Var}[N_{ri} - N_{le}] \approx \left(\frac{\sigma^2}{\mu^3} + \frac{\sigma^2}{\mu^3} \right) h.$$

3.3. *Limit distribution of $(G_{h,t})$ under H_0.* In order to compute the test statistic M and choose the rejection threshold Q, we derive a limit of the process $(G_{h,t})_{t\in\tau_h}$, choosing an asymptotic setting in which time T and window size h grow proportionally. As the limit we identify a $2h$-dependent Gaussian process $(L_{h,t})_{t\in\tau_h}$ on τ_h that does not depend on the parameters of the process Φ.

To make this asymptotic statement precise, let Φ be an element of H_0 with life times $\xi_1, \xi_2, \ldots$. We consider an extended version $(G_{h,t}^{(n)})_{t\in\tau_h}$ of $(G_{h,t})_{t\in\tau_h}$,

$$(22) \qquad G_{h,t}^{(n)} := \frac{(N_{n(t+h)} - N_{nt}) - (N_{nt} - N_{n(t-h)})}{\hat{s}(nt, nh)} \qquad \text{if } \hat{s}(nt, nh) > 0,$$

and $G_{h,t}^{(n)} := 0$ otherwise, for all $t \geq h$ and $n = 1, 2, \ldots$. Recall that N_t denotes the number of life times up to time t and the estimator $\hat{s}$ is defined in (20). We consider the processes $(G_{h,t})_{t\in\tau_h}$ and $(G_{h,t}^{(n)})_{t\in\tau_h}$ as càdlàg processes in the Skorokhod topology.

The asymptotic analysis is given by letting $n \to \infty$. To define the limit process, let $W = (W_t)_{t\geq 0}$ denote a standard Brownian motion on $[0, \infty)$. For $h > 0$ we define for all $t \geq h$

$$(23) \qquad L_{h,t} := \frac{(W_{t+h} - W_t) - (W_t - W_{t-h})}{\sqrt{2h}}.$$

The process $(L_{h,t})_{t\geq h}$ is a $2h$-dependent Gaussian process, with zero mean and autocovariance given as

$$(24) \qquad \Sigma_v^h := \Sigma_{u,u+v}^h := \begin{cases} 1 - \dfrac{3}{2h}|v|, & \text{if } |v| \in [0, h], \\ -1 + \dfrac{1}{2h}|v|, & \text{if } |v| \in (h, 2h], \\ 0, & \text{if } |v| \geq 2h, \end{cases}$$

for all suitable u, v (Figure 5). Note that the autocovariance only depends on the window size h and the time lag v of two elements $L_{h,t}$ and $L_{h,t+v}$.

In the Appendix we show the following process convergence, extending results obtained by Steinebach and Eastwood (1995).
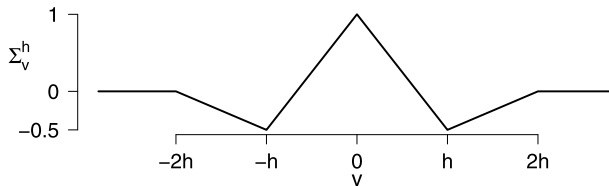

FIG. 5. *The autocovariance structure Σ_v^h of $(L_{h,t})_{t\geq h}$ as a function of the time lag v for a fixed window size h.*

THEOREM 3.1. *Let $T > 0$ and $h \in (0, T/2]$ be a window size. Let Φ be an element of the null hypothesis. Then for the processes $(G_{h,t}^{(n)})$ and $(L_{h,t})$ defined in* (22), *and* (23), *as $n \to \infty$, we have*

$$(25) \qquad (G_{h,t}^{(n)})_{t \in \tau_h} \xrightarrow{d} (L_{h,t})_{t \in \tau_h},$$

where $\xrightarrow{d}$ denotes weak convergence in the Skorokhod topology.

4. Multiple filter algorithm (MFA) for change point detection. In Section 3 the first part of the MFT was presented as a test for the null hypothesis of rate stationarity versus the alternative of at least one rate change. After rejection of the null hypothesis, we intend to identify the number and location of change points. To this end, we propose an algorithm that combines the results of multiple window sizes. It consists of a procedure for change point detection on the basis of individual windows (single filter algorithm—SFA, Section 4.1) and a multiple filter algorithm (MFA) for the combination of individual windows (Section 4.2).

4.1. *Single filter algorithm (SFA).* For the detection of change points with a single window of size h, we apply a common method to the scaled filtered derivative process $(R_{h,t})_{t \in \tau_h}$, which successively estimates change points from the maxima of the process [see the contributions Antoch and Hušková (1994), Basseville and Nikiforov (1993), Bertrand (2000), Bertrand, Fhima and Guillin (2011)]. Similar procedures have been shown to give consistent estimates of the number and location of the change points under mild conditions in Gaussian sequence change point models [Hušková and Slabý (2001), Muhsal (2013)].

The SFA for one $h \in H$ works as follows. First, observe the maximum of the process $(R_{h,t})_{t \in \tau_h}$. If $\max_t R_{h,t} > Q$, this indicates deviations from rate stationarity. The time $\hat{c}_1$ at which this maximum is taken is an estimate of a change point because the maxima are expected at the change points if the difference between change points is sufficiently large (Figure 6A). More precisely, one should note that the sample path of $(R_{h,t})_{t \in \tau_h}$ is a step function, so that the set of maximizers is an interval. We define $\hat{c}_1$ as the infimum of this interval

$$\hat{c}_1 := \inf \Big\{ \arg \max_{t \in \tau_h} R_{h,t} \Big\}.$$

Second, we observe that a change point which occurs at time c affects the behavior of the process $(R_{h,t})_{t \in \tau_h}$ within the h-neighborhood of c,

$$(26) \qquad B_h(c) := (c - h, c + h) \cap \tau_h \qquad \text{(Figure 6A),}$$

while leaving all points outside of $B_h(c)$ unaffected. Therefore, the h-neighborhood of $\hat{c}_1$ is omitted in the subsequent analysis. If the remaining process $(R_{h,t})_{t \in \tau_h \setminus B_h(c)}$ outside of $B_h(c)$ exceeds Q, this indicates another deviation from rate stationarity because a change point at c cannot cause this deviation. Therefore, we successively identify change points as the maxima of $(R_{h,t})_t$ outside the union of all $B_h(\hat{c}_i)$ of detected change points, until the process $(R_{h,t})_t$ is smaller than Q in all remaining intervals (Figure 6B).
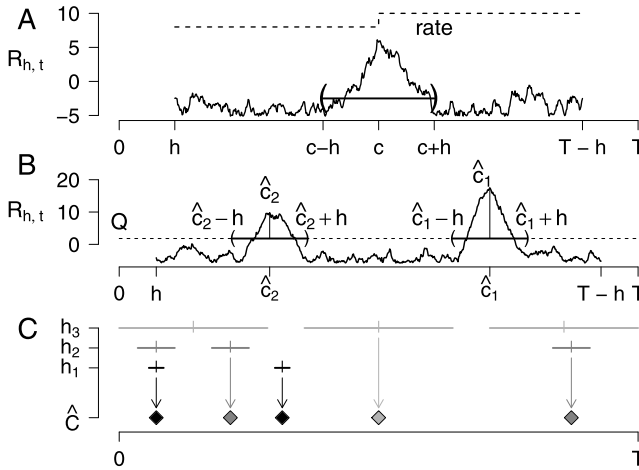
FIG. 6. *The SFA and MFA. (A) A change point at time c affects the process $(R_{h,t})_t$ within the h-neighborhood of c, and the maximum of $(R_{h,t})_t$ is expected at c. (B) The SFA successively searches for maxima of $(R_{h,t})_t$. When the maximum is larger than Q, the maximizer $\hat{c}_1$ is the first change point estimator. Then the h-neighborhood of $\hat{c}_1$ is omitted, and the procedure is iterated on the remaining process until the maximum remains smaller than Q (underlying process different from A). (C) Schematic representation of the MFA for a window set $H = \{h_1, h_2, h_3\}$ (underlying process different from A and B). The change points detected by SFA are marked as vertical bars, and their h_i-neighborhoods are indicated by horizontal lines (h_1, black, h_2, gray, h_3, light gray). The MFA first accepts all change points detected with the smallest window h_1 (black diamonds). Among the change points detected by h_2 (gray), only the first one is rejected because its h_2-neighborhood contains an accepted change point. Among the change points detected by the largest window (light gray), only the second one is added to the list of accepted change points because its h_3-neighborhood does not contain formerly accepted change points. Diamonds indicate finally accepted change points.*

4.2. *Multiple filter algorithm* (*MFA*). We now propose a multiple filter algorithm with which the results of the SFA of different windows can be combined. This integrates the advantages of multiple time scales because large windows are more likely to detect small rate changes and small windows can be more sensitive to fast changes. In particular, using only a large window of size h, the SFA can fail or mislocate change points c_1, c_2 with distance smaller than h. This suggests to prefer change point estimates of smaller windows.

The MFA can be summarized as follows (Figure 6C). Let $H = \{h_1, h_2, \ldots, h_n\}$ be the set of involved windows, with $h_1 < \cdots < h_n$. Derive the threshold Q for this set H as described in Section 3. For all h_i, detect change points via SFA. Let $\widehat{C}_i := \{\hat{c}_1^i, \ldots, \hat{c}_{k_i}^i\}$ denote the set of change points estimated with window h_i. Then, define a set of accepted change points $\widehat{C}$, which is first set to $\widehat{C} := \widehat{C}_1$, that is, all change points estimated by the smallest window. Among the change points $\widehat{C}_2$ associated with h_2, only those are added to $\widehat{C}$ whose h_2-neighborhood does not include a formerly accepted change point $c_j^1 \in \widehat{C}$. The remaining estimates $c_j^2 \in \widehat{C}_2$ are assumed to be affected by change points that have already been estimated and
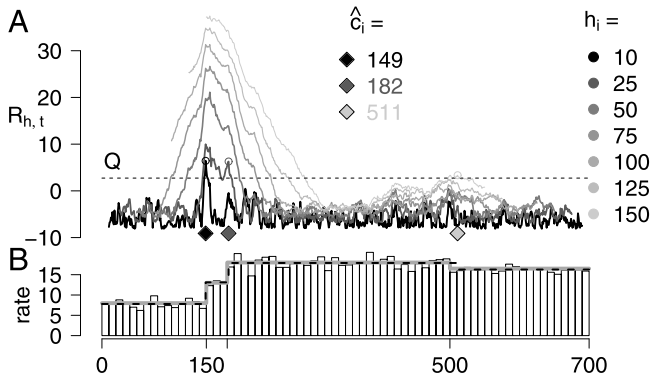
FIG. 7. *Application of the MFT to a simulated point process on* $(0, 700]$ *with three change points at* $c_1 = 150$, $c_2 = 180$ *and* $c_3 = 500$. *The life time distributions were* $\exp(8)$ *in* $[0, c_1]$, $\Gamma(2, 26)$ *in* $[c_1, c_2]$, $\exp(18)$ *in* $[c_2, c_3]$ *and* $\Gamma(2, 33)$ *in* $[c_3, T]$, *corresponding to rates of* $8, 13, 18$ *and* 16.5 *in the respective intervals.* (A) *Gray curves indicate the processes* $(R_{h,t})$ *for window sizes* $h \in H = \{10, 25, 50, 75, 100, 125, 150\}$. *The simulated threshold* $Q = 2.75$ *is indicated by the dashed line. The estimated change points* $\hat{c}_1 = 149$, $\hat{c}_2 = 182$, $\hat{c}_3 = 511$ *are marked by diamonds. As indicated by the grayscale of the diamonds, each change point was detected by a different window size.* (B) *Rate histogram of the underlying point process with real* (*gray*) *and estimated* (*dashed*) *rate profiles.*

therefore omitted in the further analysis. This procedure is iterated by successively increasing the window sizes up to h_n.

4.3. *Application to a simulated point process.* Figure 7 illustrates the application of the MFA to a simulated point process with three change points. All change points are detected by the MFA, and the estimated change points correspond closely to the true change points. Consequently, the rate estimates agree closely with the true rates.

Figure 7 also shows that different window sizes were used for the detection of different change points: while the first change point was detected by the smallest window $h_1 = 10$, the second was detected by $h_2 = 25$ and the third by $h_3 = 150$. This supports the idea of combining several windows: if change points are close together (e.g., c_1 and c_2 in Figure 7), small windows are preferable because large windows tend to be affected by both change points, and thus lead to imprecise estimates. On the other hand, small rate changes require large windows, which have a higher test power. Indeed, none of the individual windows could detect all change points (data not shown).

4.4. *Choosing the window set* H. The previous example and the simulations that follow in Section 5.2 show that multiple filters can increase the probability to detect change points. This is because rate changes in fast and slow time scales can be detected simultaneously using multiple windows. However, using too many
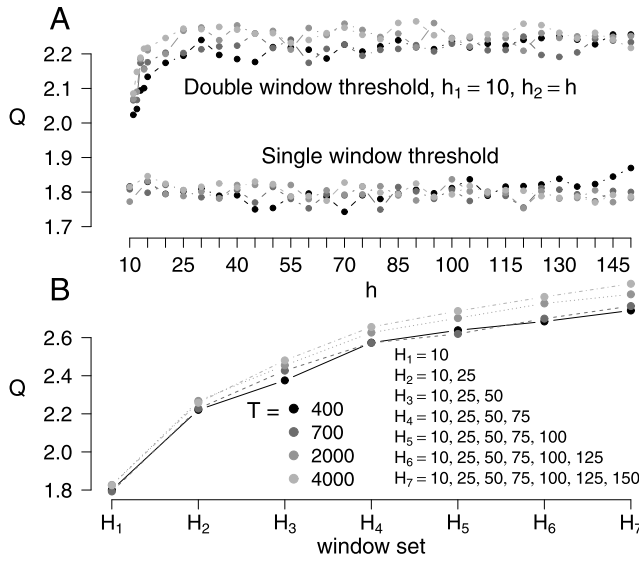
FIG. 8. *Dependence of Q on H and T for $\alpha = 5\%$. (A) For SFA, the threshold Q basically does not depend on T or h (connected points at about $Q \approx 1.8$, color codes for T as in B). Choosing $h_1 = 10$, the second window h_2 increases Q to about 2.23, with a stronger increase for larger h_2 (points at about $Q = 2.2$, color codes for T as in B). (B) Adding more windows from the set $H = \{10, 25, 50, 75, 100, 125, 150\}$ leads to only slight increases in Q. Dependence on the recording time T is weak. 10,000 simulations were used to calculate the empirical mean and standard deviation for the standardization of the limit process $(L_{h,t})_t$ [equation (14)].*

windows increases the threshold Q applied for change point detection, which can also decrease the test power in certain settings. Therefore, we discuss here in which way Q depends on the window set H and give recommendations for the choice of H.

Because Q depends only on T, H and α, we investigate its dependency on T and H for $\alpha = 5\%$. Figure 8A shows that if only one window is used, the single window threshold Q does essentially not depend on h or T. Because the test statistic is normed for every h [equation (12)], any window size h and any simulated time T results in a threshold of about $Q \approx 1.8$. In order to study the influence of one additional window on Q, we fix $h_1 = 10$ and illustrate the double window threshold Q for $h_2 \in \{11, 12, \ldots, 15, 20, \ldots, 150\}$ in Figure 8A. The threshold Q increases to about 2.23. Smaller h_2 close to $h_1 = 10$ lead to smaller increases than larger windows because the processes $(R_{h_1,t})_t$ and $(R_{h_2,t})_t$ show higher correlation if $|h_1 - h_2|$ is small. In Figure 8B we successively add windows of increasing size to the set $H_7 = \{10, 25, 50, 75, 100, 125, 150\}$. The increase in Q from $H_1 = \{h_1\}$ to $H_2 = \{h_1, h_2\}$ is about the same as from H_2 to H_7. Similarly, adding more windows between 10 and 150 would only slightly increase Q (data not shown).

Because additional windows have minor impact on Q, we recommend the following window choice: the smallest window h_1 should be restricted such that the

asymptotic significance level is approximately kept. To this end, Section 5.1.1 investigates the empirical significance level for stationary Gamma processes with different regularity and rate parameters. The maximal window $h_{\max}$ is only limited by $T/2$. The choice of the grid between h_1 and $h_{\max}$ can be guided by the following principles: choosing a narrow grid can detect change points in a broad class of time scales. However, it will also slightly increase the threshold Q, and thus reduce the probability to detect change points at all. Additionally, it increases the computational effort required for the performance of the test. Here, we study the performance for the window set $H = \{10, 25, 50, 75, 100, 125, 150\}$.

5. Evaluation of the MFT.

5.1. *Practical applicability of the MFT.*

5.1.1. *Empirical significance level in simulations.* As discussed in Section 3.3, the proposed MFT is an asymptotic procedure, providing asymptotic significance level α. Therefore, we use simulations in order to investigate under which conditions the asymptotic significance level is kept for small rates in the finite setting. We simulate rate stationary renewal processes with Gamma-distributed life times in order to investigate the empirical significance level of the asymptotic MFT. We focus on the parameters $T = 700$, $H = \{10, 25, 50, 75, 100, 125, 150\}$ and an asymptotic significance level $\alpha = 5\%$.

Figure 9 shows the empirical significance level obtained in 10,000 simulations as a function of the mean (μ) and standard deviation (σ) of the independent and Gamma-distributed life times. Under high irregularity, that is, if σ is high, the test remains conservative. With increasing regularity, the rate required to obtain an empirical significance level of 5% is increasing. For low rates and high regularity, the percentage of false positives of the MFT tends to be slightly larger than the asymptotic significance level. In the very extreme case of almost perfect regularity and

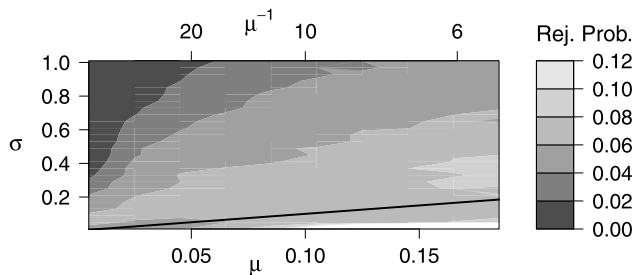FIG. 9. *Simulated rejection probability of the MFT for processes with i.i.d. Gamma-distributed life times ($T = 700$, $H = \{10, 25, 50, 75, 100, 125, 150\}$, 10,000 simulations). For high irregularity, the test tends to be conservative. With increasing regularity, the rate required to keep the asymptotic 5% significance level increases. μ and σ denote the mean and standard deviation of life times.*

low rates (white area in the bottom right corner), the MFT should not be applied because the empirical significance level is largely enhanced. In all but these extreme parameter combinations, the detection of more than one change point was very unlikely under the null hypothesis (detection of at least 2 change points in <1%, of at least 3 change points in <0.1% of 1000 simulations, data not shown). Thus, the detection of more than one change point can almost always be considered a strong indication of rate nonstationarity.

In summary, one needs to keep in mind for practical applications that the error rate can be slightly enhanced for regular processes with low rates. However, a false estimation of a nonexisting change point is not problematic if one primarily intends to split up the time series into rate stationary sections. If the significance level needs to be kept strictly even for small rates, the window size needs to be increased. This has the same effect as increasing the rate because the approximation of $(G_{h,t})_t$ to the limit process $(L_{h,t})_t$ [equation (3.1)] mainly depends on the mean number of events per window.

5.1.2. *Comparison of the MFT to a bootstrap test.* The preceding section shows that the MFT should be treated carefully in situations with limited rates and high regularity because the asymptotic significance level is not precisely kept. Therefore, one might consider deriving Q with a bootstrap procedure, as suggested, for example, by Hušková and Slabý (2001). The distribution of M can then be derived directly by permutation of the life times and recalculation of M in the permuted process. By construction, this procedure yields an empirical significance level of 5% if the underlying process is a classical, stationary renewal process. However, it has two shortcomings: first, it requires high computational effort because the process $(R_{h,t})_t$ [equation (15)] needs to be recalculated for every realization. Second, permutation can only be applied if the life times are independent and identically distributed.

Therefore, we compare the MFT with a bootstrap test when the underlying process does not comply with the assumption of independent and identically distributed life times, that is, when the underlying process is a rate stationary RPVV but not a classical renewal process. To this end, we simulate rate stationary processes with Gamma-distributed life times. The variance of life times changes every $g/2$ life times, alternating between two values. As shown in Section 2.2 (Figure 2C), the resulting process is an RPVV.

In order to reduce computational effort for the bootstrap test, we replace $R_{h,t}$ by only computing $|N_{\mathrm{ri}}(t, h) - N_{\mathrm{le}}(t, h)|$, the absolute difference of the number of events in the left and right windows, for every h and t, and derive the maximum of these values as a test statistic. The 95%-quantile of the distribution of this test statistic is then estimated in permutations, and the null hypothesis is rejected if the maximum is larger than its estimated quantile.

Table 1 shows the resulting significance levels for the MFT and the bootstrap procedure. The MFT roughly keeps the 5% significance level in all simulated scenarios, whereas the bootstrap test rejects the null hypothesis in about 3%, 7% and

Comparison of the significance level of the MFT and a bootstrap test for simulated RPVVs. Here, the distribution of life times changes every $g/2$ life times from $\Gamma(0.5, 15)$ to $\Gamma(5, 150)$, leading to alternations between regular and irregular patterns. The grid size is (A) $g = 5000$, (B) $g = 10{,}000$, (C) $g = 20{,}000$. 1000 *simulations with* $H = \{10, 25, 50, 75, 100, 125, 150\}$ *and* $T = 700$ *at level* $\alpha = 5\%$ *were performed in all cases*, 1000 *permutations were used for the construction of the bootstrap threshold*

$\Gamma(0.5, 15) \rightsquigarrow \Gamma(5, 150)$	MFT	Bootstrap
(A) $g = 5000$	$(5.9 \pm 0.7)\%$	$(3.0 \pm 0.5)\%$
(B) $g = 10{,}000$	$(4.7 \pm 0.7)\%$	$(6.6 \pm 0.8)\%$
(C) $g = 20{,}000$	$(5.5 \pm 0.7)\%$	$(15.1 \pm 1.1)\%$

15% of the simulations. This indicates, as expected, that permutation tests are not necessarily robust against changes in the variance of life times and should therefore not be applied under such conditions.

5.1.3. *True change points do not increase the frequency of falsely detected change points.* The previous paragraphs show that the proposed MFT keeps the asymptotic significance level also in empirical point processes with a finite time horizon, that is, rejecting the null hypothesis of stationary rate with probability about α. In contrast, the proposed MFA for change point detection is a heuristic procedure that is not associated with a specific significance level. However, as mentioned in Section 4.1, the SFA is a common method which yields consistent change point estimates under mild conditions in Gaussian models [Hušková and Slabý (2001)]. In addition, we explain here why the MFA, after taking into account the typical number of falsely detected change points (false positives, FP), should not overestimate the number of true change points. More precisely, a true change point does not increase the number of FPs. This is because a true change point can only affect its h-neighborhood, which is cut out in the SFA after detection. Outside this h-neighborhood, the remaining process should resemble a process derived under the null hypothesis, and thus produce about as many FPs as under the null hypothesis with the same threshold Q. For the MFA with multiple windows, a similar argument holds because change points are only added when no accepted change point lies within their h-neighborhood (cf. Section 4.2). Thus, one change point should usually lead to at most one detection.

In order to support these considerations, Table 2 shows simulation results of Gamma processes of length $T = 700$ with a change point at $c = 350$ in which we investigate the number of correctly and of falsely detected change points. A change point is called correctly detected if its h-neighborhood overlaps a true change point, whereas h corresponds to the window used for detection in the MFA. Rate

TABLE 2

Simulation results of $\Gamma(2, \lambda)$-processes of length $T = 700$ with a rate change at $c = 350$. Life times are $\Gamma(2, 24)$-distributed on $[0, 350]$ and $\Gamma(2, \lambda_i)$-distributed on $[350, 700]$ with $\lambda_i \in \{25, 26, 28, 30\}$. The respective rates are given on the left, 10,000 simulations per scenario

Rates	Detection prob. of true cp	Mean number of FPs per process	% of processes with ≥1 FP
$12 \rightsquigarrow 12.5$	0.119	0.051	4.9
$12 \rightsquigarrow 13$	0.653	0.048	4.6
$12 \rightsquigarrow 14$	0.996	0.050	4.9
$12 \rightsquigarrow 15$	0.999	0.048	4.6

changes of different heights are simulated in order to account for different detection probabilities of the inserted change point (first column). In this setting, the MFA does not falsely detect more change points than under the null hypothesis. The number of FPs (second column) and the number of processes with at least one falsely detected change point (third column) even decrease slightly because after cutting h-neighborhoods, the remaining process is shorter, and thus less likely to cross the threshold by chance.

5.2. *Multiple filters increase the detection probability.* We have already seen in the example in Section 4.3 that multiple windows can increase the probability to detect a change point. One explanation is that the simultaneous use of multiple filters avoids the problem of choosing the most appropriate single window size. But, more importantly, the combination of multiple filters is advantageous because large windows have a higher detection probability, whereas small windows can be more precise or sensitive to fast changes. Accordingly, we show here in simulations that the MFA can even detect more change points than the best single window.

In order to quantify this effect, we investigate the following random change point model. We simulate processes Φ on $(0, 700]$ in which the rate fluctuates between four different values. The model includes rate changes of different size and in different time scales. Each process Φ is a piecewise composition of four independent renewal processes $\Phi_1, \ldots, \Phi_4$ with Gamma-distributed life times with event rates $\mu_1^{-1} = 14$, $\mu_2^{-1} = 12$, $\mu_3^{-1} = 10$ and $\mu_4^{-1} = 9$. The change points for switches between the processes Φ_1 to Φ_4 are given by a stationary renewal process Φ_c on $(0, T]$ with change points $c_1, \ldots, c_{|\Phi_c|}$. In order to simulate change points in different time scales, the life times of Φ_c are uniformly distributed on $[0, 100]$. The observed process Φ is constructed from $\Phi_1, \ldots, \Phi_4$ as follows: set $\Phi||_{(0,c_1]} := \Phi_1|_{(0,c_1]}$, that is, start in process Φ_1. At the first change point c_1 choose independently and uniformly a process from $\{\Phi_2, \Phi_3, \Phi_4\}$ and jump into this process, such that, for example, $\Phi|_{(c_1,c_2]} = \Phi_2|_{(c_1,c_2]}$. Third, jump back deterministically to Φ_1 at c_2, that is, set $\Phi|_{(c_2,c_3]} = \Phi_1|_{(c_2,c_3]}$. Repeat the procedure, choosing
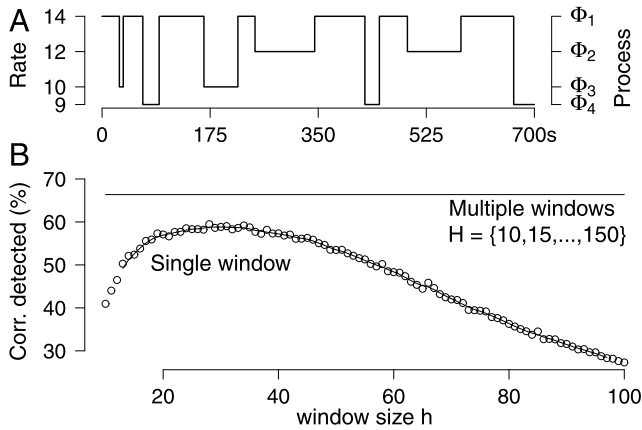
FIG. 10. *Multiple filtering increases detection rate.* (A) *A random realization of the rate of a process Φ used in the simulations. Intervals between change points are independent and unif(0, 100]-distributed. Simulated processes $\Phi_1, \ldots, \Phi_4$ have independent and $\Gamma(2, \lambda_i)$-distributed life times with rate parameters $\lambda_1 = 28$, $\lambda_2 = 24$, $\lambda_3 = 20$ and $\lambda_4 = 18$, leading to rates $\mu_1^{-1} = 14$, $\mu_2^{-1} = 12$, $\mu_3^{-1} = 10$ and $\mu_4^{-1} = 9$ (indicated on the left). (B) Mean relative frequency of correct change point detections in simulated processes as a function of the chosen window size. The points represent the mean percentages of correct detections derived in 1000 simulations using SFA. The curve shows a filtered average. The window size with maximal detection probability of about 0.59 is about $\tilde{h} \approx 28$. The horizontal line marks the mean relative detection probability of about 0.66 for a set of multiple windows $H = \{10, 15, \ldots, 150\}$.*

uniformly a process from $\{\Phi_2, \Phi_3, \Phi_4\}$ at odd-valued change points and returning to Φ_1 at even-valued change points. An example of the rate of the resulting process Φ is shown in Figure 10A.

Figure 10B indicates the percentage of correctly detected change points in 1000 simulations of the described processes. A change point is called correctly detected if its h-neighborhood overlaps a true change point, whereas h corresponds to the window used for detection in the MFA. In order to identify the best individual window, the detection rate for the SFA is shown as a function of the window size $h \in \{10, 11, \ldots, 100\}$. The percentage of correct detections is maximal at about 59% for a window size of about $\tilde{h} = 28$. Using the MFA with a set of multiple windows chosen here arbitrarily as $H = \{10, 15, \ldots, 150\}$, the correct detection rate increases to about 66%.

6. Application to spike train recordings.

6.1. *Data analysis.* In this section we apply the proposed MFT to a data set of 72 empirical spike train recordings that were reported partly in Schiemann et al. (2012). The recording time T was 540–900 seconds per spike train, and the mean firing rate was about 6 spikes per second. The significance level was set to $\alpha = 5\%$.

In order to choose the set of windows, we use the results from Section 5.1.1, Figure 9. Briefly, a mean number of about 100–200 events in the smallest window is required in order to keep the asymptotic significance level for point processes with medium irregularity. Therefore, we choose a minimal window of $h_1 = 25$ for a mean rate of 6 Hz and $H = \{25, 50, 75, 100, 125, 150\}$.

Figure 11 shows two spike train analyses in which multiple change points have been detected. As indicated by the different grayscales, different window sizes were used for change point estimation. From the set of 72 spike trains, 62 were identified as nonstationary. In 50 spike trains, at least two change points were detected, and in 37 spike trains, more than one window was necessary for the detection of these change points. Across all spike trains, the mean rate of detected change points was about 0.32 per minute. The lengths of intervals between detected change points followed a right-skewed distribution with median 75 s and
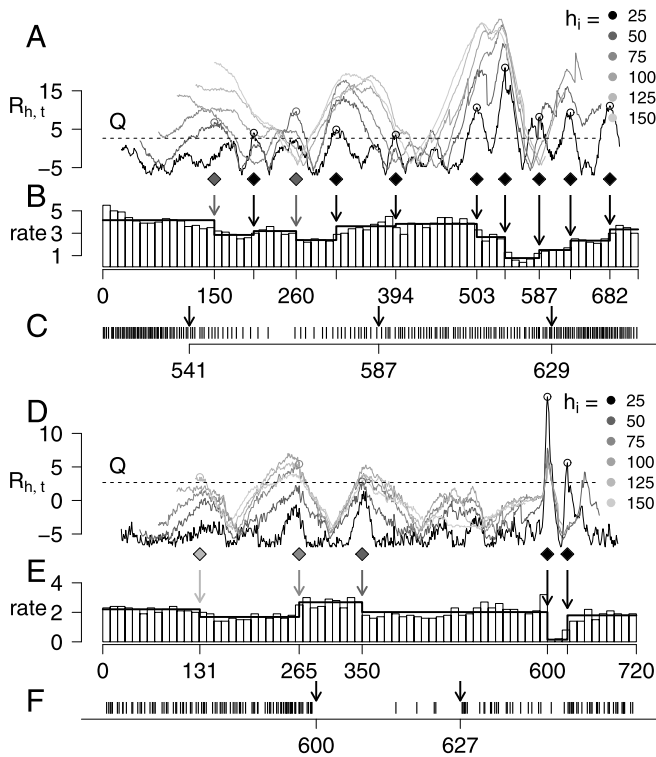

FIG. 11. *Application of the MFT to two spike train recordings;* $T = 720$, $H = \{25, 50, 75, 100, 125, 150\}$, $\alpha = 5\%$. (A) *and* (D) *The scaled processes* $(R_{h,t})_t$. *Grayscales of the different window sizes are indicated on the right. The dashed line marks the threshold* Q, *the detected change points are marked by diamonds. In spike train 1,* 10 *change points are detected with two different windows; in spike train 2,* 5 *change points are detected with four different windows.* (B) *and* (E) *Rate histograms of the spike trains. Black step function indicates estimated firing rate.* (C) *and* (F) *Short sections of the spike trains. Arrows mark the estimated change points.*
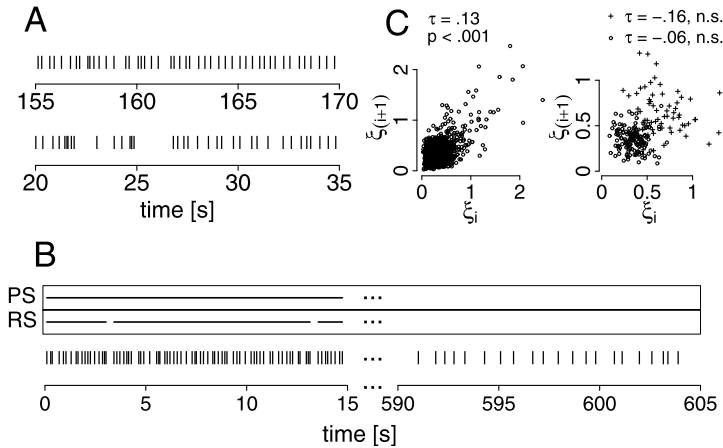
FIG. 12. *Examples of data analyses applying the MFT.* (A) *Sections of spike train* 1 (*above*) *and* 2 (*below*) *indicating changes in the variance of the life times. Left parts of illustrated sections show higher irregularity than right parts.* (B) *Application of Poisson Surprise* (*PS*) *and Rank Surprise* (*RS*) *algorithms to spike train* 1, *using standard parameters* (*only sections with surprise value* $S > 10$ *marked as bursts in both methods*). *Horizontal lines indicate bursts identified with PS and RS methods, respectively. Because sections with high rate* (*left*) *and low rate* (*right*) *are jointly analyzed, high-rate sections are identified as long bursts.* (C) *Analyzing serial correlations globally yields spurious positive rank correlation* (*left*), *whereas serial correlations within different sections* (*right, two sections indicated by point characters*) *can be slightly negative and nonsignificant.*

quartiles $q_1 = 44$ s and $q_3 = 123$ s. The height of a detected rate change, measured as the difference of estimated rates $\hat{\mu}_1^{-1}$ and $\hat{\mu}_2^{-1}$ at the change point in relation to their mean, $|\hat{\mu}_1^{-1} - \hat{\mu}_2^{-1}|/(0.5(\hat{\mu}_1^{-1} + \hat{\mu}_2^{-1}))$, ranged between about 0.5% and 173%. As one can see from the illustrations of the rate profiles in Figure 11B and E, the estimated rate profile corresponds well to a rate estimate that is obtained from visual inspection. Figure 12A illustrates that both spike trains show varying variance in their inter-spike intervals.

The identification of changes in the firing rate within neuronal spike trains can facilitate their interpretation and avoid pitfalls. Most importantly, the detected change points can be used for the separation of a spike train into sections of virtually stationary firing rate. This is important for multiple analysis techniques that assume rate stationarity for the description and statistical analysis of single or multiple spike trains, for example, techniques that study temporal coordination between processes [e.g., Grün, Diesmann and Aertsen (2002), Schneider (2008), Staude, Rotter and Grün (2008)]. Here, we show two simple analysis examples for individual spike trains.

First, variability of variance in the inter-spike intervals in dopamine (DA) neurons is often expressed as a switching of firing between a low-rate single spike background pattern and short events with relatively many spikes, so-called "bursts" (cf. also Figure 12A, bottom spike train: higher irregularity in the left part). For

DA neurons, burst firing has been shown to possess important behavioral significance, as it is coupled to an increase of DA release [Gonon (1988), Redgrave et al. (2010), Schiemann et al. (2012)]. Such bursts usually span very short periods with up to about 10 spikes and can thus not be detected with the asymptotic MFT, which requires about 100–200 spikes per window. However, the MFT can be an essential preprocessing step in burst detection when existing methods require rate stationarity. In two common methods for burst detection, bursts are described as short periods with "surprisingly many" spikes [Gourévitch and Eggermont (2007), Legéndy and Salcman (1985)]. These methods, called Poisson Surprise (PS) and Rank Surprise (RS), assume rate stationary Poisson or renewal processes and identify the "surprising" nature of a burst by comparison to the overall mean life time. If periods of different rates are jointly analyzed, the number of spikes in high-rate sections that are assigned to bursts can be much larger than when applying the algorithms to separate sections with approximately stationary firing rate. Figure 12B illustrates this effect by exemplary application of PS and RS burst detection algorithms to spike train 1, for which visual inspection indicates nonbursty firing activity (see also Figures 11C, 12A, top panel). The horizontal lines in Figure 12B indicated by PS and RS indicate the bursts identified by applying the two methods to the whole spike train. Almost all spikes in the high-rate section are assigned to long "bursts" consisting of 50 and more spikes (in illustrated sections: PS: one long burst with 160 spikes, RS: three long bursts with 18, 54 and 38 spikes). This is, however, inconsistent with the assumed physiological function and short duration of DA bursts. In agreement with these considerations, practically no bursts are identified in spike train 1 when applying the MFT first and separately analyzing the sections with different rates (PS: no bursts, RS: one burst with three spikes, not in illustrated section). Thus, by separating between multiple longer sections of different and unknown firing rates, the present algorithm for change point detection complements burst detection methods which aim at separating the two states "bursty" and "nonbursty" [e.g., Tokdar et al. (2010)].

Second, rate changes might cause potential misinterpretations of serial correlations of life times, which has also been discussed in the context of neuronal spike train analysis by Farkhooi, Strube-Bloss and Nawrot (2009). Consider a renewal process consisting of two periods with different rates. In each period, correlation between adjacent life times is zero, but in the high-rate section, short life times follow short life times, and in the low-rate section, long life times follow long life times. This induces a positive correlation in the global analysis. A similar result is obtained in spike train 1, in which we exemplarily analyze the correlation between adjacent life times with Kendall's rank correlation τ (Figure 12C). A global analysis falsely indicates a significant positive correlation (left panel, $\tau = 0.13$, $p < 0.001$) due to rate changes, whereas most correlations in individual sections are slightly negative and not significantly different from zero. The right part of Figure 12C shows two separate data pieces with different rates and slightly negative correlations and illustrates how the joint analysis of such data sets can produce a

spurious positive global correlation. Because serial correlations may reflect intrinsic neuronal properties [Benda and Herz (2003)], the application of the MFT as a preprocessing step can also be helpful in this context.

Finally, apart from improving statistical analysis by detecting periods of roughly constant rate, the detected rate changes themselves might contain important information. For example, in addition to bursts, periods of very low rate ("pauses," see, e.g., Figure 11F) may also have behavioral relevance. A recent study showed that the duration of these periods in DA neurons can be associated with the expression of fear [Mileykovskiy and Morales (2011)], and a modeling study demonstrated that synchronized pauses in spiking activity of many DA neurons can reduce information transmission in DA type 2 receptors [Dreyer et al. (2010)]. In addition to pauses, more complex change point sequences, such as multiple successive increases in the firing rate, could reflect specific prolonged changes in the typical DA activity that have been described recently [Howe et al. (2013)].

6.2. *Practical issues and R-code.* In practice, the described procedures can be applied easily. Depending on a rough estimate of the overall rate and irregularity of the process, one needs to choose the smallest window such that the asymptotic properties are kept. One can then choose a set of windows up to the largest interesting time scale. Then, the threshold Q can be estimated by repeated simulation of the limit process $(L_{h,t})_t$ [equation (14)].

In the supplementary material Messer et al. (2014) we provide an R code that performs these steps efficiently within one single routine and returns an illustration comparable to Figure 11. It also suggests a set of window sizes for a given time series of events. The code can be applied easily, using as input only a time series of events and (optional) a significance level and a set of windows, and returning a set of estimated change points.

7. Discussion. In this paper we have developed a multiple filter technique for the detection of change points in the event rate of time series. Motivated by the problem that rate stationarity of the underlying processes is crucial to many statistical analysis techniques, the multiple filter test (MFT) tests the null hypothesis of rate stationarity against the alternative of finitely many change points. In a second step, a multiple filter algorithm (MFA) identifies and locates an unspecified number of change points in the rate of the process. In addition, it includes a graphical representation in which strong deviations from rate stationarity can be visualized.

As a first extension to existent approaches, we introduce a general class of point processes called renewal processes with varying variance (RPVV). In addition to standard renewal assumptions reflected, for example, in Poisson or Gamma processes, an RPVV assumes that the variance of life times can show a certain degree of variability, which includes, for example, mixtures of Gamma processes in the simplest case. We propose RPVVs in order to account for the high variability of

patterns observed empirically and to allow for the detection of rate changes irre-spective of variance changes, which may be analyzed in subsequent, separate steps when rate changes have been identified.

In order to test the null hypothesis of rate stationarity against the alternative of finitely many change points, we extend a standard filtered derivative method which compares the number of events in adjacent windows in a moving window manner. Due to the general RPVV assumptions, statistical significance of deviations from rate stationarity cannot be tested by standard bootstrap approaches because the life times are not necessarily identically distributed. Therefore, we extend an asymp-totic result of Steinebach and Eastwood (1995) to RPVVs and show that the limit $(L_{h,t})_t$ of the filtered derivative process is a $2h$-dependent and zero mean Gaus-sian process. Notably, this limit $(L_{h,t})_t$ is independent of the underlying RPVV parameters such as the rate or the variances. By using the limit process, thresholds for testing the statistical significance of deviations from rate stationarity can be obtained by simulation.

As a second extension to existent approaches, we combine multiple window sizes $h \in H$ in order to detect rate changes at fast and slow time scales simultane-ously. In the present asymptotic setting, multiple window sizes can be combined easily because the set of processes $\{(G_{h,t})_t | h \in H\}$ depends on one underlying RPVV. In the same way, the set of limit processes $\{(L_{h,t})_t | h \in H\}$ depends on one underlying Brownian motion. In addition, the use of multiple windows requires two considerations: first, the statistical properties of $(L_{h,t})_t$ depend on the window size h. Therefore, we standardize the processes in order to give similar impact to every window size h. Second, change point detection requires an extended algo-rithm that combines the change points detected by multiple windows. Our multi-ple filter algorithm is based on the idea of preferring change points estimated by smaller windows to those estimated by larger windows. In a random change point model with multiple time scales that we used here for simulation, the MFA could detect more rate changes than the best individual window.

The presented methods can be particularly relevant for practical applications. First, the general assumptions of RPVVs cover a high variability of patterns ob-served in empirical time series. Second, multiple filtering can take into account that rate changes in empirical time series can occur at fast and slow time scales si-multaneously. In practice, one should keep in mind that the MFA always estimates a step function even when applied to a rate profile with gradual changes, and that very short time scales, for example, bursts with a few spikes, cannot be investi-gated by this asymptotic method. Third, in order to enable an easy application of the MFA, we provide an R code that includes all necessary steps within one single routine. It can be computed efficiently, and it also includes a graphical illustration of the resulting filtered derivative processes, in which large values indicate devia-tions from rate stationarity. In an exemplary application of the MFA to single unit neuronal recordings, we illustrate that the detection of rate changes can be impor-tant for the understanding of neuronal information processing and show that the

MFA can be a useful preprocessing step for data analysis techniques that assume rate stationarity.

In summary, we believe that the present multiple filter technique can be useful for the estimation of change points in the event rate of time series of events. It may be used as a universal preprocessing step whenever statistical analysis methods are sensitive to deviations from rate stationarity.

APPENDIX

In this Appendix we prove Theorem 3.1. Main ingredients of this proof are first the convergence of the normalized counting process $(N_t)_{t \geq 0}$ which is shown in Section A.2 (cf. Proposition A.6), and second the consistency of the estimator $(\hat{s}^2)_{t \in \tau_h}$ defined in (20). This is shown in Section A.3 (cf. Proposition A.13). First, in Section A.1 elementary facts are collected which are later needed repeatedly and for which we do not claim originality. The pieces are finally put together in Section A.4 to prove Theorem 3.1.

The following notation is used: for $\tau > 0$ the set of all real-valued continuous functions on $[0, \tau]$ is denoted by $C[0, \tau]$ and the set of all càdlàg functions by $D[0, \tau]$. We abbreviate the metric induced by the supremum norm by $d_{\|\cdot\|}$, the Skorokhod metric on $D[0, \tau]$ by d_{SK}. Analogously, we define $C[0, \infty)$ and use the metric $d_{\|\cdot\|}$ which induces the topology of compact convergence. Further, we use $D[0, \infty)$ and $D[h, T - h]$ with d_{SK}. Note that convergence in $(D[0, \infty), d_{\|\cdot\|})$ implies convergence in $(D[0, \infty), d_{SK})$.

A.1. Technical preliminaries. The lemmas in this subsection have different assumptions on the renewal processes occurring. However, note that the assumptions of all lemmas in this subsection are fulfilled for an RPVV as in Definition 2.1.

First, we want to assure that the number of events N_t in an RPVV tends to infinity almost surely (a.s.), while explosion is avoided.

LEMMA A.1. *Let $\{\xi_i\}_{i \geq 1}$ be a sequence of independent, positive, integrable random variables, interpreted as the life times of a point process on the positive line, and $(N_t)_{t \geq 0}$ the associated counting process as in* (1). *Then we have almost surely*

$$(27) \qquad N_t \to \infty \qquad (t \to \infty).$$

If the $\{\xi_i\}_{i \geq 1}$ are square integrable and satisfying conditions (2) *and* (5), *then for all $t \geq 0$ we have almost surely*

$$(28) \qquad N_t < \infty.$$

PROOF. For (27) note that N_t is increasing in t. For all fixed $k > 0$ we have

$$\{N_n < k\} = \left\{ \sum_{i=1}^{k} \xi_i > n \right\} \downarrow \bigcap_{n \geq 1} \left\{ \sum_{i=1}^{k} \xi_i > n \right\} \qquad (n \to \infty).$$

Since the ξ_i are integrable, we have $P(\xi_i < \infty) = 1$. Continuity from above (applied twice) yields

$$P\left(\bigcap_{n \in \mathbb{N}} \{N_n < k\}\right) = \lim_{n \to \infty} P\left(\sum_{i=1}^{k} \xi_i > n\right) \leq \lim_{n \to \infty} P\left(\bigcup_{i=1}^{k} \left\{\xi_i > \frac{n}{k}\right\}\right)$$

$$\leq \lim_{n \to \infty} \sum_{i=1}^{k} P\left(\xi_i > \frac{n}{k}\right) = 0.$$

This implies (27).

For (28) first note that (2) and (5) imply Kolmogorov's conditions (7)–(9) with ξ_i^2 replaced by ξ_i. Hence, we have the SLLN for $\{\xi_i\}_{i \geq 1}$, that is, $(1/n)\sum_{i=1}^{n} \xi_i \to \mu$ a.s. as $n \to \infty$. This implies $\sum_{i=1}^{n} \xi_i \to \infty$ a.s. and

$$P(N_t < \infty) = P\left(\bigcup_{n=1}^{\infty} \left\{\sum_{i=1}^{n} \xi_i > t\right\}\right) = P\left(\lim_{n \to \infty} \sum_{i=1}^{n} \xi_i > t\right) = 1. \qquad \square$$

Now we show that the number of events in successively increased windows, scaled with the widths of the windows, tends to the stationary rate $1/\mu$ almost surely.

LEMMA A.2. *Let $\{\xi_i\}_{i \geq 1}$ be a sequence of independent, positive, square-integrable random variables satisfying conditions (2) and (5), which are interpreted as the life times of a point process on the positive line, and $(N_t)_{t \geq 0}$ the associated counting process as in (1). Then for all $0 \leq s < t$ we have, as $n \to \infty$, almost surely*

$$\frac{N_{nt} - N_{ns}}{n(t - s)} \longrightarrow \frac{1}{\mu}.$$

PROOF. As in the proof of Lemma A.1, conditions (2) and (5) imply the SLLN for $\{\xi_i\}_{i \geq 1}$, that is, with $S_n = \sum_{i=1}^{n} \xi_i$ for $n \geq 1$, we have $S_n/n \to \mu$ a.s. for $n \to \infty$. By Lemma A.1 we have $N_t \to \infty$ a.s. as $t \to \infty$, hence, $S_{N_t}/N_t \to \mu$ 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 a.s., we obtain $N_t/t \to 1/\mu$ a.s. as $t \to \infty$. This implies, as $n \to \infty$, almost surely

$$\frac{N_{nt} - N_{ns}}{n(t - s)} = \frac{t}{t - s} \frac{N_{nt}}{nt} - \frac{s}{t - s} \frac{N_{ns}}{ns} \longrightarrow \frac{t}{t - s} \frac{1}{\mu} - \frac{s}{t - s} \frac{1}{\mu} = \frac{1}{\mu}. \qquad \square$$

The next result will secure that the events in the different windows will evolve properly in time.

LEMMA A.3. *Let $(N_t)_{t \geq 0}$ be a counting process with $N_0 = 0$ such that for some $\mu > 0$ and for all $0 \leq s < t$ we have $N_{nt} - N_{ns} \sim n(t-s)/\mu$ almost surely. Further, let $V_1, V_2, \ldots$ be a sequence of independent random variables that satisfies the SLLN. Then for all $0 \leq s < t$ we have, as $n \to \infty$, almost surely*

$$\frac{1}{N_{nt} - N_{ns}} \sum_{i=N_{ns}+1}^{N_{nt}} V_i \longrightarrow c.$$

PROOF. Note that choosing $s = 0$ in the statement of the lemma implies $N_t \sim t/\mu$ a.s., such that we find $N_t \to \infty$ as $t \to \infty$. Then we calculate (for $N_{ns} > 0$, the case $N_{ns} = 0$ being similar)

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

so that, for $n \to \infty$,

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

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

$$\longrightarrow \frac{t}{t-s} \left(c - \frac{s}{t} c \right) = c \qquad \text{a.s.} \qquad \square$$

COROLLARY A.4. *Let $(v_i)_{i \geq 1}$ be a sequence in $\mathbb{R}$ with $(1/n) \sum_{i=1}^{n} v_i \to c$ as $n \to \infty$. Then for all $0 \leq s < t$, as $n \to \infty$, we have*

$$\frac{1}{n(t-s)} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} v_i \to c.$$

Finally, we provide a result related to Lemma A.3 for the Lindeberg condition which will be used below to apply the Lindeberg–Feller CLT for triangular schemes.

LEMMA A.5. *Let $\{\xi_i\}_{i \geq 1}$ be a sequence of independent, square-integrable random variables satisfying conditions (2), (3) and (4). Then, for all $0 \leq s < t$ and all $\varepsilon > 0$ we have, as $n \to \infty$,*

$$\frac{1}{s_n^2(s,t)} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \mathbb{E}\big[(\xi_i - \mu)^2 \mathbb{1}_{\{(\xi_i - \mu)^2 > \varepsilon^2 s_n^2(s,t)\}}\big] \longrightarrow 0,$$

where $s_n^2(s,t) := \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \mathrm{Var}(\xi_i)$.

PROOF.   Let $0 \leq s < t$. Condition (3) and Corollary A.4 imply, as $n \to \infty$,

$$s_n^2(s,t) \sim (t-s)\sigma^2 n \sim \frac{t-s}{t} \sum_{i=1}^{\lfloor nt \rfloor} \mathrm{Var}(\xi_i).$$

For $\varepsilon > 0$ set $\eta := \varepsilon\sqrt{(t-s)/(2t)}$. It follows the existence of an element $n_0 = n_0(s,t) \in \mathbb{N}$ so that for all $n > n_0$ and an appropriate null sequence $o(1)$ we have

$$\varepsilon^2 s_n^2(s,t) = (1+o(1))\varepsilon^2 \frac{t-s}{t} \sum_{i=1}^{\lfloor nt \rfloor} \mathrm{Var}(\xi_i) > \eta^2 \sum_{i=1}^{\lfloor nt \rfloor} \mathrm{Var}(\xi_i).$$

Thus, for $n > n_0$ we obtain

$$\frac{1}{s_n^2(s,t)} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \mathbb{E}\big[(\xi_i - \mu)^2 \mathbb{1}_{\{(\xi_i-\mu)^2 > \varepsilon^2 s_n^2(s,t)\}}\big]$$

$$\leq (1+o(1)) \frac{t}{t-s} \frac{1}{\sum_{i=1}^{\lfloor nt \rfloor} \mathrm{Var}(\xi_i)} \sum_{i=1}^{\lfloor nt \rfloor} \mathbb{E}\big[(\xi_i - \mu)^2 \mathbb{1}_{\{(\xi_i-\mu)^2 > \varepsilon^2 s_n^2(s,t)\}}\big]$$

$$\leq (1+o(1)) \frac{t}{t-s} \frac{1}{\sum_{i=1}^{\lfloor nt \rfloor} \mathrm{Var}(\xi_i)} \sum_{i=1}^{\lfloor nt \rfloor} \mathbb{E}\big[(\xi_i - \mu)^2 \mathbb{1}_{\{(\xi_i-\mu)^2 > \eta^2 \sum_{i=1}^{\lfloor nt \rfloor} \mathrm{Var}(\xi_i)\}}\big],$$

and the last expression tends to zero due to condition (4). $\square$

A.2. Convergence of the rescaled counting process.   In this subsection we show that the counting process $(N_t)_{t \geq 0}$ as in (1) properly normalized converges weakly to a standard Brownian motion.

For an RPVV Φ with parameters μ and σ^2, the rescaled version of the corresponding counting process $(N_t)_{t \geq 0}$ is given by

$$(29) \qquad\qquad Z_t^{(n)} := \frac{N_{nt} - nt/\mu}{\sqrt{n}\sqrt{\sigma^2/\mu^3}}, \qquad t \geq 0.$$

The present subsection is devoted to the proof of this proposition:

PROPOSITION A.6.   *Let Φ be an RPVV with associated parameters μ and σ^2. Further, let $(W_t)_{t \geq 0}$ be a standard Brownian motion. Then, in $(D[0,\infty), d_{\mathrm{SK}})$ we have the convergence, as $n \to \infty$, in distribution*

$$(Z_t^{(n)})_{t \geq 0} \xrightarrow{d} (W_t)_{t \geq 0}.$$

For the proof of Proposition A.6 note that we have the following result from Billingsley (1999), Theorem 14.6:

PROPOSITION A.7. *Let $\{\xi_i\}_{i\geq 1}$ be a sequence of positive random variables and $(W_t)_{t\geq 0}$ be a standard Brownian motion. Assume the existence of positive constants μ and σ, so that the rescaled process $(X_t^{(n)})_{t\geq 0}$ defined via*

$$(30) \qquad X_t^{(n)} := \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{[nt]}(\xi_i - \mu), \qquad t \geq 0,$$

converges weakly to $(W_t)_{t\geq 0}$ in $(D[0,\infty), d_{SK})$. Then, the rescaled counting process $Z^{(n)} := (Z_t^{(n)})_{t\geq 0}$ defined in (29) converges weakly to $(W_t)_{t\geq 0}$ in $(D[0,\infty), d_{SK})$.

Since convergence in $(D[0,\infty), d_{\|\cdot\|})$ implies convergence in $(D[0,\infty), d_{SK})$, Proposition A.6 is proved if the conditions in Proposition A.7 are satisfied. Thus, it remains to show the following proposition:

PROPOSITION A.8. *Let Φ be an RPVV with associated parameters μ and σ^2 and corresponding life times $\{\xi_i\}_{i\geq 1}$. For $n = 1, 2, \ldots$ let the processes $X^{(n)}$ be defined as in (30). Then it holds in $(D[0,\infty), d_{\|\cdot\|})$ as $n \to \infty$ that*

$$\left(X_t^{(n)}\right)_{t\geq 0} \xrightarrow{d} (W_t)_{t\geq 0}.$$

For the proof of Proposition A.8 we first show that $(X_t^{(n)})_{t\in[0,\tau]}$ converges weakly to $(W_t)_{t\in[0,\tau]}$ in $(D[0,\tau], d_{\|\cdot\|})$ for $\tau > 0$, which is the subject of the following Lemma A.9. Afterward, we present the proof of Proposition A.8, which then merely consists of extending the result of Lemma A.9 from the interval $[0,\tau]$ to $[0,\infty)$.

LEMMA A.9. *Let Φ be an RPVV with associated parameters μ and σ^2 and corresponding life times $\{\xi_i\}_{i\geq 1}$ and $\tau > 0$. For $n = 1, 2, \ldots$ let the processes $(X_t^{(n)})_{t\geq 0}$ be defined as in (30). Then it holds in $(D[0,\tau], d_{\|\cdot\|})$ as $n \to \infty$ that*

$$\left(X_t^{(n)}\right)_{t\in[0,\tau]} \xrightarrow{d} (W_t)_{t\in[0,\tau]}.$$

For the proof of Lemma A.9 we use the following construction of processes which connects $(X_t^{(n)})_{t\geq 0}$ and its restriction $(X_t^{(n)})_{t\in[0,\tau]}$ to the setting of RPVVs.

CONSTRUCTION A.10. *Let Φ be an RPVV with corresponding parameters μ and σ^2 and life times $\{\xi_i\}_{i\geq 1}$. Let $(X_t^{(n)})_{t\geq 0}$ be constructed from $\{\xi_i\}_{i\geq 1}$ as in (30). For $n = 1, 2, \ldots$ and $\tau > 0$ let the restriction of time to $[0,\tau]$ be denoted by $Y^{(n)} := (X^{(n)})_{t\in[0,\tau]}$. Further denote the restriction of the standard Brownian motion as $Y = (W_t)_{t\in[0,\tau]}$.*

To prepare the proof of Lemma A.9 note that we have the following Theorem A.11 from Pollard (1984), Section V, Theorem 19 (where we adjust the time interval appropriately):

THEOREM A.11. *Let* $\tau > 0$ *and* $Y, Y^{(1)}, Y^{(2)}, \ldots$ *be random elements of* $(D[0, \tau], d_{\|\cdot\|})$, *each with independent life times. Suppose Y has continuous sample paths. Then, as* $n \to \infty$, *we have* $Y^{(n)} \xrightarrow{d} Y$ *in* $(D[0, \tau], d_{\|\cdot\|})$ *if and only if*

1. $Y_0^{(n)} \xrightarrow{d} Y_0$.

2. *For all* s, t *with* $0 \leq s < t \leq \tau$ *we have* $Y_t^{(n)} - Y_s^{(n)} \xrightarrow{d} Y_t - Y_s$.

3. *For all* $\varepsilon > 0$ *there exist* $\alpha > 0, \beta > 0$ *and* $n_0 \in \mathbb{N}$, *such that* $P(|Y_t^{(n)} - Y_s^{(n)}| < \varepsilon) \geq \beta$ *for all* $t, s \in [0, \tau]$ *with* $0 \leq t - s < \alpha$ *and all* $n \geq n_0$.

PROOF OF LEMMA A.9. We apply Theorem A.11 to our setting of RPVVs: the $Y, Y^{(1)}, Y^{(2)}, \ldots$ from Construction A.10 have independent increments and Y has continuous sample paths. We now verify that $Y, Y^{(1)}, Y^{(2)}, \ldots$ from Construction A.10 fulfill conditions 1–3 of Theorem A.11: condition 1 is clear. For condition 2 note that for all $n \geq 1$ and all $0 \leq s < t \leq \tau$ the increment $Y_t^{(n)} - Y_s^{(n)}$ is the sum of elements of a triangular scheme. The nth row of this scheme is of the type $\{(\xi_{i_{s_n}} - \mu)/\sigma\sqrt{n}, (\xi_{i_{s_n}+1} - \mu)/\sigma\sqrt{n}, \ldots, (\xi_{i_{t_n}} - \mu)/\sigma\sqrt{n}\}$, hence, it consists of independent random variables. For the variance of the increments we have

$$\text{Var}(Y_t^{(n)} - Y_s^{(n)}) = \frac{1}{n\sigma^2} \sum_{i=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor} \text{Var}(\xi_i)$$

$$= (t - s)\frac{1}{\sigma^2}\frac{1}{n(t-s)} \sum_{i=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor} \text{Var}(\xi_i) \longrightarrow t - s,$$

for $n \to \infty$, where we use condition (3) and Corollary A.4.

Due to condition (4) and Lemma A.5, the Lindeberg condition is satisfied for the corresponding triangle scheme, so that the Lindeberg–Feller CLT implies, as $n \to \infty$,

$$Y_t^{(n)} - Y_s^{(n)} \xrightarrow{d} \mathcal{N}(0, t - s).$$

Now, for condition 3 let $\varepsilon > 0$. For all $0 \leq s < t \leq \tau$ Chebyshev's inequality implies

$$P(|Y_t^{(n)} - Y_s^{(n)}| < \varepsilon) = 1 - P(|Y_t^{(n)} - Y_s^{(n)}| \geq \varepsilon) \geq 1 - \frac{1}{\varepsilon^2}\text{Var}(Y_t^{(n)} - Y_s^{(n)})$$

$$= 1 - \frac{1}{\varepsilon^2}\left((t - s)\frac{1}{\sigma^2}\frac{1}{n(t-s)} \sum_{i=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor} \text{Var}(\xi_i)\right)$$

$$\geq 1 - c_\varepsilon(t - s) =: \beta,$$

where we use condition (5), so that the constant c_ε does not depend on s, t and n. Now choose $\alpha > 0$ sufficiently small such that $\beta > 0$.

Hence, all conditions of Theorem A.11 are satisfied, thus, we obtain that the processes $Y^{(n)}$ converge weakly to $Y = (W_t)_{t\in[0,\tau]}$ in $(D[0,\tau], d_{\|\cdot\|})$ for $n \to \infty$. □

Finally, we extend Lemma A.9 to the time interval $[0, \infty)$ and hence prove Proposition A.8. We use the following theorem from Pollard (1984), Section V, Theorem 23:

THEOREM A.12. *Let $X, X^{(1)}, X^{(2)}, \ldots$ be random elements of $D[0, \infty)$, with $X \in C$ a.s., for some separable set $C \subset (D[0, \infty), d_{\|\cdot\|})$. Then, with convergence $n \to \infty$, the following statements are equivalent:*

$$(31) \qquad X^{(n)} \xrightarrow{d} X \qquad in\ \big(D[0, \infty), d_{\|\cdot\|}\big),$$

$$(32) \qquad \big(X_t^{(n)}\big)_{t\in[0,\tau]} \xrightarrow{d} (X_t)_{t\in[0,\tau]} \qquad in\ \big(D[0, \tau], d_{\|\cdot\|}\big)\ for\ all\ \tau > 0.$$

PROOF OF PROPOSITION A.8. We apply Theorem A.12: let $X, X^{(n)}, Y$ and $Y^{(n)}$ be derived from Construction A.10. Note that $C[0, \infty)$ is a closed, separable subset of $(D[0, \infty), d_{\|\cdot\|})$; see Pollard (1984), page 108. Condition (32) has been shown in Lemma A.9. Hence, Theorem A.12 applies and we obtain Proposition A.8. □

A.3. Consistency of the estimators. Here we show the almost sure uniform convergence of our estimator $(\hat{s})_{t\in\tau_h}$ defined in equation (20). This will be needed for the proof of Theorem 3.1 to exchange the denominator of $G_{h,t}^{(n)}$ with an empirical normalization by application of Slutsky's theorem. Note that for an a.s. constant stochastic process in $D[h, T - h]$, say, with constant c, we write $(c)_{t\in\tau_h}$.

We have the following consistency result for our estimator $(\hat{s})_{t\in\tau_h}$:

PROPOSITION A.13. *Let Φ be an RPVV with corresponding parameters μ and σ^2. Let $T > 0$, $h \in (0, T/2)$ and $\hat{s}^2(t, h)$ be as defined in equation (20). Then we have in $(D[h, T - h], d_{\|\cdot\|})$, as $n \to \infty$, almost surely*

$$\left(\frac{\hat{s}^2(nt, nh)}{n}\right)_{t\in\tau_h} \longrightarrow \left(\frac{2h\sigma^2}{\mu^3}\right)_{t\in\tau_h}.$$

PROOF. We show the uniform a.s. convergence of $(\hat{\mu}_{\mathrm{le}})_{t\in\tau_h}$ and $(\hat{\mu}_{\mathrm{ri}})_{t\in\tau_h}$ to the constant μ in Lemma A.15, and the uniform a.s. convergence of $(\hat{\sigma}_{\mathrm{le}}^2)_{t\in\tau_h}$ and $(\hat{\sigma}_{\mathrm{ri}}^2)_{t\in\tau_h}$ to the constant σ^2 in Lemma A.16. Uniform a.s. convergence interchanges with sums in general and with products if the limits are constant. Hence, Lemmas A.15 and A.16 and the form of the estimator $\hat{s}^2$ in (20) imply the assertion. □

In the rest of the section we show the uniform a.s. convergence of the estimators $(\hat{\mu}_{\mathrm{ri}})_{t \in \tau_h}$ and $(\hat{\sigma}_{\mathrm{ri}}^2)_{t \in \tau_h}$, respectively, $(\hat{\mu}_{\mathrm{le}})_{t \in \tau_h}$ and $(\hat{\sigma}_{\mathrm{le}}^2)_{t \in \tau_h}$ (see Lemmas A.15 and A.16), as needed in the latter proof. We start with a uniform a.s. result for the scaled counting process $(N_t)_{t \geq 0}$.

LEMMA A.14. *Let Φ be an RPVV with associated mean μ. Let $T > 0$, $h \in (0, T/2]$. Then we have in $(D[h, T - h], d_{\|\cdot\|})$ a.s. as $n \to \infty$ that*

$$(33) \qquad \left(\frac{N_{n(t+h)} - N_{nt}}{nh/\mu} \right)_{t \in \tau_h} \longrightarrow (1)_{t \in \tau_h},$$

$$(34) \qquad \left(\frac{N_{nt} - N_{n(t-h)}}{nh/\mu} \right)_{t \in \tau_h} \longrightarrow (1)_{t \in \tau_h}.$$

PROOF. We show the first statement (33). The second one (34) follows analogously.

We even prove that in $(D[0, T - h], d_{\|\cdot\|})$ it holds a.s. as $n \to \infty$ that

$$(35) \qquad \left(\frac{N_{n(t+h)} - N_{nt}}{nh/\mu} \right)_{t \in [0, T-h]} \longrightarrow (1)_{t \in [0, T-h]}.$$

It is sufficient to show that almost surely

$$(36) \qquad \begin{aligned} &\lim_{n \to \infty} \sup_{t \in [0, T-h]} \frac{N_{n(t+h)} - N_{nt}}{nh/\mu} \leq 1 \quad \text{and} \\ &\lim_{n \to \infty} \inf_{t \in [0, T-h]} \frac{N_{n(t+h)} - N_{nt}}{nh/\mu} \geq 1. \end{aligned}$$

In order to see the left inequality, we decompose the interval $(0, nT]$ into equidistant sections of length $n\varepsilon$. We use the notation

$$(37) \qquad \left| \lceil x \rceil \right| := \lceil x \rceil + 1, \qquad \left| \lfloor x \rfloor \right| := \lfloor x \rfloor - 1, \qquad x > 0.$$

Then each window $(nt, n(t + h)]$ for $t \in (h, T - h]$ is overlapped by one of the finitely many windows $(kn\varepsilon, kn\varepsilon + n|\lceil h/\varepsilon \rceil|\varepsilon]$ for $k = 0, 1, \ldots, \lceil T/\varepsilon \rceil$. Therefore, we find for all $\varepsilon > 0$

$$\sup_{t \in [0, T-h]} N_{n(t+h)} - N_{nt} \leq \max_{k=0,1,\ldots,\lceil T/\varepsilon \rceil} N_{kn\varepsilon + n|\lceil h/\varepsilon \rceil|\varepsilon} - N_{kn\varepsilon}.$$

Thus,

$$\sup_{t \in [0, T-h]} \frac{N_{n(t+h)} - N_{nt}}{nh/\mu}$$

$$\leq \max_{k=0,1,\ldots,\lceil T/\varepsilon \rceil} \frac{N_{kn\varepsilon + n|\lceil h/\varepsilon \rceil|\varepsilon} - N_{kn\varepsilon + nh}}{nh/\mu} + \max_{k=0,1,\ldots,\lceil T/\varepsilon \rceil} \frac{N_{kn\varepsilon + nh} - N_{kn\varepsilon}}{nh/\mu}.$$

The first summand in the latter display becomes small, since $n\lceil h/\varepsilon\rceil\varepsilon \to nh$ for $\varepsilon \downarrow 0$. More precisely, for every $\delta > 0$, we can appropriately choose $\varepsilon > 0$, so that

$$\max_{k=0,1,\ldots,\lceil T/\varepsilon\rceil} \frac{N_{kn\varepsilon+n\lceil h/\varepsilon\rceil\varepsilon} - N_{kn\varepsilon+nh}}{nh/\mu} \to \frac{\delta}{h}$$

a.s. as $n \to \infty$. The second summand in the latter display converges to 1 a.s. for $n \to \infty$. This is because, due to Lemma A.2, the convergence in Lemma A.14 is already known to hold a.s. for finitely many $t \in [0, T - h]$. Thus, we find a.s. that

$$\lim_{n\to\infty} \sup_{t\in[0,T-h]} \frac{N_{n(t+h)} - N_{nt}}{nh/\mu} \le \frac{\delta}{h} + 1.$$

Since $\delta > 0$ is arbitrary, for small $\varepsilon \downarrow 0$ we obtain a.s. that

$$\lim_{n\to\infty} \sup_{t\in[0,T-h]} \frac{N_{n(t+h)} - N_{nt}}{nh/\mu} \le 1.$$

For the right inequality of (36), we use the same decomposition of the interval $(0, nT]$ into equidistant sections of length $n\varepsilon$. Then each window $(nt, n(t + h)]$ for $t \in (h, T - h]$ overlaps one of the finitely many windows $(kn\varepsilon, kn\varepsilon + n\lfloor h/\varepsilon\rfloor\varepsilon]$ for $k = 0, 1, \ldots, \lfloor (T - h)/\varepsilon\rfloor$. One can apply arguments as for the proof of the left inequality of (36) to find that a.s. for $n \to \infty$

$$\lim_{n\to\infty} \inf_{t\in[0,T-h]} (N_{n(t+h)} - N_{nt})/(nh/\mu) \ge 1.$$

The assertion follows.   □

Next we show the uniform a.s. convergence of the estimators $(\hat{\mu}_{\mathrm{ri}})_{t\in\tau_h}$, $(\hat{\mu}_{\mathrm{le}})_{t\in\tau_h}$, $(\hat{\sigma}_{\mathrm{ri}}^2)_{t\in\tau_h}$ and $(\hat{\sigma}_{\mathrm{le}}^2)_{t\in\tau_h}$. We use that uniform a.s. convergence interchanges with sums in general and with products if the limits are constant. Recall the notation

$$\gamma = \gamma_{\mathrm{ri}}(nt, nh) = \{\xi_i : S_i, S_{i+1} \in (nt, n(t + h)], i = 1, 2, \ldots\}$$
$$= \{\xi_i : i = N_{nt} + 2, \ldots, N_{n(t+h)}\}.$$

We find our empirical quantities from equations (18) and (19) as

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

(38)
$$\text{if } N_{n(t+h)} - N_{nt} > 1,$$

and $\hat{\mu}_{\mathrm{ri}} = 0$ otherwise, and

$$\hat{\sigma}_{\mathrm{ri}}^2 = \hat{\sigma}_{\mathrm{ri}}^2(nt, nh) = \frac{1}{N_{n(t+h)} - N_{nt} - 2} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} (\xi_i - \hat{\mu})^2$$

(39)
$$\text{if } N_{n(t+h)} - N_{nt} > 2,$$

and $\hat{\sigma}_{\mathrm{ri}}^2 = 0$ otherwise.

LEMMA A.15. *Let Φ be an RPVV with associated mean μ. Let $T > 0$, $h \in (0, T/2]$ and further $\hat{\mu}_{\mathrm{le}}$ and $\hat{\mu}_{\mathrm{ri}}$ be defined as in* (18). *Then it holds in* $(D[h, T - h], d_{\|\cdot\|})$ *a.s. as $n \to \infty$ that*

$$\left(\hat{\mu}_{\mathrm{le}}(nt, nh)\right)_{t \in \tau_h} \longrightarrow (\mu)_{t \in \tau_h}, \qquad \left(\hat{\mu}_{\mathrm{ri}}(nt, nh)\right)_{t \in \tau_h} \longrightarrow (\mu)_{t \in \tau_h}.$$

PROOF. Conditions (2) and (5) imply Kolmogorov's conditions (7)–(9) with ξ_i^2 there replaced by ξ_i. Hence, we have the SLLN for $\{\xi_i\}_{i \geq 1}$. Lemmas A.2 and A.3 imply the strong consistency for every fixed t, that is, almost surely as $n \to \infty$

$$(40) \qquad \hat{\mu}_{\mathrm{ri}}(nt, nh) = \frac{1}{N_{n(t+h)} - N_{nt} - 1} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i \longrightarrow \mu.$$

Applying Slutsky's theorem with Lemma A.2, we obtain for every t a.s. as $n \to \infty$

$$(41) \qquad \frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i \longrightarrow \mu.$$

In particular, the a.s. convergence holds for finitely many t simultaneously. In order to show the uniform a.s. convergence of $(\hat{\mu}_{\mathrm{ri}}(nt, nh))$, we first show that in $(D[0, T - h], d_{\|\cdot\|})$ it holds a.s. as $n \to \infty$ that

$$(42) \qquad \left(\frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i\right)_{t \in [0, T-h]} \longrightarrow (\mu)_{t \in [0, T-h]}.$$

Note that as in the proof of Lemma A.14, for (42) it is sufficient to show that almost surely

$$(43) \qquad \begin{aligned} &\lim_{n \to \infty} \sup_{t \in [0, T-h]} \frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i \leq \mu \quad \text{and} \\ &\lim_{n \to \infty} \inf_{t \in [0, T-h]} \frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i \geq \mu. \end{aligned}$$

We use the same decomposition of the interval $(0, nT]$ into equidistant sections of length $n\varepsilon$ as in the proof of Lemma A.14. In order to see the left inequality of (43), let $\varepsilon > 0$. Since the life times are nonnegative, with the notation (37), we can bound

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

$$\leq \max_{k=0,1,\ldots,\lceil T/\varepsilon \rceil} \frac{\mu}{nh} \sum_{i=N_{kn\varepsilon}}^{N_{kn\varepsilon + n\lceil \lceil h/\varepsilon \rceil \rceil \varepsilon}} \xi_i$$

$$\leq \max_{k=0,1,\dots,\lceil T/\varepsilon\rceil} \frac{\mu}{nh} \sum_{i=N_{kn\varepsilon}+nh}^{N_{kn\varepsilon}+n\lceil h/\varepsilon\rceil\varepsilon} \xi_i + \max_{k=0,1,\dots,\lceil T/\varepsilon\rceil} \frac{\mu}{nh} \sum_{i=N_{kn\varepsilon}}^{N_{kn\varepsilon}+nh} \xi_i$$

$$\leq \frac{\mu}{h}\bigl(\lvert\lceil h/\varepsilon\rceil\rvert\varepsilon - h\bigr) + \max_{k=0,1,\dots,\lceil T/\varepsilon\rceil} \frac{\mu}{nh} \sum_{i=N_{kn\varepsilon}}^{N_{kn\varepsilon}+nh} \xi_i.$$

The first summand in the previous line is independent of n and tends to zero as $\varepsilon \downarrow 0$. Further, for every $\varepsilon > 0$, the second summand converges to μ a.s. as $n \to \infty$, according to equation (41). Therefore, the first inequality in (43) holds. The second inequality in (43) can be shown similarly, hence, (42) holds. In particular, we obtain the convergence in $(D[h, T-h], d_{\|\cdot\|})$.

By Slutsky's theorem, (42) and Lemma A.14 yield in $(D[h, T-h], d_{\|\cdot\|})$ a.s. as $n \to \infty$ that

$$(44) \qquad \left(\frac{1}{N_{n(t+h)} - N_{nt} - 1} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i\right)_{t\in\tau_h} \longrightarrow (\mu)_{t\in\tau_h},$$

which is the uniform a.s. consistency of $(\hat{\mu}_{\mathrm{ri}})_{t\in\tau_h}$. In the same way we can conclude the uniform a.s. consistency of $(\hat{\mu}_{\mathrm{le}})_{t\in\tau_h}$. $\quad\square$

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

LEMMA A.16. *Let Φ be an RPVV with variance σ^2. Let $T > 0$, $h \in (0, T/2]$ and further $\hat{\sigma}_{\mathrm{le}}^2$ and $\hat{\sigma}_{\mathrm{ri}}^2$ be defined as in* (19). *Then in $(D[h, T-h], d_{\|\cdot\|})$ a.s. as $n \to \infty$ we have*

$$\bigl(\hat{\sigma}_{\mathrm{le}}^2(nt, nh)\bigr)_{t\in\tau_h} \longrightarrow \bigl(\sigma^2\bigr)_{t\in\tau_h}, \qquad \bigl(\hat{\sigma}_{\mathrm{ri}}^2(nt, nh)\bigr)_{t\in\tau_h} \longrightarrow \bigl(\sigma^2\bigr)_{t\in\tau_h}.$$

PROOF. For $N_{n(t+h)} - N_{nt} > 2$ we decompose

$$\hat{\sigma}_{\mathrm{ri}}^2(nt, nh) = \frac{1}{N_{n(t+h)} - N_{nt} - 2} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} (\xi_i - \hat{\mu}_{\mathrm{ri}})^2$$

$$= \frac{1}{N_{n(t+h)} - N_{nt} - 2} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i^2$$

$$+ \left[-2\hat{\mu}_{\mathrm{ri}}\left(\frac{1}{N_{n(t+h)} - N_{nt} - 2} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i\right) + \hat{\mu}_{\mathrm{ri}}^2\right].$$

The expression in the squared brackets as a process in $t \in \tau_h$ converges to $(-\mu^2)_{t\in\tau_h}$ a.s. in $(D[h, T-h], d_{\|\cdot\|})$ due to the consistency of $(\hat{\mu}_{\mathrm{ri}})_{t\in\tau_h}$; see Lemma A.15.

It remains to show that in $(D[h, T - h], d_{\|\cdot\|})$ a.s. as $n \to \infty$

$$
(45) \qquad \left(\frac{1}{N_{n(t+h)} - N_{nt} - 2} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i^2 \right)_{t \in \tau_h} \longrightarrow (\sigma^2 + \mu^2)_{t \in \tau_h}.
$$

We abbreviate $\sigma_i^2 := \mathrm{Var}(\xi_i)$ and center $(\xi_i^*)^2 := \xi_i^2 - (\sigma_i^2 + \mu^2)$, so that

$$
\frac{1}{N_{n(t+h)} - N_{nt} - 2} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i^2 = \frac{1}{N_{n(t+h)} - N_{nt} - 2} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} (\xi_i^*)^2
$$

$$
+ \left[\left(\frac{1}{N_{n(t+h)} - N_{nt} - 2} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \sigma_i^2 \right) + \mu^2 \right].
$$

For fixed t the term in the squared brackets converges to $\sigma^2 + \mu^2$ a.s., as $n \to \infty$, by condition (3) and Lemma A.3. Furthermore, condition (6) now writes $(1/n) \sum_{i=1}^n (\xi_i^*)^2 \to 0$ almost surely. Hence, Lemma A.3 implies for fixed t

$$
(46) \qquad \frac{1}{N_{n(t+h)} - N_{nt} - 2} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} (\xi_i^*)^2 \longrightarrow 0 \qquad \text{a.s.,}
$$

as $n \to \infty$.

Thus, for finitely many t we have the convergence in (45) a.s. toward $\sigma^2 + \mu^2$. In order to obtain the convergence in $(D[h, T - h], d_{\|\cdot\|})$, we proceed as in the proofs of Lemmas A.14 and A.15 and show a.s. as $n \to \infty$ that

$$
(47) \qquad \left(\frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i^2 \right)_{t \in \tau_h} \longrightarrow \sigma^2 + \mu^2.
$$

We again prove this claim even for $t \in [0, T - h]$. Hence, it suffices to show a.s. that

$$
(48) \qquad
\begin{aligned}
&\lim_{n \to \infty} \sup_{t \in [0, T-h]} \frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i^2 \leq \sigma^2 + \mu^2 \quad \text{and} \\
&\lim_{n \to \infty} \inf_{t \in [0, T-h]} \frac{\mu}{nh} \sum_{i=N_{nt}+2}^{N_{n(t+h)}} \xi_i^2 \geq \sigma^2 + \mu^2.
\end{aligned}
$$

As in the previous proofs, for an $\varepsilon > 0$, we decompose the time interval $[0, nT]$ into equidistant sections of length $n\varepsilon$ and, with notation (37), bound

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

$$
\leq \max_{k=0,1,\ldots,\lceil T/\varepsilon \rceil} \frac{\mu}{nh} \sum_{i=N_{kn\varepsilon}}^{N_{kn\varepsilon + n\lceil h/\varepsilon \rceil \varepsilon}} \xi_i^2
$$

$$\leq \max_{k=0,1,\dots,\lceil T/\varepsilon\rceil} \frac{\mu}{nh} \sum_{i=N_{kn\varepsilon+nh}}^{N_{kn\varepsilon+n\lceil h/\varepsilon\rceil\varepsilon}} \xi_i^2 + \max_{k=0,1,\dots,\lceil T/\varepsilon\rceil} \frac{\mu}{nh} \sum_{i=N_{kn\varepsilon}}^{N_{kn\varepsilon+nh}} \xi_i^2.$$

For $\delta := \lceil h/\varepsilon\rceil\varepsilon - h + \varepsilon$ we find a.s. for $n \to \infty$,

$$\max_{k=0,1,\dots,\lceil T/\varepsilon\rceil} (N_{kn\varepsilon+n\lceil h/\varepsilon\rceil\varepsilon} - N_{kn\varepsilon+nh})/(\delta n/\mu) \to 1.$$

Then for $n \to \infty$ the first summand in the latter display converges to $(\delta/h)(\sigma^2 + \mu^2)$ a.s. and the second summand to $\sigma^2 + \mu^2$ a.s., since we have the convergence (45) for finitely many t. Since δ can be chosen arbitrarily small, we find the first inequality of (48). The second follows analogously, and the convergence in (47) follows. There, we exchange the normalization according to Lemma A.14 and obtain (45). Thus, the a.s. uniform consistency of the variance estimator $(\hat{\sigma}_{\mathrm{ri}}^2)_{t\in\tau_h}$ is proven. The uniform a.s. convergence of $(\hat{\sigma}_{\mathrm{le}}^2)_{t\in\tau_h}$ is obtained analogously. $\quad\square$

A.4. Proof of Theorem 3.1. Finally, we put the pieces of the previous subsections together to prove Theorem 3.1:

PROOF OF THEOREM 3.1. Let Φ be an RPVV with associated parameters μ and σ^2 and conditions as is Theorem 3.1. The associated counting process is denoted by $(N_t)_{t\geq 0}$; cf. (1). Further, let $T > 0$ and $h \in (0, T/2]$ denote a window size.

From Proposition A.6 we have that the normalization of $(N_t)_{t\geq 0}$ given by

$$Z_t^{(n)} = \frac{N_{nt} - nt/\mu}{\sqrt{n}\sqrt{\sigma^2/\mu^3}}, \qquad t \geq 0,$$

converges, as $n \to \infty$ in distribution in $(D[0,\infty), d_{\mathrm{SK}})$ to a standard Brownian motion:

$$(49) \qquad\qquad \left(Z_t^{(n)}\right)_{t\geq 0} \xrightarrow{d} (W_t)_{t\geq 0}.$$

Now, for technical reasons we define an auxiliary process, for $t \geq 0$ and $h \in (0, T/2]$, by

$$\Gamma_{t,h}^{(n)} := \frac{(N_{n(t+h)} - N_{nt}) - (N_{nt} - N_{n(t-h)})}{\sqrt{2hn\sigma^2/\mu^3}}.$$

In comparison with the $G_{t,h}^{(n)}$ defined in (22), note that the $\Gamma_{t,h}^{(n)}$ are normalized deterministically with the order of the estimator $\hat{s}$ used for normalization in (22). Now, we apply the continuous mapping theorem as follows: the map $\varphi : (D[0,\infty), d_{\mathrm{SK}}) \to (D[h, T-h], d_{\mathrm{SK}})$ defined by

$$f(t) \overset{\varphi}{\longmapsto} \frac{(f(t+h) - f(t)) - (f(t) - f(t-h))}{\sqrt{2h}}$$

is continuous. With the process $(Z^{(n)})_{t \geq 0}$ defined in (29), we have $\varphi((Z^{(n)})_{t \geq 0}) = (\Gamma_{h,t}^{(n)})_{t \in \tau_h}$. Furthermore, the process $(L_{h,t})_{t \in \tau_h}$ defined in (23) is distributed as $\varphi((W_t)_{t \geq 0})$ with a standard Brownian motion $(W_t)_{t \geq 0}$. Hence, the convergence (49) and the continuous mapping theorem imply the weak convergence in Skorokhod topology of $(\Gamma_{h,t}^{(n)})_{t \in \tau_h}$ to $(L_{h,t})_{t \in \tau_h}$.

By Proposition A.13 we have in $(D[h, T-h], d_{\|\cdot\|})$ a.s. as $n \to \infty$ that

$$(50) \qquad \left(\frac{\hat{s}(nt, nh)^2}{n} \right)_{t \in \tau_h} \longrightarrow \left(\frac{2h\sigma^2}{\mu^3} \right)_{t \in \tau_h}.$$

Since we have the relation

$$G_{h,t}^{(n)} = \frac{\sqrt{2nh\sigma^2/\mu^3}}{\hat{s}(nt, nh)} \Gamma_{h,t}^{(n)},$$

we conclude by Slutsky's theorem with (50) that in $(D[h, T-h], d_{\mathrm{SK}})$ it holds

$$(G_{h,t}^{(n)})_{t \in \tau_h} \xrightarrow{d} (L_{h,t})_{t \in \tau_h}.$$

This is the assertion. $\quad \square$

SUPPLEMENTARY MATERIAL

Supplement to "A multiple filter test for the detection of rate changes in renewal processes with varying variance" (DOI: 10.1214/14-AOAS782SUPP; .r). We provide the R-Code for the multiple filter algorithm.

REFERENCES

ANTOCH, J. and HUŠKOVÁ, M. (1994). Procedures for the detection of multiple changes in series of independent observations. In *Asymptotic Statistics* (*Prague*, 1993). *Contrib. Statist.* 3–20. Physica, Heidelberg. MR1311925

BASSEVILLE, M. and NIKIFOROV, I. V. (1993). *Detection of Abrupt Changes*: *Theory and Application*. Prentice Hall, Englewood Cliffs, NJ. MR1210954

BENDA, J. and HERZ, A. V. M. (2003). A universal model for spike-frequency adaptation. *Neural Comput.* **15** 2523–2564.

BERTRAND, P. (2000). A local method for estimating change points: The "hat-function." *Statistics* **34** 215–235. MR1802728

BERTRAND, P. R., FHIMA, M. and GUILLIN, A. (2011). Off-line detection of multiple change points by the filtered derivative with p-value method. *Sequential Anal.* **30** 172–207. MR2801138

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

BRODSKY, B. E. and DARKHOVSKY, B. S. (1993). *Nonparametric Methods in Change-Point Problems*. Kluwer Academic, Dordrecht. MR1228205

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

BROWN, E. N., KASS, R. E. and MITRA, P. P. (2004). Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nat. Neurosci.* **7** 456–461.

CSÖRGŐ, M. and HORVÁTH, L. (1987). Asymptotic distributions of pontograms. *Math. Proc. Cambridge Philos. Soc.* **101** 131–139. MR0877707

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, Chichester. MR2743035

DREYER, J. K., HERRIK, K. F., BERG, R. W. and HOUNSGAARD, J. D. (2010). Influence of phasic and tonic dopamine release on receptor activation. *J. Neurosci.* **30** 14273–14283.

FARKHOOI, F., STRUBE-BLOSS, M. and NAWROT, M. P. (2009). Serial correlation in neural spike trains: Experimental evidence, stochastic modelling, and single neuron variability. *Phys. Rev. E* **79** 021905.

GONON, F. G. (1988). Nonlinear relationship between impulse flow and dopamine released by rat midbrain dopaminergic neurons as studied by in vivo electrochemistry. *Neuroscience* **24** 19–28.

GOURÉVITCH, B. and EGGERMONT, J. J. (2007). A nonparametric approach for detection of bursts in spike trains. *J. Neurosci. Methods* **160** 349–358.

GRÜN, S., DIESMANN, M. and AERTSEN, A. (2002). "Unitary Events" in multiple single-neuron activity. II. Non-Stationary data. *Neural Comput.* **14** 81–119.

GRÜN, S., RIEHLE, A. and DIESMANN, M. (2003). Effect of cross-trial nonstationarity on joint-spike events. *Biol. Cybernet.* **88** 335–351.

GRÜN, S. and ROTTER, S., eds. (2010). *Analysis of Parallel Spike Trains*. *Springer Series in Computational Neuroscience* **7**. Springer, New York.

HOWE, M. W., TIERNEY, P. L., SANDBERG, S. G., PHILLIPS, P. E. M. and GRAYBIEL, A. M. (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* **500** 575–579.

HUŠKOVÁ, M. and SLABÝ, A. (2001). Permutation tests for multiple changes. *Kybernetika* (*Prague*) **37** 605–622. MR1877077

JOHNSON, D. H. (1996). Point process models of single-neuron discharges. *J. Comput. Neurosci.* **3** 275–299.

JONES, M. C., MARRON, J. S. and SHEATHER, S. J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* **91** 401–407. MR1394097

KASS, R. E., VENTURA, V. and BROWN, E. N. (2005). Statistical issues in the analysis of neuronal data. *J. Neurophysiol.* **94** 8–25.

KENDALL, D. G. and KENDALL, W. S. (1980). Alignments in two-dimensional random sets of points. *Adv. in Appl. Probab.* **12** 380–424. MR0569434

LEGÉNDY, C. R. and SALCMAN, M. (1985). Bursts and recurrences of bursts in the spike trains of spontaneously active striate cortex neurons. *J. Neurophysiol.* **53** 926–939.

MESSER, M., KIRCHNER, M., SCHIEMANN, J., ROEPER, J., NEININGER, R. and SCHNEIDER, G. (2014). Supplement to "A multiple filter test for the detection of rate changes in renewal processes with varying variance." DOI:10.1214/14-AOAS782SUPP.

MILEYKOVSKIY, B. and MORALES, M. (2011). Duration of inhibition of ventral tegmental area dopamine neurons encodes a level of conditioned fear. *J. Neurosci.* **31** 7471–7476.

MUHSAL, B. (2013). Change-point methods for multivariate autoregressive models and multiple structural breaks in the mean. Dissertation, available at http://nbn-resolving.org/urn:nbn:de:swb:90-355368.

NAWROT, M. P., AERTSEN, A. and ROTTER, S. (1999). Single-trial estimation of neuronal firing rates: From single-neuron spike trains to population activity. *J. Neurosci. Methods* **94** 81–92.

NAWROT, M. P., BOUCSEIN, C., RODRIGUEZ MOLINA, V., RIEHLA, A., AERTSEN, A. and ROTTER, S. (2008). Measurement of variability dynamics in cortical spike trains. *Journal of Neuroscience Methods* **169** 347–390.

PERKEL, D. H., GERSTEIN, G. L. and MOORE, G. P. (1967a). Neuronal spike trains and stochastic point processes. I. The single spike train. *Biophys. J.* **7** 391–417.

PERKEL, D. H., GERSTEIN, G. L. and MOORE, G. P. (1967b). Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys. J.* **7** 419–440.

PETROV, V. V. (1995). *Limit Theorems of Probability Theory*: *Sequences of Independent Random Variables*. Clarendon, Oxford. MR1353441

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York. MR0762984

REDGRAVE, P., RODRIGUEZ, M., SMITH, Y., RODRIGUEZ-OROZ, M. C., LEHERICY, S., BERGMAN, H., AGID, Y., DELONG, M. R. and OBESO, J. A. (2010). Goal-directed and habitual control in the basal ganglia: Implications for Parkinson's disease. *Nat. Rev.*, *Neurosci*. **11** 760–772.

RIEKE, F., WARLAND, D., DE RUYTER VAN STEVENINCK, R. and BIALEK, W. (1999). *Spikes*: *Exploring the Neural Code*. MIT Press, Cambridge, MA. MR1983010

SCHIEMANN, J., KLOSE, V., SCHLAUDRAFF, F., BINGMER, M., SEINO, S., MAGILL, P. J., SCHNEIDER, G., LISS, B. and 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** 1272–1280.

SCHNEIDER, G. (2008). Messages of oscillatory correlograms: A spike train model. *Neural Comput*. **20** 1211–1238. MR2402061

SHIMAZAKI, H. and SHINOMOTO, S. (2007). A method for selecting the bin size of a time histogram. *Neural Comput*. **19** 1503–1527. MR2316960

STAUDE, B., ROTTER, S. and GRÜN, S. (2008). Can spike coordination be differentiated from rate covariation? *Neural Comput*. **20** 1973–1999. MR2419958

STEINEBACH, J. and EASTWOOD, V. R. (1995). On extreme value asymptotics for increments of renewal processes. *J. Statist*. *Plann*. *Inference* **45** 301–312. MR1342102

STEINEBACH, J. and ZHANG, H. Q. (1993). On a weighted embedding for pontograms. *Stochastic Process*. *Appl*. **47** 183–195. MR1239836

TOKDAR, S., XI, P., KELLY, R. C. and KASS, R. E. (2010). Detection of bursts in extracellular spike trains using hidden semi-Markov point process models. *J. Comput*. *Neurosci*. **29** 203–212.

M. MESSER
R. NEININGER
G. SCHNEIDER
INSTITUTE OF MATHEMATICS
GOETHE UNIVERSITY FRANKFURT
ROBERT-MAYER-STR. 10
60325 FRANKFURT
GERMANY
E-MAIL: messer@math.uni-frankfurt.de
        neiningr@math.uni-frankfurt.de
        schneider@math.uni-frankfurt.de

M. KIRCHNER
INSTITUTE OF MEDICAL BIOMETRY
    AND INFORMATICS
HEIDELBERG UNIVERSITY
IM NEUENHEIMER FELD 305
69120 HEIDELBERG
GERMANY
E-MAIL: kirchner@imbi.uni-heidelberg.de

J. SCHIEMANN
INSTITUTE OF NEUROPHYSIOLOGY
NEUROSCIENCE CENTER
GOETHE UNIVERSITY FRANKFURT
THEODOR-STERN-KAI 7
60590 FRANKFURT
GERMANY
AND
CENTRE FOR INTEGRATIVE PHYSIOLOGY
UNIVERSITY OF EDINBURGH
GEORGE SQUARE
EDINBURGH EH8 9XD
UNITED KINGDOM
E-MAIL: Julia.Schiemann@ed.ac.uk

J. ROEPER
INSTITUTE OF NEUROPHYSIOLOGY
NEUROSCIENCE CENTER
GOETHE UNIVERSITY FRANKFURT
THEODOR-STERN-KAI 7
60590 FRANKFURT
GERMANY
E-MAIL: roeper@em.uni-frankfurt.de