

A multi-scale approach for testing and detecting peaks in time series

Michael Messer , Hendrik Backhaus , Ting Fu , Albrecht Stroh & Gaby Schneider

To cite this article: Michael Messer , Hendrik Backhaus , Ting Fu , Albrecht Stroh & Gaby Schneider (2020): A multi-scale approach for testing and detecting peaks in time series, *Statistics*, DOI: [10.1080/02331888.2020.1823980](https://doi.org/10.1080/02331888.2020.1823980)

To link to this article: <https://doi.org/10.1080/02331888.2020.1823980>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 22 Oct 2020.



Submit your article to this journal [↗](#)



Article views: 135





View related articles [↗](#)



View Crossmark data [↗](#)

A multi-scale approach for testing and detecting peaks in time series

Michael Messer ^a, Hendrik Backhaus^b, Ting Fu^c, Albrecht Stroh ^{b,c} and Gaby Schneider^d

^aInstitute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria; ^bLeibniz Institute for Resilience Research, Mainz, Germany; ^cInstitute for Pathophysiology, University Medical Center of the Johannes Gutenberg-University, Mainz, Germany; ^dInstitute of Mathematics, Goethe University Frankfurt, Frankfurt, Germany

ABSTRACT

An approach is presented that combines a statistical test for peak detection with the estimation of peak positions in time series. Motivated by empirical observations in neuronal recordings, we aim at investigating peaks of different heights and widths. We use a moving window approach to compare the differences of estimated slope coefficients of local regression models. We combine multiple windows and use the global maximum of all different processes as a test statistic. After rejection, a multiple filter algorithm combines peak positions estimated from multiple windows. Analysing neuronal activity recorded in anaesthetized mice, the procedure could identify significant differences between two brain states concerning peak occurrences and intermediate down states showing no peaks. This suggests that the method can be useful in the analysis of time series showing variability of peak shapes. The method is implemented in the R-package MFT (available on CRAN).

ARTICLE HISTORY



Received 26 August 2019
Accepted 5 June 2020


KEYWORDS

Brain states; linear regression; multi-scale; neuronal ensembles; peak detection

1. Introduction

In many applications in time series analysis, one can observe more or less regularly occurring fluctuations between periods of lower activity and short peaks of higher activity. We consider here the context of spiking activity of a local ensemble of neuronal cells in layer II/III of primary visual cortex, accumulated within a one-dimensional time series. In such ensembles, spontaneous rhythmic activity can emerge (see Figure 5), consisting of peaks, and intermediate periods of lower activity which will be called down states here. Such functional brain states play an important role in the way sensory inputs are processed and govern neuronal excitability across different spatial scales. We can discern between an awake-like persistent state [1], and a sleep-like slow wave state [2,3]. In particular, the variability of neuronal responses results from the interplay between the ensemble brain state, ongoing spontaneous activity and sensory inputs [4]. Thus, functional brain states

CONTACT Michael Messer  michael.messer@tuwien.ac.at  Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria

 Supplemental data for this article can be accessed here. <https://doi.org/10.1080/02331888.2020.1823980>

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

are possibly confined to local circuits or established as distributed patterns, and may crucially shape the neural processes from perception to action [5]. To disentangle this complex interplay, an unbiased, data-driven approach to classify the current ensemble brain state – as put forward in this study – is mandatory.

It is therefore the aim of the present paper to provide a statistical method with which differences between such time series can be quantified. To that end, we also require a model for time series showing such peaks and down states that should, first, allow for the development of rigorous statistical tools. Second, the model should be able to capture a certain variability in the length and height of peaks and down states present in the data. This is because the associated widths and heights of the peaks can be variable, particularly across functional brain states (see Figure 5). It is therefore the aim of this paper to present a method for the detection of peaks of various heights and widths in sequences of random variables in order to describe functional differences between different brain states in the occurrences of peaks and down states within a neuronal ensemble.

There is a vast literature on the statistical analysis and detection of peaks in time series. On the one hand, procedures for the detection of changes in regression models are well studied in mathematical statistics, with the aim of consistently detecting single breaks [6–10] or multiple breaks [11–13] in regression functions, also see the textbook [14]. Particularly the detection of peaks, meaning a jagged deviation from baseline activity, is studied in various disciplines. For financial time series, [15] proposed an online algorithmic approach based on the difference of slope parameters in a moving window analysis in order to maximize the profitability of investment strategies. We also mention methods from signal processing which often focus on the design of algorithms for peak detection, using e.g., neural networks [16], Hidden Markov Models [17] or wavelet transform [18], to name but a few. For a review in the context of ECG waveform analysis see, e.g., [19]. In the context of neuronal network rhythmic fluctuations between up states (i.e., peaks) and down states, Seamari et al. [20] proposed a method that relies on the crossing points of two moving windows whose optimal size depends on the typical duration of the up state, or peak, that is to be detected. Such signal processing approaches often evaluate the performance with visual inspection or by application to large data sets in which peaks have been labelled by expert knowledge. From a statistical point of view we mention [21] who combine the identification of multiple peaks with a correction for multiple testing in a kernel-based method, or similarly [22] who estimate changes in regression and then apply post hoc correction procedures in a filtered-derivative method. The latter two methods apply a single kernel or bandwidth, which regulates the degree of localization of the data, and thus determines the preferred width of the peak to be detected.

Here we present a method that combines a statistical test for peak detection with the estimation of the peak positions. Specifically, we aim at detecting an unknown number of peaks of unknown and potentially different heights and widths, by application of multiple bandwidths. The idea of the approach is the following.

First, in order to describe time series with downstates and peaks of different peaks and heights, we assume the time series to consist of a jagged function f_t that adds piecewise linear peaks of different heights and widths to a baseline level and is associated with independent and identically distributed errors (Section 2.1).

Second, for the identification of the positions of the peaks, we use a sliding window approach with two adjacent subwindows of fixed size h . In each subwindow, we apply linear

regression and then study the normed difference of the slope estimates between the left and right window for every time t (Section 2.2). We show asymptotic properties of the resulting difference process $(D_{h,t})_t$, which allow to define an asymptotic rejection region of a statistical test that uses the temporal maximum of this difference process for an individual window size h .

Third, in order to study peaks of different height and width, we use a combined statistical test that uses multiple bandwidths. This extends the multiple filter test (MFT, Section 2.3) which was introduced in [23] for the analysis of first-order breaks in stochastic point processes. The idea is to use the maximum of $D_{h,t}$ across time t and all windows h in a window set H as a test statistic. Because these multiple difference processes originate from the same time series, it is possible to derive a combined rejection threshold for the resulting test.

Fourth, we discuss an algorithm that locates the peaks in Section 2.4. It starts by searching for maximizers of each difference process $(D_{h,t})_t$ for individual h , providing sets of peak candidates. These candidates are then combined over all $h \in H$ by elimination of candidates that are likely to refer to the same underlying peak. We also give recommendations on the choice of the window sizes.

In Section 3, we analyse neuronal ensemble activity recorded using high-speed 2-photon calcium imaging upon expression of the genetically encoded calcium indicator GCaMP6f, to assess the optical correlate of spiking activity of the entire local neuronal ensemble with single-cell-resolution. First, we analyse the occurrences of peaks, from which we then identify down states. The results suggest interesting differences between the two investigated brain states. After the discussion in Section 4, we give proofs and auxiliary results in the Appendix. The procedure for peak detection is implemented in the R-package MFT ([24] available on CRAN).

2. Materials and methods

2.1. The model

In order to describe time series with peaks of different height and width, we use a model that describes a time series as a realization of a sequence of random variables X_1, X_2, \dots, X_T following

$$X_t = \gamma + f_t + \varepsilon_t,$$

while $\gamma \in \mathbb{R}$ denotes a baseline constant, f denotes a jagged function that adds peaks of different height and width, and i.i.d. and square-integrable random variables $(\varepsilon_t)_t$ with $\mathbb{E}[\varepsilon_t] = 0$ and $\text{Var}(\varepsilon_t) = \sigma^2$ for all t (Figure 1 (A)). For the formulation of peaks, we define the peak function $f: [0, T] \rightarrow \mathbb{R}_0^+$ as a non-negative, continuous and piecewise linear jagged function with baseline zero, see Figure 1(A) (red). More precisely, we decompose the domain $[0, T]$ into disjoint, increasingly ordered and right-closed intervals with bounds in the integers. In each interval we assume f to be a linear function $f_t = \alpha + \beta t$, with intercept α and slope β depending on the interval (see Figure 2(A), interval numbers $i = 1, 2, \dots$ indicated in the bottom of the figure). Provided the following intervals exist, we make two assumptions: first, we assume that in all intervals numbered as $i \in \{1 + 3m | m = 0, 1, 2, \dots\}$, we find α and β to vanish, i.e., $f = 0$. Second, in the intervals $i \in \{2 + 3m | m = 0, 1, 2, \dots\}$, we assume the slope β to be positive. As a consequence of continuity, the remaining intervals $i \in \{3m | m = 1, 2, \dots\}$ have negative slope, decreasing

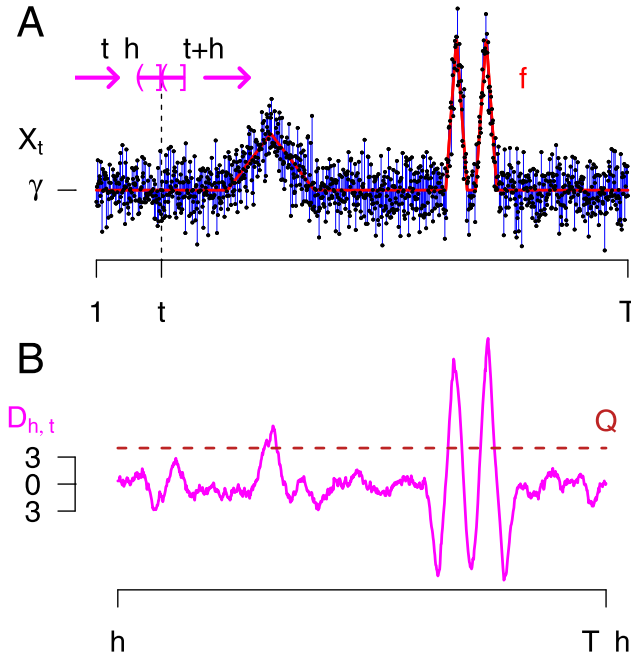


Figure 1. (A) A process $(X_t)_t$ with three peaks. The X_t 's (black points) equal the sum of the baseline constant γ , the peak function f (red) and the random errors ε_t (blue vertical lines). The double-window of size $2h$ (magenta) indicates all X_t 's from which the difference $D_{h,t}$ is calculated. (B) The difference process $(D_{h,t})_t$ (magenta) derived from the process $(X_t)_t$. Three local maxima lie above the rejection threshold Q (brown), and their maximizers lie close to the peak points of f .

back to baseline level, such that the endpoint of an interval of the second type describes a local maximum of f and is thus denoted a peak point c .

For an example with two peaks and seven intervals with constant slope see Figure 2(A). Note that the construction of f with vanishing slope at zero can be extended naturally to the case in which f starts within the area of a peak, allowing also for positive or negative slopes at the origin. The set of peak points is denoted by C . For a peak point c we call f_c the corresponding peak height. In Figure 1(A) we find three peaks. The standard deviation $\sigma > 0$ is assumed to be constant within each section of constant slope, but may change between intervals of different slope. The set of all processes $(X_t)_t$ constructed in this fashion constitutes the model set \mathcal{M} .

One aim is to construct a test of the null hypothesis $H_0 : f = 0$, i.e., $C = \emptyset$, in which $(X_t)_t$ describes a sequence of i.i.d. random variables and f shows no peak. The idea of the statistical test is to fit two simple linear regression models to two adjacent sliding windows and to use the process $(D_{h,t})_t$ of the standardized differences of estimated slope coefficients for peak identification. In Section 2.2, we define this process and recall its elementary properties.

In Section 2.3, we derive its asymptotic properties, which are used then for the derivation of the test statistic and its rejection region. In order to enable asymptotic considerations, we generalize the proposed time series model here first to an asymptotic setting, introducing

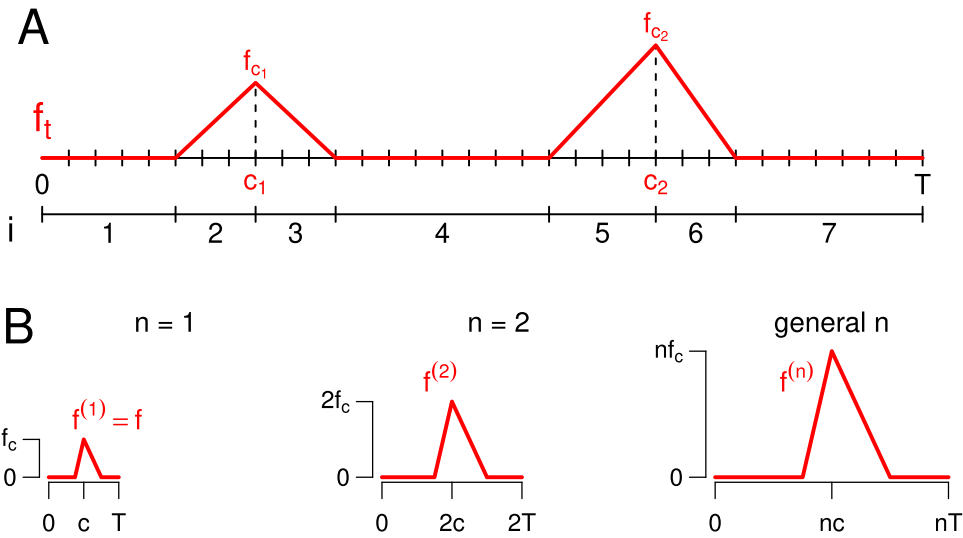


Figure 2. Illustration of the model assumptions. A. The peak function f on $[0, T]$ with peaks at c_1 and c_2 of heights f_{c_1} and f_{c_2} and the corresponding intervals $i = 1, 2, \dots$ of constant slope used in the definition of f . B. The generalized peak function $f^{(n)}$ in the situation of a single peak. By increasing n , the lengths of all linear sections are increased linearly.

$(X_t^{(n)})_{t=1, \dots, nT}$ via

$$X_t^{(n)} := \gamma + f_t^{(n)} + \varepsilon_t \quad (1)$$

with $f^{(n)} : [0, nT] \rightarrow \mathbb{R}_0^+$ given by $f_t^{(n)} := nf_t/n$. This is a triangular setting in which the time, as well as the peak height grow linearly in n , see Figure 2(B). Over different n we find a constant number of linear sections within $f^{(n)}$ while also all slopes of $f^{(n)}$ remain constant. However, by increasing n we linearly increase the number of time instants within the respective sections. For asymptotic statements we let $n \rightarrow \infty$. Oftentimes we suppress the dependence of the generalised process $(X_t^{(n)})_t$ on n and simply write $(X_t)_t$.

2.2. Derivation of the difference process

In the described asymptotic setting of the generalized process $(X_t)_t$ from (1), we pointwise fit two linear regression models. We separately consider the random variables in each of the two adjacent windows. The idea is that the switch from a positive to a negative estimated slope can indicate a peak point. This will be quantified in the difference process $(D_{h,t}^{(n)})_t$ in (9) which represents the standardized difference of the estimated slopes.

In the following, we first discuss the regression model and derive properties of least-squares estimators. This consideration does not depend on the peak process $(X_t)_t$. In a second step, we then fit $(X_t)_t$ and obtain $D_{h,t}^{(n)}$. As a consequence of the linear nature of the regression models, the properties derived for the estimators will inherit to the fit $D_{h,t}^{(n)}$ as long as each of the associated windows entirely lies within a section where the peak function f is constant.

Let $\varepsilon_1^*, \varepsilon_2^*, \dots$ denote a sequence of independent random variables with zero mean and unit variance. For a time point $t \in [h, T - h]$ and $n = 1, 2, \dots$ we assume a regression model

$$Y_k = \eta_j + \beta_j k + \sigma_j \varepsilon_k^* \quad \text{with } k \in I_j, \quad (2)$$

where $\eta_j, \beta_j \in \mathbb{R}$ and $\sigma_j > 0$ for $j \in \{\ell, r\}$, while $I_\ell := \{\lfloor n(t - h) \rfloor + 1, \dots, \lfloor nt \rfloor\}$ and $I_r := \{\lfloor nt \rfloor + 1, \dots, \lfloor n(t + h) \rfloor\}$ denote the windows of size nh to the left and right of t and $\lfloor \cdot \rfloor$ denotes the floor function.

In order to derive the least-squares estimators $\hat{\beta}_j, \hat{\eta}_j$ and $\hat{\sigma}_j^2$ for the model parameters, we use the following notation. For $j \in \{\ell, r\}$, $t \in [h, T - h]$ and $n = 1, 2, \dots$, we calculate the empirical means of the random variables in a set of indices I_j via $\bar{Y}_j := (1/(nh)) \sum_{k \in I_j} Y_k$. The empirical means of the indices are given by

$$\bar{I}_\ell := \lfloor n(t - h) \rfloor + \frac{nh + 1}{2} \quad \text{and} \quad \bar{I}_r := \lfloor nt \rfloor + \frac{nh + 1}{2} \quad (3)$$

because the nh indices of I_j are equidistantly spaced. The following lemma states the least-squares estimates for the slope and intercept.

Lemma 2.1: For $j \in \{\ell, r\}$, $t \in [h, T - h]$ and $n = 1, 2, \dots$ the estimators

$$\hat{\beta}_j := \frac{12}{nh[(nh)^2 - 1]} \sum_{k \in I_j} (k - \bar{I}_j) Y_k, \quad \text{and} \quad (4)$$

$$\hat{\eta}_j := \bar{Y}_j - \hat{\beta}_j \bar{I}_j. \quad (5)$$

are the least-squares estimators of the slope and intercept within the regression models (2).

This follows from elementary statistics, as the regression line hits the centre of mass (\bar{I}_j, \bar{Y}_j) , and because the slope is represented through the correlation of the data Y_k and the explanatory indices I_j . For that we mention the identities

$$\sum_{k \in I_j} (k - \bar{I}_j) = 0 \quad \text{and} \quad \sum_{k \in I_j} (k - \bar{I}_j)^2 = \frac{nh[(nh)^2 - 1]}{12}, \quad (6)$$

also noting that the empirical variance of the time indices I_j writes as $s^2(I_j) = (1/(nh)) \sum_{k \in I_j} (k - \bar{I}_j)^2 = [(nh)^2 - 1]/12$ which equals the variance of a fair die with nh sides, aligning with the equidistant spacing of the elements of I_j .

Let further an estimator of σ_j^2 be given by

$$\hat{\sigma}_j^2 := \frac{1}{nh - 2} \sum_{k \in I_j} (Y_k - (\hat{\eta}_j + \hat{\beta}_j k))^2. \quad (7)$$

This estimator complements the least-squares fit in the sense that $\sum_{k \in I_j} [Y_k - (\eta_j + \beta_j k)]^2$ states the criterion to be minimized. The estimator $\hat{\sigma}_j^2$ is unbiased, and in case of normally distributed errors, associated standard statistical tests can be applied using Student's or Fisher's distributions.

Dependence of $\hat{\beta}_j$, $\hat{\eta}_j$ and $\hat{\sigma}_j^2$ on t , h and n is inherited from \bar{Y}_j and \bar{I}_j . The estimators are well-defined due to the assumption $h \geq 3$. We state asymptotic normality of the slope and strong consistency of all estimators in Lemma 2.2. A proof can be found in the Appendix.

Lemma 2.2: For $\hat{\beta}_j$ from (4) it holds for $j \in \{\ell, r\}$, $t \in [h, T - h]$ and $n = 1, 2, \dots$ that $\mathbb{E}[\hat{\beta}_j] = \beta_j$ and $\text{Var}[\hat{\beta}_j] = 12\sigma_j^2/[nh((nh)^2 - 1)]$. Further, as $n \rightarrow \infty$ it holds $[\text{Var}(\hat{\beta}_j)]^{-1/2}(\hat{\beta}_j - \beta_j) \xrightarrow{d} N(0, 1)$.

It denotes \xrightarrow{d} convergence in distribution. Asymptotic normality results from the Lindeberg-Feller central limit theorem in which respect we rewrite

$$\hat{\beta}_j = \frac{12}{nh[(nh)^2 - 1]} \sum_{k \in I_j} (k - \bar{I}_j) Y_k = \beta_j + \sigma_j \frac{12}{nh[(nh)^2 - 1]} \sum_{k \in I_j} (k - \bar{I}_j) \varepsilon_k^*. \quad (8)$$

while the r.h.s. constitutes a sum of independent but not identically distributed random variables. The second identity in (6) explains the variance of $\hat{\beta}_j$, mentioning the order $\text{Var}(\hat{\beta}_j) = O(1/n^3)$.

Lemma 2.3: For the estimators in (4), (5) and (7) it holds for $j \in \{\ell, r\}$, $t \in [h, T - h]$ as $n \rightarrow \infty$ that almost surely $\hat{\beta}_j \rightarrow \beta_j$, $\hat{\eta}_j \rightarrow \eta_j$ and $\hat{\sigma}_j^2 \rightarrow \sigma_j^2$.

A proof can be found in the Appendix.

We now construct the difference process $(D_{h,t}^{(n)})_t$ by pointwise fitting the regression model (2) to the peak process $(X_t)_t$ from (1). For every time t and every n , we consider the time domains I_ℓ and I_r , and compare the estimated slopes $\hat{\beta}_\ell$ and $\hat{\beta}_r$ via

$$D_{h,t}^{(n)} := \frac{\hat{\beta}_\ell - \hat{\beta}_r}{\hat{s}_{h,t}^{(n)}} \quad (\text{Figure 1B}). \quad (9)$$

The denominator $\hat{s}_{h,t}^{(n)} := \hat{s}_{h,t}^{(n)}$ estimates the standard deviation of the numerator, i.e., $(\text{Var}(\hat{\beta}_\ell - \hat{\beta}_r))^{1/2}$ and is derived separately for every window of size h and every time t . Specifically, we set

$$(\hat{s}_{h,t}^{(n)})^2 := \frac{12}{nh[(nh)^2 - 1]} (\hat{\sigma}_\ell^2 + \hat{\sigma}_r^2) \quad (10)$$

because $\text{Var}(\hat{\beta}_j) = 12\sigma_j^2/[nh((nh)^2 - 1)]$ when the respective window lies in a section of constant slope of the underlying peak function f , compare Lemma 2.2. The asymptotic normality of the $\hat{\beta}_j$ is inherited by the difference process $(D_{h,t}^{(n)})_t$ for those time points where each of the two windows entirely lies in a section of constant slope of f as stated in the following lemma.

Lemma 2.4: Let $(X_t)_t \in \mathcal{M}$ with peak function f , and $t \in [h, T - h]$ such that both windows $(t - h, t]$ and $(t, t + h]$ entirely refer to a section of constant slope of f . Then it holds

as $n \rightarrow \infty$

$$D_{h,t}^{(n)} - \frac{\beta_\ell - \beta_r}{\hat{s}_{h,t}^{(n)}} \xrightarrow{d} N(0, 1),$$

while β_ℓ and β_r denote the slopes of f in the sections with constant slope, and with $D_{h,t}^{(n)}$ in (9) and $\hat{s}_{h,t}^{(n)}$ in (10).

The superscript in $D_{h,t}^{(n)}$ makes the dependence on n explicit. The Lemma first states that under the null hypothesis $f = 0$, i.e., $\beta_\ell = \beta_r (= 0)$, all marginals of the difference process $D_{h,t}^{(n)}$ are approximately standard normally distributed. Further, in case of involved peaks, $D_{h,t}^{(n)} \cdot \hat{s}_{h,t}^{(n)}$ is approximately normally distributed with expectation $(\beta_\ell - \beta_r)$ at those t where each of the windows entirely lies within a section of constant slope of f . In particular, $D_{h,t}^{(n)}$ will show systematic deviations from zero in the positive direction at peak points, while being systematically decreased at the on- and offset of a peak, see the form of $D_{h,t}^{(n)}$ at the rightmost peaks in Figure 1(B). In contrast, it is constructed not to be sensitive to potential deviations of the variance parameter, which suggests a certain robustness of the procedure.

In order to derive a statistical test, we now consider the stochastic process $(D_{h,t}^{(n)})_{t \in [h, T-h]}$. Because positive deviations from zero indicate peaks, a maximum of this process will serve as a test statistic. To this end, for let $(\mathcal{D}_{\mathbb{R}}[h, T-h], d_{SK})$ denote the space of all real-valued càdlàg-functions on $[h, T-h]$ endowed with the Skorokhod topology. As a consequence of the floor function in (2) we find all estimators in (4), (5) and (7), and thus $(D_{h,t}^{(n)})_t$ to constitute step-processes in $(\mathcal{D}_{\mathbb{R}}[h, T-h], d_{SK})$. In Section 2.3, we prove weak convergence of $(D_{h,t}^{(n)})_t$, which is then used for the construction of the statistical test.

2.3. The statistical test

The main idea of the statistical test of the null hypothesis of no peaks, $H_0 : f = 0$, is to use the maximum of the difference process $(D_{h,t}^{(n)})_t$ as a test statistic, where one maximizes both over time t and over multiple windows h . We therefore first study the asymptotic behaviour of the process $(D_{h,t}^{(n)})_t$ for fixed h in Lemma 2.5 and then explain how different window sizes can be used simultaneously.

Proposition 2.5: *Let $(X_t)_t \in \mathcal{M}$ and $f = 0$. Then it holds in $(\mathcal{D}_{\mathbb{R}}[h, T-h], d_{SK})$ as $n \rightarrow \infty$*

$$(D_{h,t}^{(n)})_t \xrightarrow{d} (L_{h,t})_t,$$

where $L_{h,t}$ is given in (12). The proof can be found in the Appendix.

The process $(L_{h,t})$ is a $2h$ -dependent Gaussian process with zero mean and unit variance. It is represented as a functional of a bivariate Gaussian process $GP(0, \Gamma)$ which does not depend on the model parameters as follows. Let $GP(0, \Gamma) := (\mathcal{G}_1(t), \mathcal{G}_2(t))_{t \geq 0}$ be a Gaussian process with continuous sample paths, expectation zero and covariance matrix

$\Gamma = (\Gamma_{s,t})_{s,t \geq 0}$ given by

$$\Gamma_{s,t} := \begin{pmatrix} (s \wedge t)^3/3 & (s \wedge t)^2/2 \\ (s \wedge t)^2/2 & s \wedge t \end{pmatrix}, \quad (11)$$

where both s and t are considered time indices and \wedge denotes the minimum. Note that the second component $(\mathcal{G}_2(t))_{t \geq 0}$ describes a standard Brownian motion, and that the components have independent increments $\text{Cov}([\mathcal{G}_i(t_4) - \mathcal{G}_i(t_3)], [\mathcal{G}_j(t_2) - \mathcal{G}_j(t_1)]) = 0$ for $i, j \in \{1, 2\}$ and $0 < t_1 \leq t_2 \leq t_3 \leq t_4$. The limit process $(L_{h,t})_{t \in [h, T-h]}$ is then given by

$$L_{h,t} := \left(\frac{6}{h^3}\right)^{1/2} \left[\left([\mathcal{G}_1(t) - \mathcal{G}_1(t-h)] - \left(t - \frac{h}{2}\right)[\mathcal{G}_2(t) - \mathcal{G}_2(t-h)] \right) - \left([\mathcal{G}_1(t+h) - \mathcal{G}_1(t)] - \left(t + \frac{h}{2}\right)[\mathcal{G}_2(t+h) - \mathcal{G}_2(t)] \right) \right]. \quad (12)$$

Intuitively, we find that the difference of the round brackets corresponds to the comparison of the left and the right slopes in $D_{h,t}^{(n)}$. Also note that the summands of $\hat{\beta}_j$ in (8) are weighted and rescaled: these are $(k - \bar{I}_j)\varepsilon_k^*$, while the part $k\varepsilon_k^*$ is associated with \mathcal{G}_1 , and in $\bar{I}_j\varepsilon_k^*$ the first factor \bar{I}_j refers to $t \mp h/2$, and the factor ε_k^* is associated with the Brownian motion \mathcal{G}_2 .

Making use of the described weak process convergence of $(D_{h,t}^{(n)})_t$, a rejection threshold Q can be derived as follows. First, to investigate peaks occurring on different time scales, we consider multiple processes $\{(D_{h,t}^{(n)})_{t \in [h, T-h]} | h \in H\}$ induced by a set H of finitely many window sizes. The statistical test is therefore called Multiple Filter Test (MFT). As described above, we consider the maximum M across time and window sizes as a test statistic. Continuous mapping implies convergence of this maximum, i.e.,

$$M^{(n)} = \max_h \max_t D_{h,t}^{(n)} \xrightarrow{d} \max_h \max_t L_{h,t},$$

while all limit processes $\{(L_{h,t})_{t \in [h, T-h]} | h \in H\}$ are functionals of the same Gaussian process $GP(0, \Gamma)$. For that we mention the joint convergence of the multiple peak processes as stated in Corollary A.3 given in the Appendix.

Due to independence from model parameters, a quantile Q of the distribution $M^{(n)}$ can be approximated in simulations. $M := M^{(1)}$ serves as a test statistic, and the null hypothesis $f = 0$ is rejected if $M > Q$. In Figure 3(B), $f = 0$ is rejected as M exceeds Q (brown dotted line). After the rejection of the null hypothesis, the estimation of the peak positions relies on the successive localization of maximizers of $(D_{h,t})_{h,t}$ that exceed Q . A respective multiple filter algorithm is proposed in Subsection 2.4.

Note that Lemma 2.4 implies asymptotic power 1 as the threshold Q is fixed while $M^{(n)}$ increases at any peak point with order $n^{3/2}$ (in case the smallest window does not overlap neighbouring peaks). Also, power 1 is kept when in the asymptotic setup slopes are allowed to decrease, but only slower than $n^{-3/2}$.

2.4. The multiple filter algorithm

In case of a rejection of the null hypothesis, we propose a multiple filter algorithm (MFA, Section 2.4.1) for the estimation of the peak positions, which is adapted here from [23]

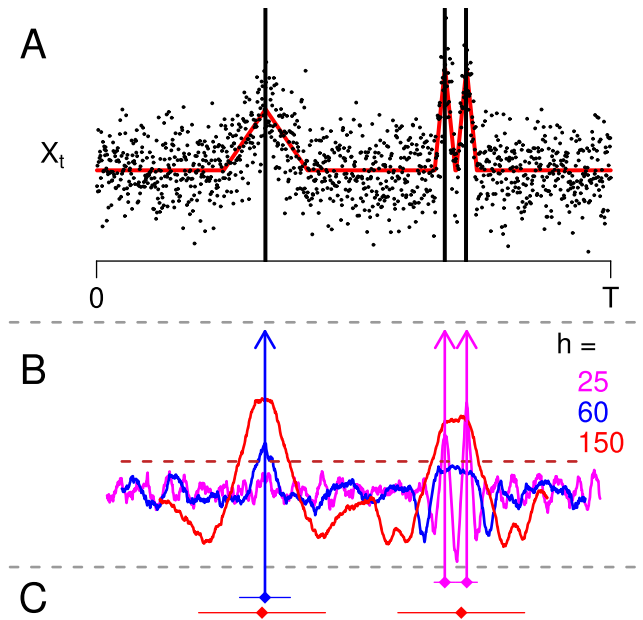


Figure 3. Illustration of the MFA for peak detection. (A) A process $(X_t)_t$ with three peaks. Peak positions indicated by vertical bars. Colors as in Figure 1. (B) The test statistic M exceeds the threshold Q , leading to a rejection of the null hypothesis. Curves indicate the processes $(D_{h,t})_t$ for windows $h \in \{25, 60, 150\}$. (C) Peak candidates of individual windows h , indicated by diamonds in different colours, are derived as successive maximizers of $D_{h,t}$. Peak candidates estimated from window h are included in the set of estimated peaks unless their h -neighbourhood contains a peak candidate already estimated with a smaller window. Arrows indicate final estimated peaks.

and similar to comparable algorithms [e.g., 25]. In Section 2.4.2, its performance is evaluated in simulations, also establishing rules for an appropriate choice of the set of window sizes H .

2.4.1. The MFA

The MFA consists of two steps. First, a separate set of peak candidates is estimated from each process $(D_{h,t})_t$ for each individual window $h \in H$ as follows. As a peak leads to a high positive value of $(D_{h,t})_t$, the argument of the maximum $\max_t D_{h,t}$ is the first peak candidate. Then we note that within the h -neighbourhood of a peak, the value of $D_{h,t}$ will tend to be positive due to a positive expected slope difference between the two windows. Therefore, we eliminate the h -neighbourhood of the estimated peak candidate and then proceed analogously, i.e., add the next maximizer of $(D_{h,t})_t$ to the set of peak candidates and eliminate its h -neighbourhood, until the process $(D_{h,t})_t$ stays below the rejection threshold Q .

In the second part of the algorithm, the sets of peak candidates of the different window sizes are integrated into one set of accepted peak estimates. As large windows may be affected by more than one peak, we start with the set of peak candidates estimated with the smallest window. We then add peak candidates of successively larger windows h if their respective h -neighbourhood does not contain a peak in the set of previously accepted peak estimators.

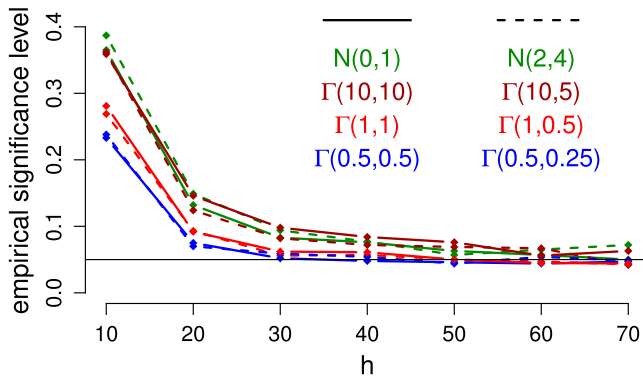


Figure 4. Empirical significance level, i.e., number of simulations in which the null hypothesis was falsely rejected, in simulated time series with $T = 1000$ i.i.d. random variables as a function of the window size $h \in \{10, 20, \dots, 70\}$ with distributions $N(\mu, \sigma^2)$ and $\Gamma(\text{shape}, \text{rate})$ for different parameter combinations indicated in the legend, 1000 simulations per parameter combination. The rejection threshold Q was derived separately in 100.000 simulations according to an asymptotic significance level $\alpha = 0.05$ indicated as a horizontal line.

An example similar to Figure 1 is shown in Figure 3. A process $(X_t)_t$ with three peaks (vertical lines) of different heights, widths and distances are analysed for peaks (Figure 3(A)). Applying three different window sizes yields three sets of peak candidates indicated as diamonds in Figure 3(C). Only the smallest window can detect the peaks in close succession but fails to detect the lower peak, which is estimated by the medium window here. This stresses the importance of multiple window sizes: For narrow peaks in close succession, a large window may not even exceed threshold when positive and negative slopes cancel out (see blue curve of $(D_{h,t})_t$ for medium window size). As a consequence, one should also note that sensitivity to peaks that are narrower than the smallest used window size is limited.

2.4.2. Choice of window size

Here we use simulations in order to evaluate the performance of the MFA in practical scenarios with finite data sections and small window choices. In particular, due to the asymptotic nature of the theoretical results, we aim at deriving guidelines for the choice of the smallest window that can be used in order to approximate the asymptotic significance level sufficiently closely.

To that end, we simulated data consisting of $T = 1000$ realizations of i.i.d. random variables, distributed according to a Normal $N(\mu, \sigma^2)$ or a Gamma $\Gamma(\text{shape}, \text{rate})$ distribution with different parameters. Figure 4 shows the fraction of simulations per parameter combination in which the null hypothesis was falsely rejected when applying the MFT at asymptotic significance level $\alpha = 0.05$, as a function of the chosen window size. According to the simulations, the smallest window should not be smaller than about $h = 50$ in order to approximately keep the asymptotic significance level.

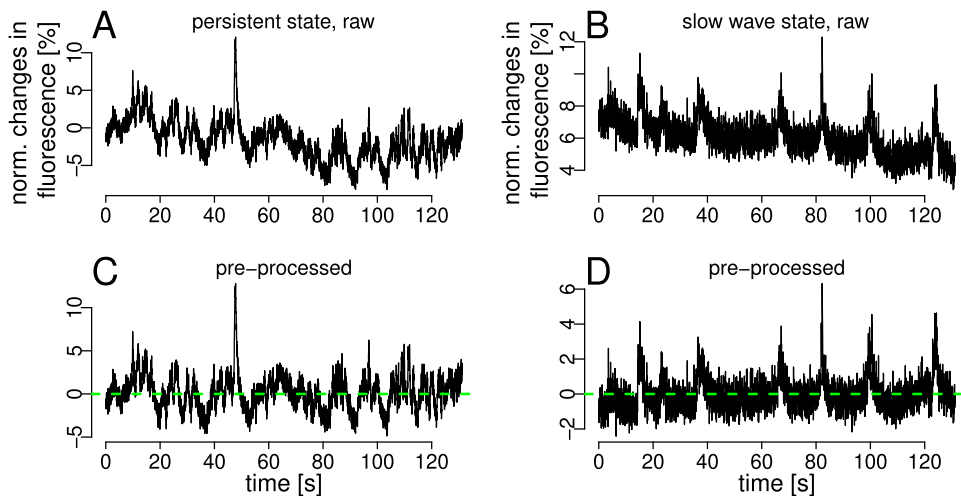


Figure 5. Data pieces of whole-field measurements of the same animal in the same field of view under persistent (A) and slow wave (B) brain state. (C,D) The same time series after application of a filter.

3. Results

Here we apply the proposed MFT and the MFA for peak detection to a neurophysiological data set consisting of two sets of paired whole field ensemble-wide recordings obtained with high-speed two-photon calcium imaging upon expression of the genetically encoded calcium indicator GCaMP6f and a sampling rate of 30.5 Hz. Specifically, the whole field activity of a cortical area in layer II/III of the primary visual cortex of two mice was recorded under two brain states: persistent state, resembling awake-like condition, and slow wave state, resembling slow wave sleep condition. The two states were induced by medetomidine sedation (persistent state) and isoflurane anaesthesia (slow wave state).

While during the slow wave state, whole field ensemble-wide recordings typically showed prominent alternating activity with few clear peaks that are clearly visible, activity during the persistent state showed less prominent fluctuations (cmp. Figure 5). It is of interest to quantify this difference, and also to discriminate periods of up and down states in order to investigate whether these may impact the timing and occurrence of individual neuronal spiking. Therefore, we analysed these two sets of paired time series with respect to differences in their occurrences of estimated peaks. We first applied a simple filtering routine to the raw data set to eliminate a trend in the mean caused by the fading of the calcium image. In consequence, the preprocessed time series consist of the residuals of the original time series from a filtered mean derived with a Gaussian kernel with a large standard deviation of 400 data points (i.e., about 11.3 s, Figures 5(C,D)).

3.1. Analysis of estimated peak occurrences

We then applied the MFA to the four filtered time series in order to estimate peaks, using a window set of $H = \{50, 75, 100, 150\}$ data points (i.e., about 1.5 – 5 s). Across all four time series, about 400 peaks were estimated. In slow wave state, these were mostly matching nicely with visual inspection (see Figure 6(B)), while in persistent state, detected peaks were

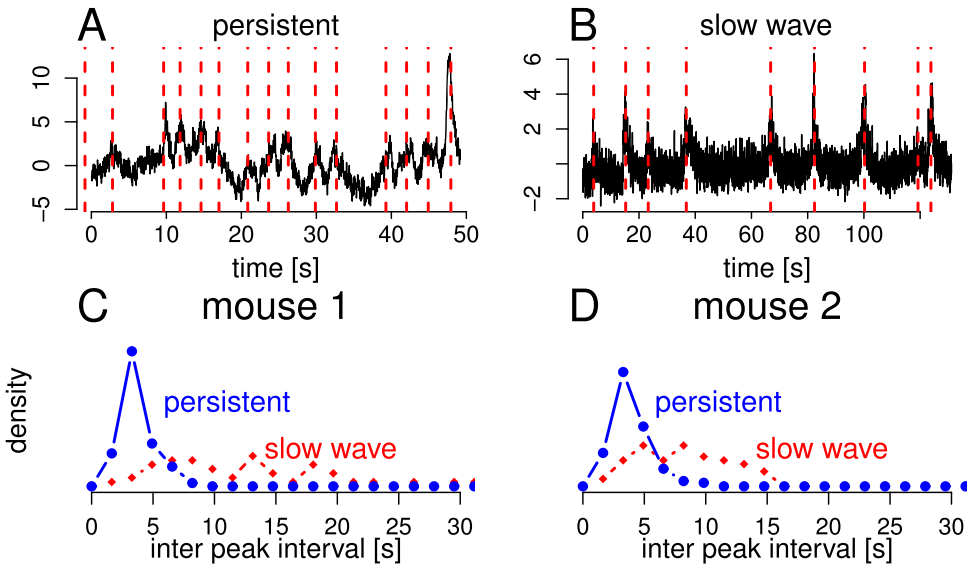


Figure 6. Examples of detected peaks in persistent (A) and slow wave (B) state. (C,D) Distributions of intervals between successively estimated peaks in the two recorded mice, in slow wave (red) and persistent (blue) state.

less pronounced and much more frequent (Figure 6(A)). Serial correlations of successive interval lengths between estimated peaks were not significantly different from zero. The empirical distributions of interval lengths between successive estimated peaks are shown in Figure 6(C,D) for the two mice in both, persistent (P) and slow wave (SW) state. In both animals, the mean inter peak interval (IPI) was longer during slow wave state than during persistent state (mean IPI 14.5 s (SW) vs. 3.6 s (P) in Mouse 1 and 8.2 s (SW) vs. 3.9 s (P) in Mouse 2). Differences between the distributions of IPIs were highly significant in both animals (Wilcoxon rank sum test, $p < 0.001$ for both animals).

3.2. Identification of down states

While one major difference between slow wave and persistent state may be the frequency of peak occurrences, another may be the dynamic of changes between peaks and baseline level, or the activity between the peaks. In order to measure periods of network quiescence, i.e., down states in between peaks, we applied a simple algorithm for the estimation of the shape of the peaks and of the duration between each successive pair of estimated peaks, which was consistent with the theoretical assumptions for the peak function f in Section 2.

Let the ordered times of the estimated peaks be given by $\hat{c}_1, \dots, \hat{c}_m$. Between a pair $(\hat{c}_i, \hat{c}_{i+1})$, we fitted a piecewise linear function

$$g_i(t) = \begin{cases} \beta_{i,0} - \beta_{i,1}(d_{i,1} - t) & t \in [\hat{c}_i, d_{i,1}] \\ \beta_{i,0} & t \in [d_{i,1}, d_{i,2}] \\ \beta_{i,0} + \beta_{i,2}(t - d_{i,2}) & t \in [d_{i,2}, \hat{c}_{i+1}] \end{cases} \quad (13)$$

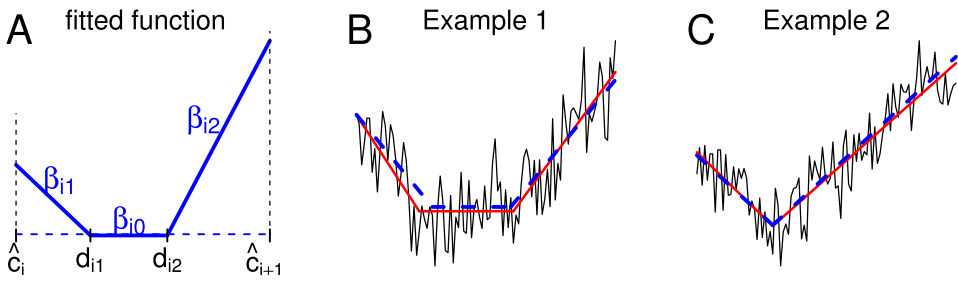


Figure 7. (A) Illustration of down state identification algorithm via least squares fitting of a piecewise linear function with negative slope $\beta_{i,1}$ in $[\hat{c}_i, d_{i,1}]$, with zero slope at baseline level $\beta_{i,0}$ in $[d_{i,1}, d_{i,2}]$ and with positive slope $\beta_{i,2}$ in $[d_{i,2}, \hat{c}_{i+1}]$. The parameters $\beta_{i,0}$, $\beta_{i,1}$, $\beta_{i,2}$ as well as $d_{i,1}$ and $d_{i,2}$ are estimated by least-squares fitting. (B,C) Examples of fitted functions (red) fitted to data (black) simulated with independent and normally distributed errors and a mean given by the blue dashed lines.

to the observed process (see Figure 7(A)). This function describes a period of low activity (denoted here as down state) of height $\beta_{i,0}$ in the interval $[d_{i,1}, d_{i,2}]$ between peaks i and $i + 1$, a linear decrease in peak i of slope $\beta_{i,1} < 0$ up to the start of the down state $d_{i,1}$ and a linear increase at peak $i + 1$ of slope $\beta_{i,2} > 0$. As all peaks showed a certain width and in order to reduce variability of parameter estimates, we assumed that $\hat{c}_i + 10 \leq d_{i,1} \leq d_{i,2} \leq \hat{c}_{i+1} - 10$.

We estimated the parameters $\beta_{i,0}$, $\beta_{i,1}$, $\beta_{i,2}$ and $d_{i,1}$, $d_{i,2}$ for every i by minimizing the sum of squared residuals between the fitted function and the respective section of the data $X_{[\hat{c}_i, \hat{c}_{i+1}]}$, the estimates are denoted with hats.

Using this algorithm, we fitted all almost 400 IPIs estimated in the four empirical time series with the above routine. In most cases, the fit corresponded closely to visual inspection, but about 10% of the IPIs were excluded from further analysis in all cases in which no clear downstate could be identified, i.e., cases in which $\hat{\beta}_{i,2} < 0$ or $\hat{\beta}_{i,1} < 0$ or in which the slope coefficients were not statistically significant from zero on the 5%-level under standard assumptions. As the absolute duration of the down state may potentially depend on the absolute duration of the respective IPI and is therefore likely to show the same differences as the IPI lengths between slow wave and persistent state, we then focused on the relative down state duration, i.e., the relation between the down state duration and the respective IPI duration, measured as

$$\frac{(\hat{d}_{i,2} - \hat{d}_{i,1})}{(\hat{c}_{i+1} - \hat{c}_i)}.$$

As Figure 8 shows, also this relative duration of the down states in relation to the IPI length was considerably larger in the slow wave state than in the persistent state. While the mean relative estimated down state duration was about twice as high in slow wave as in persistent in both animals (0.52 in slow wave and 0.24 in persistent in Mouse 1 and 0.34 in slow wave and 0.17 in persistent in Mouse (2)), the median relative down state duration was even about five times as high in slow wave as in persistent state (0.58 in slow wave and 0.14 in persistent state in Mouse 1 and 0.33 in slow wave and 0.04 in persistent state in Mouse (2)). The differences between distributions were statistically highly significant (Wilcoxon rank sum test, $p < 0.0001$ in both animals).

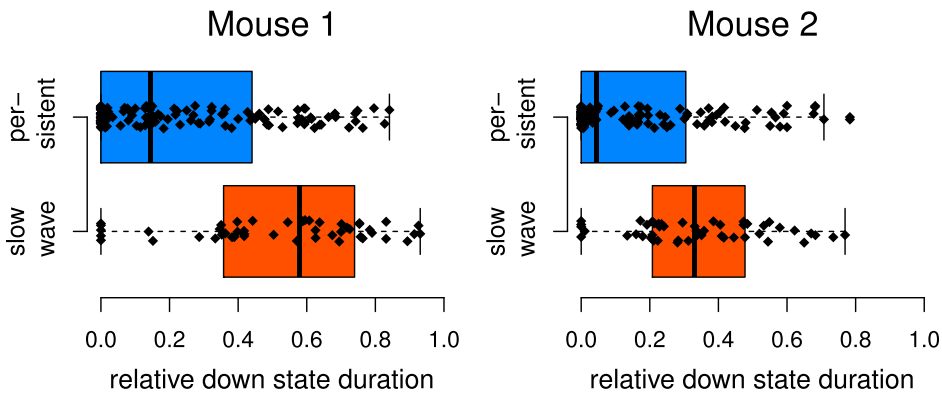


Figure 8. Estimated relative down state durations in the two animals in persistent state (blue) and in slow wave state (red).

In addition, we measured the fraction of IPIs in which the down state was estimated shorter than 10% of the IPI duration, i.e., IPIs with practically no estimated intermediate down state. The fraction of such intervals was as much as about 50% in the persistent state (43% in Mouse 1 and 54% in Mouse 2), but only about 10% in slow wave state in both animals (Fisher's exact test, $p < 0.0001$ for both animals).

These results suggest that the visual impression of the difference between the slow wave state and the persistent state can be quantified by application of the MFA for peak detection together with the algorithm for down state estimation. During slow wave state the signal shows a smaller number of changes between long 'down states' and peaks, or up states, while during persistent state, a higher number of 'peaks' is estimated, while intermediate clear 'down states' cannot always be identified, and if so, they are of much smaller relative duration.

4. Discussion

This article presents an approach that combines a statistical test for peak detection with the estimation of the peak positions in time series. Motivated by empirical observations in neuronal recordings, we aim at simultaneously investigating peaks of different heights and widths. We use a simple model in which discrete and equidistant observations are assumed independent and with an expectation function given by a piecewise linear jagged function f . No parametric assumptions on the errors are required.

The test for the null hypothesis of the absence of peaks in f slides two adjacent moving windows across the time series and compares the scaled difference of estimated slope coefficients within the two windows. The resulting difference process shows systematic positive deviations at peaks, i.e., when the slope decreases instantaneously. By a combination of multiple windows, the procedure allows for an analysis of peaks of different heights and widths, and we use the global maximum of all difference processes as a test statistic.

Statistical significance is evaluated using a limit result about the difference processes. As the respective limit processes do not depend on the parameters of the underlying time series, they can be used to derive the rejection threshold for the test statistic by simulations. After rejection of the null hypothesis, a multiple filter algorithm first estimates peak

positions for every individual window by successively searching for maximizers of the difference processes. In a second step, the results are combined across individual windows. An additional consideration of the growth rates of the parameters could allow for an investigation of the performance of the algorithm, as in [26]. Here, the exact derivation of the limit behaviour of the difference process enables the construction of the asymptotic statistical test.

The chosen test statistic is designed to be sensitive to rapid decreases in the slope, such as peaks or - similar - short plateaus of the mean with a duration of about one window size, or also to smooth hills. Such patterns are likely to be detected analogously with the proposed algorithm. In addition, while the peak process used to describe the neurophysiological data assumes continuity of f , the regression model used for the analysis does not require this assumption and would therefore provide analogous results also without continuity. In addition, the test statistic is designed to be robust against changes in the variance.

The asymptotic test procedure requires that the smallest window should not be smaller than about 50. However, the null hypothesis of i.i.d. random variables can also be tested using simple bootstrap or permutation procedures for the derivation of the rejection threshold, which enables the application of smaller windows as well. This will, however, be considerably more time consuming. In addition, the proposed asymptotic procedure could also be extended to allow for potential serial correlations in the underlying time series. Assuming ergodic error sequences preserves, first, functional central limit theory as in Lemma A.1, and second, functional strong laws of large numbers as in Lemma A.2. Under such assumptions, serial correlations of the random variables need to be estimated and incorporated in the denominator of $D_{h,t}$. In the present, local method, such parameter estimation can only capture the correlation in small time lags, compare also [27]. While in theory, larger time lags can successively be captured as the windows are increased, the window sizes are necessarily fixed in practice, which limits the reliable estimation of serial correlations of larger time lags.

The proposed procedure is implemented in the R-package MFT ([24] available on CRAN), including both the asymptotic and the bootstrap method for derivation of the rejection threshold. In addition, the method can be used analogously in a two-sided manner for the detection of both, positive and negative peaks, which is also implemented.

One major advantage of the proposed method is the simultaneous use of multiple window sizes. In the present data set for example, oscillatory behaviour differs between the different brain states. This suggests that with the application of only one window, comparability would be limited because a window that is suitable for one time series might not be for the other, and vice versa. Here, the use of the same set of multiple windows in all data sets respects peaks of different shape and allows for comparability of the results.

In the statistical analysis of neuronal ensemble activity, the proposed procedure could identify peaks that are visible to the eye. We also combined it with a simple procedure for the detection of intermediate down states that was compatible with the model assumptions on the underlying function f . The results suggest specific differences between different brain states. During slow wave state the signal shows a smaller number of changes between long down states and peaks as compared to the persistent state. Also, during persistent state, a higher number of peaks is estimated, while intermediate clear down states cannot always be identified, and if so, they are of much smaller relative duration. In a further step, this standardized detection of the ensemble-wide functional state could be used to investigate

how local up and down states impact the firing activity of individual neurons. In summary, these results suggest that the proposed method for peak detection can be a useful tool in the analysis of time series showing a variability of different peak phenomena.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Priority Program 1665 of the Deutsche Forschungsgemeinschaft [grant number SCHN 1370/02-1] to MM and GS, grant number STR 1321/8-1 to AS, www.dfg.de), by the CRC 1193 of the Deutsche Forschungsgemeinschaft (to AS) and by the LOEWE Schwerpunkt CMMS – Multiscale Modelling in the Life Sciences (to GS).

Data availability statement

The neurophysiological data set of paired whole field ensemble-wide recordings obtained with high-speed two-photon calcium imaging upon expression of the genetically encoded calcium indicator GCaMP6f can be found in the supplemental material.

ORCID

Michael Messer  <http://orcid.org/0000-0002-7068-4204>

Albrecht Stroh  <http://orcid.org/0000-0001-9410-4086>

References

- [1] Steriade M, Timofeev I, Grenier F. Natural waking and sleep states: a view from inside neocortical neurons. *J Neurophysiol.* **May 2001**;85:1969–1985.
- [2] Sanchez-Vives MV, McCormick DA. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nat Neurosci.* **Oct 2000**;3:1027–1034.
- [3] Stroh A, Adelsberger H, Groh A, et al. Making waves: initiation and propagation of corticothalamic Ca²⁺ waves in vivo. *Neuron.* **2013**;6(77):1136–50.
- [4] Schwalm M, Schmid F, Wachsmuth L, Cortex-wide bold fMRI activity reflects locally-recorded slow oscillation-associated calcium waves. *Elife.* **2017**;6:e27602.
- [5] Pascual-Leone A, Walsh V. Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science.* **Apr 2001**;292:510–512.
- [6] Quandt RE. Tests of the hypothesis that a linear regression system obeys two separate regimes. *J Amer Statist Assoc.* **1960**;55:324–330.
- [7] Hinkley D. Inference about the intersection in two-phase regression. *Biometrika.* **1969**;56:12.
- [8] Brown RL, Durbin J, Evans JM. Techniques for testing the constancy of regression relationships over time. *J R Stat Soc. Ser B (Methodological).* **1975**;37(2):149–192.
- [9] MacNeill IB. Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *Annal Stat.* **1987**;6(2):422–433.
- [10] Hušková M, Picek J. Bootstrap in detection of changes in linear regression. *Sankhy: Indian J Stat.* **2005**;67:200–226.
- [11] Bai J, Perron P. Estimating and testing linear models with multiple structural changes. *Econometrica.* **1998**;66(1):47–78.
- [12] Bischoff W, Miller F. Asymptotically optimal tests and optimal designs for testing the mean in regression models with applications to change-point problems. *Ann Inst Stat Math.* **2000**;52(4):658–679.

- [13] Aue A, Horváth L, Hušková M. Segmenting mean-nonstationary time series via trending regressions. *J Econom.* **2012**;168:367–381.
- [14] Cörgös M, Horváth L, Limit theorems in change-point analysis. Chichester: John Wiley & Sons, Ltd.; 1997. (Wiley Series in Probability and Statistics), With a foreword by David Kendall.
- [15] Bramante R, Facchinetti S, Zappa D. Online detection of financial time series peaks and troughs: a probability-based approach. *Stat Anal Data Mining: ASA Data Sci J.* **2018**;12:426–433.
- [16] Vijaya G, Kumar V, Verma H. Ann-based qrs-complex analysis of ecg. *J Med Eng Technol.* **1998**;22(4):160–167.
- [17] Coast D, Stern R, Cano G, et al. An approach to cardiac arrhythmia analysis using hidden Markov models. *IEEE Trans Biomed Eng.* **1990**;37:826–836.
- [18] Mallat S, Hwang W. Singularity detection and processing with wavelets. *IEEE Trans Inform Theory.* **1992**;38:617–643.
- [19] Köhler B, Hennig C, Orglmeister R. The principles of software QRS detection. *IEEE Eng Med Biol Mag.* **2002**;21(1):42–57.
- [20] Seamari Y, Narvaez JA, Vico FJ, et al. Robust off- and online separation of intracellularly recorded up and down cortical states. *PLoS ONE.* **2007**;2(9):e888.
- [21] Schwartzman A, Gavrilov Y, Adler RJ. Multiple testing of local maxima for detection of peaks in 1d. *Annal Stat.* **2011**;39(6):3290–3319.
- [22] Bertrand PR, Fhima M, Guillin A. Off-line detection of multiple change points by the filtered derivative with p-value method. *Seq Anal.* **2011**;30(2):172–207.
- [23] Messer M, Kirchner M, Schiemann J, et al. A multiple filter test for change point detection in renewal processes with varying variance. *Ann Appl Stat.* **2014**;8(4):2027–2067.
- [24] Messer M, Albert S, Plomer S, et al. *MFT: The Multiple Filter Test for Change Point Detection*, 2019. R package version 2.0.
- [25] Niu YS, Zhang H. The screening and ranking algorithm to detect DNA copy number variations. *Ann. Appl. Stat.* **2012**;6(3):1306–1326.
- [26] Fryzlewicz P. Wild binary segmentation for multiple change-point-detection. *Ann Stat.* **2014**;42(6):2243–2281.
- [27] Messer M, Costa K, Roeper J, et al. Multi-scale detection of rate changes in spike trains with weak dependencies. *J Comput Neurosci.* **2017**;42(2):187–201.
- [28] Ethier SN, Kurtz TG, Markov processes. New York (NY): John Wiley & Sons, Inc.; 1986. (Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics), Characterization and convergence

Appendix

We give all proofs and auxiliary results. Recall that for an error ε_k we abbreviate its scaled version by $\varepsilon_k^* := \varepsilon_k / (\text{Var}(\varepsilon_k))^{1/2}$.

Proof of Lemma 2.2.: We rewrite the sum that occurs for $\hat{\beta}_j$ in (4) as

$$\sum_{k \in I_j} (k - \bar{I}_j) Y_k = \sum_{k \in I_j} (k - \bar{I}_j) (\eta_j + \beta_j k + \sigma \varepsilon_k^*) = \beta_j \sum_{k \in I_j} (k - \bar{I}_j)^2 + \sigma \sum_{k \in I_j} (k - \bar{I}_j) \varepsilon_k^*,$$

where we used the formulation of the model $Y_k = \eta_j + \beta_j k + \sigma \varepsilon_k$ and that $\sum_{k \in I_j} (k - \bar{I}_j) = 0$, see the first identity in (6). From the second identity in (6) we obtain by definition of $\hat{\beta}_j$ (cmp. equation (8))

$$\hat{\beta}_j = \frac{12}{nh[(nh)^2 - 1]} \sum_{k \in I_j} (k - \bar{I}_j) Y_k = \beta_j + \sigma_j \frac{12}{nh[(nh)^2 - 1]} \sum_{k \in I_j} (k - \bar{I}_j) \varepsilon_k^*.$$

This yields the expectation and variance as stated since the $(\varepsilon_k^*)_k$ are i.i.d. with $\mathbb{E}[\varepsilon_k^*] = 0$ and unit variance for all k . For the asymptotic normality we rewrite the last expression as

$$\frac{\hat{\beta}_j - \beta_j}{[\text{Var}(\hat{\beta}_j)]^{1/2}} = \sum_{k \in I_j} \left[\frac{12}{nh[(nh)^2 - 1]} \right]^{1/2} (k - \bar{I}_j) \varepsilon_k^*.$$

For each n , the summands on the latter display are independent with expectation zero while their variance adds up to unity. Over different n the summands constitute a triangular scheme. Since $|k - \bar{I}_j| \leq nh/2$, we can bound each summand by $c\varepsilon_k^*/\sqrt{n}$ for some positive constant c and thus the Lindeberg-Feller central limit theorem applies. \blacksquare

Proof of Lemma 2.3.: The idea of the proof is to use the strong law of large numbers and first deduce that $|\hat{\beta}_j - \beta_j| = o(1/n)$ almost surely, and then we use the definition of $\hat{\eta}_j$ and $\hat{\sigma}_j^2$. We recall the estimated slope $\hat{\beta}_j$ from (8) and mention that $|k - \bar{I}_j| \leq nh$ for all $k \in I_j$, such that we can bound

$$|\hat{\beta}_j - \beta_j| \leq \frac{\kappa}{n} \left(\frac{1}{nh} \sum_{k \in I_j} \varepsilon_k^* \right).$$

for some positive constant κ (which does not depend on t). Since the $(\varepsilon_k^*)_k$ are centred and integrable, the mean in the brackets vanishes almost surely according to the strong law of large numbers, such that almost surely as $n \rightarrow \infty$

$$|\hat{\beta}_j - \beta_j| = o\left(\frac{1}{n}\right).$$

For the estimator of the intercept we recall (5) where we find that $\hat{\eta}_j = \bar{Y}_j - \hat{\beta}_j \bar{I}_j$. Further, by definition of the model we obtain $\bar{Y}_j = \eta_j + \beta_j \bar{I}_j + \sigma_j(1/(nh)) \sum_{k \in I_j} \varepsilon_k^*$. Substituting the latter yields almost surely as $n \rightarrow \infty$

$$|\hat{\eta}_j - \eta_j| \leq |\hat{\beta}_j - \beta_j| \bar{I}_j + \sigma_j \frac{1}{nh} \sum_{k \in I_j} \varepsilon_k^* \longrightarrow 0,$$

because $\bar{I}_j = O(n)$ and $|\hat{\beta}_j - \beta_j| = o(1/n)$ almost surely, and also because the second summand tends to zero almost surely as stated before. For the estimator of σ_j^2 we also recall (7) and find $\hat{\sigma}_j^2 = (1/(nh - 2)) \sum_{k \in I_j} (Y_k - (\hat{\eta}_j + \hat{\beta}_j k))^2$. Substituting $Y_k = \eta_j + \beta_j k + \sigma_j \varepsilon_k^*$ yields

$$\hat{\sigma}_j^2 = \frac{1}{nh - 2} \sum_{k \in I_j} [(\eta_j - \hat{\eta}_j) + (\beta_j - \hat{\beta}_j)k + \sigma_j \varepsilon_k^*]^2. \quad (\text{A1})$$

Solving the square yields six summands of which all but the last vanish almost surely. To see this we write them as first $[1/(nh - 2)]nh(\eta_j - \hat{\eta}_j)^2$, second $2(\eta_j - \hat{\eta}_j)(\beta_j - \hat{\beta}_j)[1/(nh - 2)] \sum_{k \in I_j} k$, third $(\beta_j - \hat{\beta}_j)^2[1/(nh - 2)] \sum_{k \in I_j} k^2$, fourth $2(\eta_j - \hat{\eta}_j)\sigma_j[1/(nh - 2)] \sum_{k \in I_j} \varepsilon_k^*$ and fifth $2(\beta_j - \hat{\beta}_j)[1/(nh - 2)] \sum_{k \in I_j} k\varepsilon_k^*$. Then, the a.s. convergences to zero follow directly from the convergence of the mean of the $(\varepsilon_k^*)_k$, the convergence of $\hat{\eta}_j$ and the rate of $\hat{\beta}_j$ as stated before. For that we also mention that $[1/(nh - 2)] \sum_{k \in I_j} k = O(n)$, $[1/(nh - 2)] \sum_{k \in I_j} k^2 = O(n^2)$ and also that $\sum_{k \in I_j} \varepsilon_k^* = o(n)$ a.s. and $\sum_{k \in I_j} k\varepsilon_k^* = o(n^2)$ a.s.

For the last summand it holds almost surely as $n \rightarrow \infty$ that

$$\sigma_j^2 \frac{1}{nh - 2} \sum_{k \in I_j} \varepsilon_k^{*2} \longrightarrow \sigma_j^2,$$

according to the strong law of large numbers and because the $(\varepsilon_k^*)_k$ are centred with unit variance. \blacksquare

Proof of Lemma 2.4: The Lemma is a consequence of Lemmas 2.2 and 2.3 which state the asymptotic normality of the slopes within the regression models and the strong consistency of the estimators. Fitting the regression models to the process $(X_t)_t$ we obtain from Lemma 2.2 for $j \in \{1, 2\}$ as $n \rightarrow \infty$

$$\left(\frac{12\sigma_j^2}{nh[(nh)^2 - 1]} \right)^{-1/2} \cdot (\hat{\beta}_j - \beta_j) \xrightarrow{d} N(0, 1),$$

and the term in the left brackets equals the variance of $\hat{\beta}_j$. In Lemma 2.3 we show that $\hat{\sigma}_j^2 \rightarrow \sigma_j^2$ almost surely as $n \rightarrow \infty$. Thus, by definition of $\hat{s}_{h,t}^{(n)}$ in (10), we obtain by Slutsky's Lemma that $[(\hat{\beta}_\ell - \hat{\beta}_r) - (\beta_\ell - \beta_r)]/\hat{s}_{h,t}^{(n)} \xrightarrow{d} N(0, 1)$, by also using that the independence of $\hat{\beta}_\ell$ and $\hat{\beta}_r$ implies their joint convergence. The latter convergence is the statement. \square

The proof of Proposition 2.5 requires the following two auxiliary lemmas. Let $(D_{\mathbb{R}^2}[0, \infty), d_{SK})$ denote the \mathbb{R}^2 -valued càdlàg-functions on $[0, \infty)$ endowed with Skorokhod topology. \blacksquare

Lemma A.1: Let $\varepsilon_1^*, \varepsilon_2^*, \dots$ be i.i.d. random variables with zero mean and unit variance. In $(D_{\mathbb{R}^2}[0, \infty), d_{SK})$ it holds as $n \rightarrow \infty$

$$\left(\frac{1}{n^3} \right)^{1/2} \cdot \left(\sum_{k=1}^{\lfloor nt \rfloor} k \varepsilon_k^* \right)_{t \geq 0} \xrightarrow{d} GP(0, \Gamma), \quad (\text{A2})$$

with covariance function Γ given in (11).

Proof: First note that the second component states Donker's theorem. We state the covariance matrix of the left-hand side and make use of a bivariate martingale central limit theorem. Let $s \leq t$. We consider three cases. First, we calculate the temporal covariance of the first component

$$\begin{aligned} \mathbb{C}ov \left(\sum_{k=1}^{\lfloor nt \rfloor} k \varepsilon_k^*, \sum_{k=1}^{\lfloor ns \rfloor} k \varepsilon_k^* \right) &= \mathbb{C}ov \left(\sum_{k=1}^{\lfloor ns \rfloor} k \varepsilon_k^* + \sum_{k=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} k \varepsilon_k^*, \sum_{k=1}^{\lfloor ns \rfloor} k \varepsilon_k^* \right) \\ &= \sum_{k=1}^{\lfloor ns \rfloor} k^2 = \frac{s^3}{3} \cdot n^3 \cdot (1 + o(1)), \end{aligned}$$

as $n \rightarrow \infty$. Second, we calculate the temporal covariance of the second component.

$$\mathbb{C}ov \left(\sum_{k=1}^{\lfloor nt \rfloor} n \varepsilon_k^*, \sum_{k=1}^{\lfloor ns \rfloor} n \varepsilon_k^* \right) = \sum_{k=1}^{\lfloor ns \rfloor} n^2 = s \cdot n^3 \cdot (1 + o(1)).$$

Third, we consider the first component at time t and the second component at s

$$\mathbb{C}ov \left(\sum_{k=1}^{\lfloor nt \rfloor} k \varepsilon_k^*, \sum_{k=1}^{\lfloor ns \rfloor} n \varepsilon_k^* \right) = \mathbb{C}ov \left(\sum_{k=1}^{\lfloor ns \rfloor} k \varepsilon_k^*, \sum_{k=1}^{\lfloor ns \rfloor} n \varepsilon_k^* \right) = \sum_{k=1}^{\lfloor ns \rfloor} kn = \frac{s^2}{2} \cdot n^3 \cdot (1 + o(1)),$$

and analogously for the components interchanged. For $s > t$ replace s by t . This yields Γ as stated while the factors n^3 cancel due to the scaling in (A2). The summands are independent with zero expectation and the left-hand side constitutes a bivariate martingale, and convergence follows from central limit theory for martingales, see e.g., [28, Theorem 7.1.4.]. \blacksquare

Lemma A.2: Within the regression models it holds for $j \in \{\ell, r\}$ in $(D_{\mathbb{R}}[h, T - h], d_{SK})$ as $n \rightarrow \infty$ in probability $(\hat{\beta}_j)_t \rightarrow (\beta_j)_t, (\hat{\eta}_j)_t \rightarrow (\eta_j)_t$ and $(\hat{\sigma}_j^2)_t \rightarrow (\sigma_j^2)_t$.

Proof: The idea is to state functional laws of large numbers for the errors $(\varepsilon_k^*)_k$ which are used to extend the marginal results from Lemma 2.3. For $j \in \{\ell, r\}$ it holds in $(\mathcal{D}_{\mathbb{R}}[h, T-h], d_{SK})$ as $n \rightarrow \infty$

$$\begin{aligned} \left[\frac{1}{nh} \sum_{k \in I_j} \frac{k - \bar{I}_j}{nh} \varepsilon_k^* \right]_t &\xrightarrow{\mathbb{P}} (0)_t, & \left[\frac{1}{nh} \sum_{k \in I_j} \frac{k}{n} \varepsilon_k^* \right]_t &\xrightarrow{\mathbb{P}} (0)_t, \\ \left[\frac{1}{nh} \sum_{k \in I_j} \varepsilon_k^* \right]_t &\xrightarrow{\mathbb{P}} (0)_t, & \left[\frac{1}{nh} \sum_{k \in I_j} \varepsilon_k^{*2} \right]_t &\xrightarrow{\mathbb{P}} (1)_t. \end{aligned} \quad (\text{A3})$$

We comment on the first convergence for $j = \ell$. From (A2) we obtain in $(\mathcal{D}_{\mathbb{R}^2}[0, \infty), d_{SK})$

$$\left[\frac{1}{n} \left(\sum_{k=1}^{\lfloor nt \rfloor} \frac{k}{n} \varepsilon_k^*, \sum_{k=1}^{\lfloor nt \rfloor} \varepsilon_k^* \right) \right]_t \xrightarrow{\mathbb{P}} (0)_t,$$

on which we apply the continuous map $\varphi : (\mathcal{D}_{\mathbb{R}^2}[0, \infty), d_{SK}) \rightarrow (\mathcal{D}_{\mathbb{R}}[h, T-h], d_{SK})$ via

$$\begin{pmatrix} f_1(t) \\ f_2(t) \end{pmatrix}_{t \geq 0} \xrightarrow{\varphi} \left([f_1(t) - f_1(t-h)] - [\lfloor t-h \rfloor + h/2][f_2(t) - f_2(t-h)] \right)_{t \geq 0}.$$

which yields

$$\left(\frac{1}{n} \sum_{k \in I_\ell} \left[\frac{k}{n} - (\lfloor t-h \rfloor + h/2) \right] \varepsilon_k^* \right)_t \xrightarrow{\mathbb{P}} (0)_t,$$

and this implies the first statement. The other convergences follow similarly. Now we run through the proof of Lemma 2.3 replacing the marginal estimators for fixed t with their functional counterparts and making use of (A3). \blacksquare

Proof of Proposition 2.5.: Recall the formulation of $\hat{\beta}_j$ from (8) given as

$$\hat{\beta}_j = \beta_j + \sigma_j \frac{12}{nh[(nh)^2 - 1]} \sum_{k \in I_j} (k - \bar{I}_j) \varepsilon_k^*,$$

for $j \in \{\ell, r\}$. The set of indices I_j and their means \bar{I}_j are given in (2) and (3). Since $f = 0$, we find $\beta_\ell = \beta_r$ and $\sigma_\ell = \sigma_r =: \sigma$. It follows from Lemma 2.2 that the true variance of the difference of estimated slopes writes as

$$\text{Var}(\hat{\beta}_\ell - \hat{\beta}_r) = \frac{12}{nh[(nh)^2 - 1]} (\sigma_\ell^2 + \sigma_r^2) = \frac{12}{(nh)^3} \cdot 2\sigma^2 \cdot (1 + o(1)).$$

We introduce an auxiliary process $(\tilde{D}_{h,t}^{(n)})_t$ in $(\mathcal{D}_{\mathbb{R}}[h, T-h], d_{SK})$ via

$$\tilde{D}_{h,t} := \left(\frac{12}{(nh)^3} 2\sigma^2 \right)^{-1/2} (\hat{\beta}_\ell - \hat{\beta}_r) = \left(\frac{12}{2(nh)^3} \right)^{1/2} \left[\sum_{k \in I_\ell} (k - \bar{I}_\ell) \varepsilon_k^* - \sum_{k \in I_r} (k - \bar{I}_r) \varepsilon_k^* \right]. \quad (\text{A4})$$

This process describes the comparison of slopes, but it uses the true order of scaling as compared to $D_{h,t}$ in which the scaling is estimated. The purpose is now to show that in $(\mathcal{D}_{\mathbb{R}}[h, T-h], d_{SK})$ we obtain for $n \rightarrow \infty$

$$(\tilde{D}_{h,t}^{(n)})_t \xrightarrow{d} (L_{h,t})_t. \quad (\text{A5})$$

This is derived in two steps. In Lemma A.1 we stated process convergence of joint rescaled sums to a Gaussian process limit, i.e., in $(\mathcal{D}_{\mathbb{R}^2}[0, \infty), d_{SK})$ it holds as $n \rightarrow \infty$ that

$$\left(\frac{1}{n^3}\right)^{1/2} \cdot \left(\sum_{k=1}^{\lfloor nt \rfloor} k \varepsilon_k^* \right)_{t \geq 0} \xrightarrow{d} \left(\mathcal{G}_1(t) \right)_{t \geq 0} = GP(0, \Gamma), \quad (\text{A6})$$

with covariance function Γ given in (11). In the second step, we then apply a continuous function φ that maps the left-hand side of (A6) to $(\tilde{D}_{h,t})_t$ (up to a negligible factor) and the right-hand side to $(L_{h,t})_t$, which results in (A5) by continuous mapping. Define $\varphi : (\mathcal{D}_{\mathbb{R}^2}[0, \infty), d_{SK}) \rightarrow (\mathcal{D}_{\mathbb{R}}[h, T-h], d_{SK})$ via

$$\begin{aligned} \left(\frac{f_1(t)}{f_2(t)}\right)_{t \geq 0} &\xrightarrow{\varphi} \left(\left(\frac{6}{h^3}\right)^{1/2} \left[\left([f_1(t) - f_1(t-h)] - \left(t - \frac{h}{2}\right) [f_2(t) - f_2(t-h)] \right) \right. \right. \\ &\quad \left. \left. - \left([f_1(t+h) - f_1(t)] - \left(t + \frac{h}{2}\right) [f_2(t+h) - f_2(t)] \right) \right] \right)_{t \geq 0}. \end{aligned} \quad (\text{A7})$$

A comparison with the definition of the limit process in (12) directly shows, that it is obtained by applying φ to the limit of the rescaled sums in (A6). Now we apply φ to the left hand side of (A6) and obtain

$$\begin{aligned} \varphi \left(\left(\sum_{k=1}^{\lfloor nt \rfloor} k \varepsilon_k^* \right)_{t \geq 0} \right) &= \left(\left(\frac{12}{2(nh)^3} \right)^{1/2} \left[\left(\sum_{k \in I_\ell} k \varepsilon_k^* - \left(t - \frac{h}{2}\right) \sum_{k \in I_\ell} n \varepsilon_k^* \right) \right. \right. \\ &\quad \left. \left. - \left(\sum_{k \in I_r} k \varepsilon_k^* - \left(t + \frac{h}{2}\right) \sum_{k \in I_r} n \varepsilon_k^* \right) \right] \right)_{t \in [h, T-h]} \\ &= \left(\left(\frac{12}{2(nh)^3} \right)^{1/2} \left[\left(\sum_{k \in I_\ell} (k - \bar{I}_\ell^*) \varepsilon_k^* \right) - \left(\sum_{k \in I_r} (k - \bar{I}_r^*) \varepsilon_k^* \right) \right] \right)_{t \in [h, T-h]}. \end{aligned} \quad (\text{A8})$$

For that we note that $\bar{I}_j^* = \bar{I}_j + o(n)$ resulting from $n \lfloor t \rfloor = \lfloor nt \rfloor + o(n)$ and thus, it can be exchanged with \bar{I}_j by Slutsky's Lemma when letting $n \rightarrow \infty$ and considering weak process convergence. But this then equals the process $(\tilde{D}_{h,t}^{(n)})_t$ such that (A5) holds true. Then we can replace the true order of scaling with the estimator $(\hat{s}_{h,t}^{(n)})_t$ by Slutsky's Lemma using functional consistency

$$\left(\left(\frac{12}{(nh)^3} 2\sigma^2 \right)^{-1/2} \cdot \hat{s}_{h,t}^{(n)} \right)_t \xrightarrow{\mathbb{P}} (1)_t, \quad (\text{A9})$$

which follows from the consistency of the estimators $(\hat{\sigma}_j^2)_t$ for $(\sigma^2)_t$ in Lemma A.2. ■

Finally, we conclude that the convergence in Proposition 2.5 extends to joint convergence applying finitely many windows $H = \{h_1, \dots, h_m\}$. For that let $(\times_{i=1}^m \mathcal{D}_{\mathbb{R}}[h_i, T-h_i], d_{SK})$ denote the product space regarding all $(\mathcal{D}_{\mathbb{R}}[h_i, T-h_i], d_{SK})$, $i = 1, \dots, m$, endowed with the product metric.

Corollary A.3: *Let $(X_t)_t \in \mathcal{M}$ with $f = 0$ and let $H = \{h_1, \dots, h_m\}$ be a set of windows. Then it holds in $(\times_{i=1}^m \mathcal{D}_{\mathbb{R}}[h_i, T-h_i], d_{SK})$ as $n \rightarrow \infty$*

$$(D_{h_1,t}^{(n)}, \dots, D_{h_m,t}^{(n)})_t \xrightarrow{d} (L_{h_1,t}, \dots, L_{h_m,t})_t,$$

where each component of the limit process is given in (12) and all components are evaluated on the same Brownian motion.

Proof: We follow the proof of Proposition (2.5). Extending $\varphi =: \varphi_h$ from (A7) to the joint map

$$(\varphi_{h_1}, \dots, \varphi_{h_m}) : (\mathcal{D}_{\mathbb{R}^2}[0, \infty), d_{SK}) \longrightarrow (\times_{i=1}^m \mathcal{D}_{\mathbb{R}}[h_i, T - h_i], d_{SK}),$$

and applying it to (A6) (resp. Lemma A.1) yields via (A8) that in $(\times_{i=1}^m \mathcal{D}_{\mathbb{R}}[h_i, T - h_i], d_{SK})$ as $n \rightarrow \infty$

$$(\tilde{D}_{h_{1,t}}^{(n)}, \dots, \tilde{D}_{h_{m,t}}^{(n)})_t \xrightarrow{d} (L_{h_{1,t}}, \dots, L_{h_{m,t}})_t,$$

which is the extension of (A5). As (A9) directly extends to joint convergence over multiple windows, the assertion follows by Slutsky's lemma. ■