

# **Heterogeneous Multiscale Change-Point Inference and its Application to Ion Channel Recordings**

**Dissertation**

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

„Doctor rerum naturalium“

der Georg-August-Universität Göttingen

im Promotionsprogramm „Mathematical Science“

der Georg-August University School of Science (GAUSS)

vorgelegt von

**Florian Pein**

aus Hatzfeld (Eder)

Göttingen, 2017

## **Betreuungsausschuss:**

Prof. Dr. Axel Munk,  
Institut für Mathematische Stochastik, Universität Göttingen

Jun.-Prof. Dr. Andrea Krajina,  
Institut für Mathematische Stochastik, Universität Göttingen

## **Mitglieder der Prüfungskommission:**

### *Referent*

Prof. Dr. Axel Munk,  
Institut für Mathematische Stochastik, Universität Göttingen

### *Korreferentin*

Jun.-Prof. Dr. Andrea Krajina,  
Institut für Mathematische Stochastik, Universität Göttingen

### *Weitere Mitglieder der Prüfungskommission*

Prof. Dr. Claudia Steinem,  
Institut für Organische und Biomolekulare Chemie, Universität Göttingen

Prof. Dr. David Russell Luke,  
Institut für Numerische und Angewandte Mathematik, Universität Göttingen

Prof. Dr. Valentin Blomer  
Mathematisches Institut, Universität Göttingen

Dr. Michael Habeck  
Institut für Mathematische Stochastik, Universität Göttingen

**Tag der mündlichen Prüfung:** 20. Oktober 2017

# Summary

Ion channel recordings by the patch clamp technique are a major tool to quantify the electrophysiological dynamics of ion channels in the cell membrane, which is for instance important in medicine for the development of new drugs. In this work, we model these recordings as a time series which is equidistantly sampled from the convolution of a piecewise constant signal disturbed by white noise with a lowpass filter. We focus on non-parametric estimation of the underlying signal, but also discuss how to use these estimations to analyze the recordings. Estimating the underlying signal requires to detect multiple change-points in noisy and filtered Gaussian observations. The variance can be constant in time, but also a varying variance is observed in some measurements. Since this change-point regression problem is very difficult, we start with independent Gaussian observations but with heterogeneous noise. Such a model is of its own interest and has further applications for instance in genetics.

For this model, we propose the **heterogeneous simultaneous multiscale change-point estimator**, H-SMUCE. It estimates the piecewise constant function by minimizing the number of change-points over the acceptance region of a multiscale test which locally adapts to changes in the variance. The multiscale test is a combination of local likelihood ratio tests which are properly calibrated by scale dependent critical values in order to keep a global nominal level  $\alpha$ , even for finite samples.

We show that H-SMUCE controls over- and underestimation of the number of change-points at a given probability for finitely many observations. To this end, new deviation bounds for  $F$ -type statistics are derived. We also bound the implicitly defined critical values. By combining these bounds, we obtain simultaneous confidence intervals for the change-point locations and a confidence band for the whole signal. Moreover, it allows us to show that H-SMUCE achieves the optimal detection rate and estimates the number of change-points consistently for vanishing signals, even when the number of change-points is unbounded. The only extra assumption we have to suppose is that the length of the constant segments does not vanished too fast. We compare the performance of H-SMUCE with several state of the art methods in simulations and show how it can be computed efficiently by a pruned dynamic program. An R-package is provided.

In a second step we combine these multiscale regression techniques with deconvolution to obtain non-parametric estimators for the ion channel recordings. Truncating the filter kernel and pre-estimating the function values on longer constant segments enable us to perform the deconvolution locally which allows fast computation. Simulations and real data applications confirm that the proposed segmentation methods, JULES and JILTAD,

estimate the underlying signal very accurately, even when events occur on small temporal scales, where the smoothing effect of the filter hinders estimation by common methods. Moreover, JILTAD shows still good results when the noise is heterogeneous, a situation for which previously no non-parametric estimation method existed. Also these methods are implemented in *R*.

The usage of these methods is demonstrated in a biochemical study against the context of multidrug-resistant bacteria. We showed statistically significant differences for the interaction of the antibiotic ampicillin with the wild type and with the mutant G103K of the outer membrane channel PorB. These results improves the understanding of potential sources for bacterial resistance and might help to develop new drugs against it to alleviate the severe consequences of multidrug-resistant bacteria.

# Acknowledgements

First and foremost, I would like to express my deep gratitude to my supervisor Axel Munk for giving me this challenging and interesting topic that allowed me insights in non-parametric statistic, asymptotic theory, practice-oriented statistical analysis of real data and software development. I am particularly thankful for his guidance through this broad range of topics. His stimulating ideas and great intuition were always a big help.

I am very grateful to Andrea Krajina for being my second advisor, in particular for accepting to review my dissertation.

Moreover, I would like to thank Claudia Steinem, David Russell Luke, Valentin Blomer and Michael Habeck for being in my committee.

I am very, very and even more obliged to Annika Bartsch, Ole Schütte, Claudia Steinem, Niels Denkert and Michael Meinecke for sharing their measured ion channel recordings, for introducing me into this topic, for all the answered questions and for the productive and pleasant collaboration.

I am deeply indebted to Miguel Del Alamo for proofreading of this thesis, his helpful suggestions and many fruitful discussions.

I owe many thanks to Manuel Diehn for the endless hours of analyzing ion channel recordings together and for the deep insights into hidden Markov models.

Special thanks go to Timo Aspelmeier, for sharing its programming expertise, vital discussions and proofreading my code and documentation.

I am also grateful to Ivo Siekmann for giving me a different perspective on the analysis of ion channel recordings, his fruitful ideas and proofreading my documentation.

Special thanks should be given to Hannes Sieling for mentoring me during my bachelor thesis, master thesis and beginning of my Ph.D. time and for introducing me into the change-point theory.

I wish to acknowledge the help of Inder Tecuapetla-Gómez and Frank Werner who assisted me several times and shared their expertise in time series analysis and inverse problems.

I gratefully acknowledge support from the DGF grant SFB 803.

I am particularly grateful to my colleagues for the enjoyable environment, for always being supportive and for many tasty cakes. I would like to offer my special thanks to my former and current office co-workers and to all table soccer enthusiasts.

I would like to express my heartfelt thanks to my parents and friends for supporting me all the time, for teaching me to become a better person and for just being the greatest persons in my life.

THANK YOU!



# Contents

<b>1. Introduction</b>	<b>1</b>
1.1. Multiscale change-point estimation . . . . .	3
1.2. Heterogeneous change-point estimation . . . . .	4
1.3. Theoretical results for H-SMUCE . . . . .	6
1.4. H-SMUCE in action . . . . .	7
1.5. Multiscale estimation for ion channel recordings . . . . .	7
1.6. Implementation and software . . . . .	10
1.7. Application to PorB recordings . . . . .	10
<b>2. Heterogeneous Multiscale Change-Point Inference for Independent Observations</b>	<b>11</b>
2.1. Heterogeneous change-point model . . . . .	12
2.2. Methodology . . . . .	14
2.2.1. Scale dependent critical values . . . . .	15
2.3. Theory . . . . .	17
2.3.1. Finite bounds for over- and underestimation . . . . .	17
2.3.2. Confidence sets . . . . .	20
2.3.3. Asymptotic detection rates for vanishing signals . . . . .	21
2.3.4. Choice of the tuning parameters . . . . .	23
2.4. Simulations . . . . .	24
2.4.1. Simulation results . . . . .	25
2.4.2. Prior information on scales . . . . .	29
2.4.3. Robustness against model violations . . . . .	29
<b>3. Ion Channel Recordings</b>	<b>32</b>
3.1. Model . . . . .	32
3.1.1. Homogeneous ion channel model . . . . .	32
3.1.2. Heterogeneous ion channel model . . . . .	36
3.2. Requirements on the estimation . . . . .	38
3.3. Multiscale regression for filtered observations and its problems . . . . .	43
3.4. J-SMURF . . . . .	43
3.4.1. J-SMURF for homogeneous noise . . . . .	44
3.4.2. J-SMURF for heterogeneous noise . . . . .	45
3.5. JULES . . . . .	46
3.5.1. Piecewise constant approximation . . . . .	46

3.5.2.	Postfiltering . . . . .	47
3.5.3.	Heterogeneous noise . . . . .	48
3.6.	Local deconvolution . . . . .	48
3.6.1.	Long segments . . . . .	48
3.6.2.	Maximum likelihood estimation . . . . .	49
3.6.3.	Regularization . . . . .	49
3.6.4.	Iterative grid search . . . . .	50
3.7.	JILTAD . . . . .	51
3.7.1.	J-SMURF step . . . . .	52
3.7.2.	Local testing . . . . .	52
3.7.3.	Local test statistic for homogeneous noise . . . . .	54
3.7.4.	Local test statistic for heterogeneous noise . . . . .	54
3.7.5.	Multiple dependent rejections . . . . .	57
3.7.6.	Critical values . . . . .	57
3.7.7.	Local deconvolution . . . . .	58
3.8.	Analysis of the gating dynamics . . . . .	58
3.9.	Simulations . . . . .	64
3.9.1.	Data generation . . . . .	65
3.9.2.	Parameter choices . . . . .	65
3.9.3.	Isolated peak with homogeneous noise . . . . .	66
3.9.4.	Isolated peak with heterogeneous noise . . . . .	69
3.9.5.	Separation of two consecutive peaks . . . . .	72
3.9.6.	Hidden Markov model . . . . .	75
3.9.7.	Hidden Markov model with heterogeneous noise . . . . .	77
3.10.	Analysis of Gramicidin A recordings . . . . .	80
3.10.1.	Estimation of the underlying signals . . . . .	80
3.10.2.	Analysis of flickering dynamics . . . . .	81
<b>4.</b>	<b>Computation and software</b>	<b>88</b>
4.1.	Computation of the multiscale regression estimator . . . . .	88
4.1.1.	Speed up for small interval sets . . . . .	89
4.2.	Computation of the vector of critical values . . . . .	90
4.2.1.	Scale balancing by penalization . . . . .	90
4.2.2.	Scale balancing by weights . . . . .	91
4.2.3.	Storing the results of the Monte-Carlo simulations . . . . .	92
4.3.	The R package stepR . . . . .	93
4.4.	The R package clampSeg . . . . .	94
4.5.	Computation time . . . . .	94
<b>5.</b>	<b>Application</b>	<b>100</b>
5.1.	Results for the recordings with ampicillin . . . . .	101

5.2. Analysis of PorB recordings with heterogeneous noise . . . . .	103
<b>6. Conclusion and outlook</b>	<b>107</b>
6.1. Conclusion . . . . .	107
6.2. Outlook . . . . .	108
6.2.1. Varying voltage . . . . .	108
6.2.2. Fixed levels . . . . .	111
6.2.3. Parallel programming . . . . .	112
<b>A. Appendix</b>	<b>113</b>
A.1. Proof of Lemma 1 . . . . .	113
A.2. Proof of Lemma 3 . . . . .	113
A.3. Exponential deviation bounds . . . . .	115
A.4. Proofs of Section 2.3 . . . . .	118
A.5. Proofs of Section 3 . . . . .	127
A.6. Proofs of Section 4 . . . . .	129
<b>Bibliography</b>	<b>137</b>



# 1. Introduction

Ion channels are pore-forming proteins in the cell membrane that allow ions to pass the membrane, which itself is impermeable to ions, along the electrochemical gradient (Alberts et al., 2007). In this function they are mandatory for several vital processes like excitation of muscle and nerve cells, building up a resting membrane potential, energy conversion and regulation of the osmotic activity of cells (Hille, 2001). The amount of ions that can pass an ion channel is not constant in time. Often caused by external stimuli such as voltage, ligand binding or mechanical stress (Chung et al., 2007; Purves et al., 2008), the pores of an ion channel can open and close, called *gating*. Also the passage of larger proteins can block temporarily the ion pathway (Raj Singh et al., 2012). Understanding these processes is crucial, e.g., for developing drugs against several diseases such as cancer or epilepsy (Kass, 2005; Overington et al., 2006; Kim, 2014). In this work, another example is given in Section 5, where we present our joint analysis with the Steinem lab (Institute of Organic and Biomolecular Chemistry, University of Göttingen) of the interaction of the antibiotic ampicillin with the outermembrane porin PorB (Bartsch et al., 2017). An example of recorded observations is shown in Figure 1.1.

A major tool for a quantitative analysis of the gating dynamics is the *patch clamp* technique (Neher and Sakmann, 1976). It allows to measure the conductance, the recorded current divided by the applied voltage, of a single ion channel in time (Sakmann and Neher, 1995). For this work E. Neher and B. Sakmann received the Nobel prize in Physiology or Medicine in 1991. Very roughly described, a single ion channel is inserted in the (often artificial) membrane surrounded by an electrolyte with an electrode to measure the current while a constant voltage is applied. For a more detailed explanation of its various configurations see (Sakmann and Neher, 1995) and the references therein. Such recordings can be modeled as a time series, which is equidistantly sampled from the convolution of a piecewise constant signal disturbed by Gaussian white noise with a lowpass filter, for more details see Section 3.1.

In this work, we focus on non-parametric estimation of the underlying signal, since from these estimations typical summary statistics and confidence statements for important channel characteristics such as the number of states, their conductance levels and the dwell time distribution in the different states can be obtained (Colquhoun, 1987; Sakmann and Neher, 1995; Hotz et al., 2013). Like in Figure 1.1, in many applications at least some events are very short in time, often even shorter than the filter length. Estimating such events is particularly challenging, since the amplitudes of them are smoothed by the filter. Moreover, in some measurements, see for instance Figure 3.3, the noise

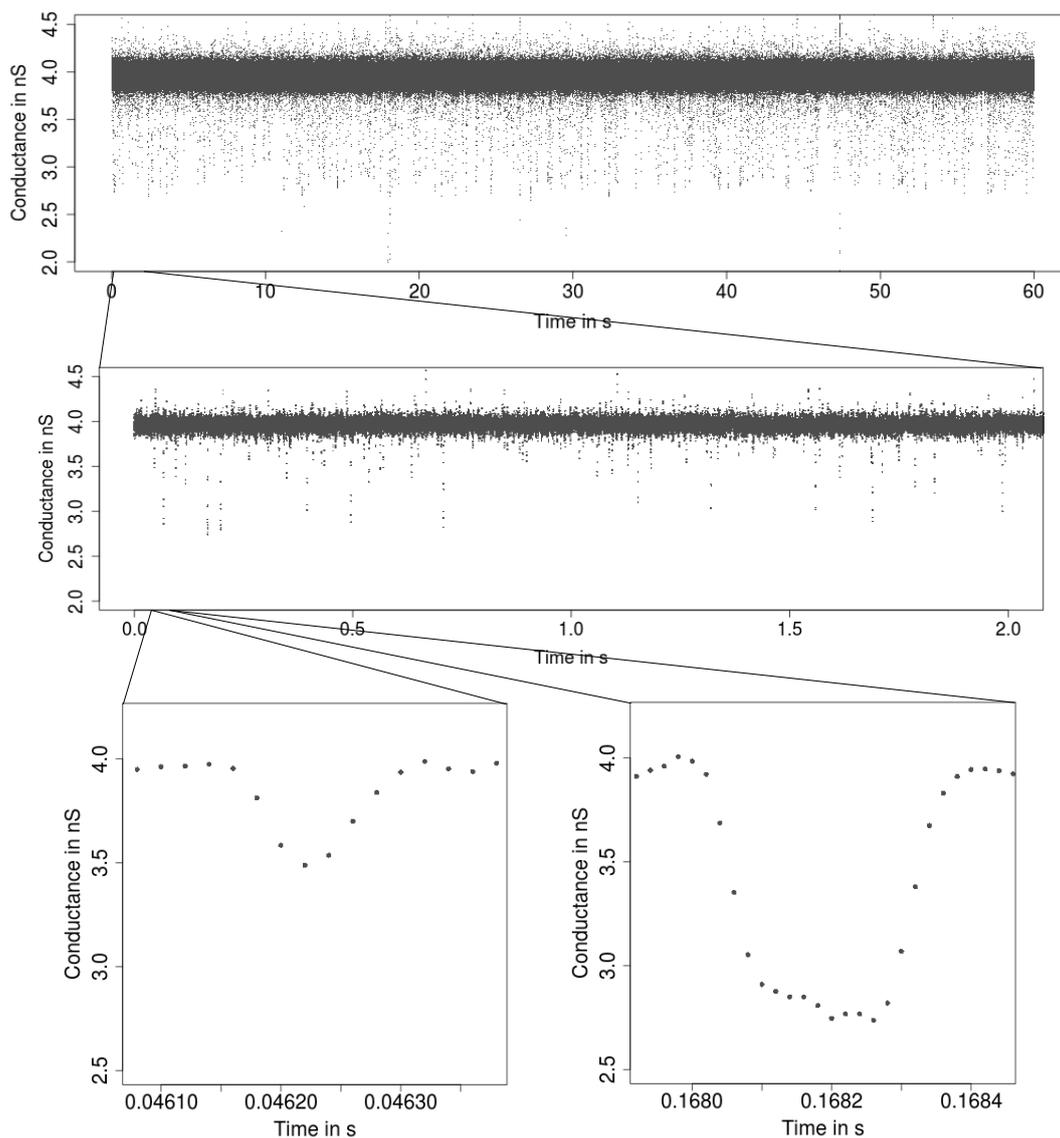


Figure 1.1.: Observations (grey points) of a representative conductance time series of PorB wild type with 1 mM ampicillin recorded by the patch clamp technique using black lipid membranes at 80 mV.

level is larger on segments with a larger conductance. This phenomenon is called *open channel noise*, since a larger conductance results from an open pore. Although for ion channel recordings many non-parametric methods exist, see for instance (Basseville and Benveniste, 1983; Colquhoun, 1987; Sakmann and Neher, 1995; VanDongen, 1996; Hotz et al., 2013; Gnanasambandam et al., 2017), to the best of our knowledge, no method can deal with short events and heterogeneous noise satisfactorily, despite its importance. This is discussed in more detail in Section 3.2.

In summary, we aim to estimate a piecewise constant function, containing very short events, from filtered, potentially heterogeneous observations, while the correlation structure is still explicitly given by the known filter. For this task we will extend a *multiscale*

approach, which was proposed by Frick et al. (2014a) for the simpler model of unfiltered observations and homogeneous noise. The general methodology is presented in the next section.

## 1.1. Multiscale change-point estimation

In this section we present the simultaneous **multiscale change-point estimator**, SMUCE, from (Frick et al., 2014a) in a more general formulation. We assume that  $n$  observations  $\mathbf{Y} = (Y_1, \dots, Y_n)$  are given by the function values of a piecewise constant function  $\mu : [0, 1] \rightarrow \mathbb{R}$  at equidistant design points  $x_i = i/n$  plus standard Gaussian errors  $\epsilon = (\epsilon_1, \dots, \epsilon_n)$  scaled by a known, global standard deviation, i.e.,

$$Y_i = \mu(i/n) + \sigma_0 \epsilon_i, \quad i = 1, \dots, n. \quad (1.1)$$

Here, an equidistant design in the unit interval is assumed, but extensions to other fixed designs are straightforward. For ion channel recordings we use for instance instead of the unit interval the physical time since recording started.

The estimator SMUCE from (Frick et al., 2014a) is defined as the restricted maximizer  $\hat{\mu}$  of a functional  $L(\mathbf{Y}, \mu)$ . Often the restricted maximum likelihood estimator is used, because of its efficiency. Thereby, the candidate set of all piecewise constant functions  $\mathcal{M}$ , see (2.4), is restricted to all solutions  $C(\mathbf{Y}, \mathbf{q})$  of the (non-convex) optimization problem to minimize the number of change-points under the constraint that the candidate function is accepted by a multiscale test. In formulas,

$$\hat{\mu} := \operatorname{argmax}_{\mu \in C(\mathbf{Y}, \mathbf{q})} L(\mathbf{Y}, \mu), \quad (1.2)$$

where

$$C(\mathbf{Y}, \mathbf{q}) := \{\operatorname{argmin}_{\mu \in \mathcal{M}} |\mu|_0 \text{ s.t. } T_n(\mathbf{Y}, \mu, \mathbf{q}) \leq 0\}, \quad (1.3)$$

with  $|\mu|_0$  the number of change-points of  $\mu$ ,  $\mathbf{q} = (q_1, \dots, q_n)$  a vector of *critical values*, see below, and *multiscale test* statistic

$$T_n(\mathbf{Y}, \mu, \mathbf{q}) := \max_{\left[\frac{i}{n}, \frac{j}{n}\right] \in \mathcal{I}(\mu)} T_i^j(\mathbf{Y}, \mu([i/n, j/n])) - q_{j-i+1}. \quad (1.4)$$

Here, the *interval set*  $\mathcal{I}$  is a subset of the set of all intervals with start and endpoints at the design points and  $\mathcal{I}(\mu) \subset \mathcal{I}$  is the subset of all intervals on which the candidate function  $\mu$  is constant. We denote by  $\mu([i/n, j/n])$  the function value of  $\mu$  on the interval  $[i/n, j/n]$ . Furthermore,  $T_i^j$  is a *local test statistic* depending on the model. Often the corresponding likelihood ratio test statistic is used due to its power. We test over all intervals in  $\mathcal{I}(\mu)$ , whether  $\mu([i/n, j/n])$  is the expectation of the observations or not. This guarantees that the estimate  $\hat{\mu}$  describes the data everywhere locally well and detects in this sense all

significant change-points. On the other hand, minimizing the number of change-points guarantees not to include additional artificial change-points.

For this model the the vector of critical values  $\mathbf{q}$  can be obtained by scale penalization, see (Dümbgen and Spokoiny, 2001; Dümbgen and Walther, 2008; Frick et al., 2014a). More precisely, different scales are balanced by a penalty function  $p_{j-i+1,n}(\cdot)$  leading to the *penalized multiscale statistic*

$$T_n^{p_{j-i+1,n}}(\mathbf{Y}, \mu) := \max_{\left[\frac{i}{n}, \frac{j}{n}\right] \in \mathcal{I}(\mu)} p_{j-i+1,n}(T_i^j(\mathbf{Y}, \mu([i/n, j/n]))). \quad (1.5)$$

Then, the *global quantile*  $q_\alpha$  at significance level  $\alpha$  is defined as the  $(1 - \alpha)$  quantile of  $T_n^{p_{j-i+1,n}}(\boldsymbol{\epsilon}, 0)$ , with 0 denoting the function identical to zero, and the critical values are obtained by  $q_l = p_{l,n}^{-1}(q_\alpha)$ ,  $l = 1, \dots, n$ . Note that the same threshold is used for all intervals of the same length as no a-priori information on the change-point locations is assumed. The most common penalization

$$p_{l,n}(t) = \sqrt{2t} - \sqrt{2 \log(n \exp(1)/l)} \quad (1.6)$$

implies for many models that the penalized multiscale statistic converges in distribution to an almost surely finite limit distribution, see (Dümbgen and Spokoiny, 2001; Dümbgen and Walther, 2008; Frick et al., 2014a), which guarantees appropriate scale balancing. We will see later that this is not true when we assume heterogeneous noise and that for this model a different approach is required.

Such an estimator has many good estimation properties like to overestimate the number of observations only with probability  $\alpha$ , to estimate the number of change-points consistently and to detect vanishing signals with the optimal rate, see (Frick et al., 2014a, (14), Corollary 1, Theorems 3, 5, 6). In addition, confidence statements like confidence intervals for the change-point locations and a confidence band for the whole signal are obtained, see (Frick et al., 2014a, Corollaries 2 and 3, Theorem 7). Hence, in this work we extend this methodology to Gaussian observations with heterogeneous noise and to filtered observations. For filtered observations we have to combine the multiscale estimator with a deconvolution approach. Since a model with filtered observations, heterogeneous noise and events on all scales is very difficult, we start with unfiltered, independent observations, but with heterogeneous noise. Although this model is already quite challenging, it still allows to establish a substantial theory.

## 1.2. Heterogeneous change-point estimation

For this task we present in Section 2 the **H**eterogeneous **S**imultaneous **M**ultiscale **C**hange-point **E**stimator, H-SMUCE, which we proposed in (Pein et al., 2017c). Motivated by the ion channel recordings we assume that the unknown variance of the observations can only change at the same locations as the unknown signal, but does not have to, in particular

homogeneous observations are still part of the model. More precisely, the global standard deviation in (1.1) is replaced by a piecewise constant function that has the same change-point locations as the signal  $\mu$  or less. We stress that this restriction is only required for our theory. For practical applications, where such an assumption is often violated, we show in simulations in Section 2.4.3 that H-SMUCE is robust against a violation of this assumption, i.e., a change in variance may occur without a change in the signal.

This is contrary to methods that consider changes in the variance as relevant structural changes of the underlying data, even when the expectation does not change, and look for changes in the expectation and in the variance. In this spirit are local search methods, such as binary segmentation (Scott and Knott, 1974; Vostrikova, 1981) if the corresponding single change-point detection method takes the heterogeneous variance into account, but also global methods can achieve this goal, e.g., the pruned exact linear time method, PELT, (Killick et al., 2012). For a Bayesian approach in this context see (Du et al., 2016) and the references therein. In addition, methods which search for more general structural changes in the distribution potentially apply to this set-up as well, see for instance (Csörgo and Horváth, 1997; Arlot et al., 2012; Matteson and James, 2014; Zou et al., 2014; Haynes et al., 2017) among others.

In contrast to this setting, H-SMUCE in Section 2 considers the variance as a nuisance parameter and we primarily seek for changes in the signal  $\mu$ . Hence, we aim for statistically efficient estimation of the signal  $\mu$ , but still being robust against heterogeneous noise. Obviously, this cannot be achieved by methods addressing the first setting. Although of great practical relevance, this situation has only rarely been considered and in particular no rigorous theory exists, to our knowledge. The cross-validation method LOOVF (Arlot and Celisse, 2011) and cumSeg (Muggeo and Adelfio, 2011) have been designed specifically to be robust against heterogeneous noise. Moreover, also circular binary segmentation (CBS), see (Venkatraman and Olshen, 2007), applies to this.

It is important that our model is also fundamental different as the model in Enikeeva et al. (2016), since they assume that the variance has to change when the signal changes. Moreover, they assume that the variances are known. This increases potentially the detection power as under this assumption variance changes can be used for finding signal changes, as well. The exact gain is quantified in their paper and will be briefly discussed in Section 2.1. In contrast, in our model variance changes can potentially even have an adverse effect, since the nuisance parameter  $\sigma^2(\cdot)$  hinders estimation.

We define H-SMUCE as the multiscale constrained maximum likelihood estimator like in (1.2) with  $L(\mathbf{Y}, \mu)$  being the likelihood function, see (2.3), and local test statistic  $T_i^j(\mathbf{Y}, m_{ij})$  being the likelihood ratio test statistic

$$T_i^j(\mathbf{Y}, m_{ij}) := (j - i + 1) \frac{(\bar{Y}_{ij} - m_{ij})^2}{2\hat{s}_{ij}^2}, \quad (1.7)$$

with  $\bar{Y}_{ij} := (j-i+1)^{-1} \sum_{l=i}^j Y_l$  and local variance estimate  $\hat{s}_{ij}^2 := (j-i)^{-1} \sum_{l=i}^j (Y_l - \bar{Y}_{ij})^2$ . In comparison to the homogeneous problem this simply replaces the constant variance by an estimate of the local variance, i.e., we use instead of a Gauss test a t-test, with squared test statistic.

To allow fast computation and to simplify the asymptotic analysis, we choose for the interval system  $\mathcal{I}$  the *dyadic partition*  $\mathcal{D}$ , see (2.8), which contains at most  $n - 1$  intervals. Nevertheless, our methodology can easily be adapted to other intervals systems, see Remark 2.

So far the adaption to heterogeneous variance was not very difficult, but it remains the most challenging task. For the construction of a multiscale test a vector of critical values  $\mathbf{q}$ , that combines the local tests appropriately, has to be chosen. To this end, note that in the heterogeneous case scale penalization as in the homogeneous case (Dümbgen and Spokoiny, 2001; Dümbgen and Walther, 2008; Frick et al., 2014a) does not balance scales anymore appropriately. In particular, this will give a multiscale statistic which diverges asymptotically, since due to the local variance estimation the test statistic fails to have subgaussian, but still has subexponential, tails. To overcome these burdens we introduce in Section 2.2.1 scale dependent critical values such that the multiscale test has significance level  $\alpha$ , see (2.12), and such that the different scales are balanced appropriately by weights  $\beta_1, \dots, \beta_n$ , see (2.13) and (2.14). More precisely, these weights determine the ratios between the rejection probabilities of the multiscale test on a corresponding scale. Existence and uniqueness of the so defined scale dependent critical values is shown in Lemma 1 and explicit bounds are given in Lemma 3. The weights also allow to incorporate prior scale information, see Section 2.3.4.

### 1.3. Theoretical results for H-SMUCE

Using the so defined vector of critical values  $\mathbf{q}$  allows to obtain several confidence statements which are a main feature of H-SMUCE. First of all, we show in Section 2.3 that the probability to overestimate the number of change-points is bounded by the significance level  $\alpha$  of the multiscale test uniformly over  $\mathcal{S}$  in (2.2),  $\mathbb{P}(\hat{K} > K) \leq \alpha$ , see Theorem 5. More specifically, we show the *overestimation* bound

$$\sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{P}(\hat{K} > K + 2k) \leq \alpha^{k+1}, \quad \forall k \in \mathbb{N}_0, \quad (1.8)$$

see Theorem 6. In Theorem 7 we provide an exponential bound for the underestimation of the number of change-points by H-SMUCE,  $\mathbb{P}(\hat{K} < K)$ . To this end, we show new exponential deviation bounds for  $F$ -statistics, see Section A.3, which might be of interest by its own. Combining the over- and the underestimation bound provides upper bounds for the errors  $\mathbb{P}(\hat{K} \neq K)$  and  $\mathbb{E}[|\hat{K} - K|]$ . For a fixed signal both bounds vanish super polynomially in  $n$  if  $\alpha = \alpha_n \searrow 0$  when the weights are chosen appropriately, see Remark 8. Consequently, the estimated number of change-points converges almost surely to the

true number, see Theorem 9. Further, these exponential bounds enable us to obtain a confidence band for the signal  $\mu$  as well as confidence intervals for the locations of the change-points, for an illustration see Figures 1.2 and 2.1. We show that the diameters of the confidence intervals decrease asymptotically as fast as the (optimal) sampling rate up to a log factor. All confidence statements hold uniformly over  $\mathcal{S}_{\Delta,\lambda} \subset \mathcal{S}$ , all functions with minimal signal to noise ratio  $\geq \Delta$  and minimal scale  $\geq \lambda := \min_{k=0,\dots,K} \tau_{k+1} - \tau_k$ , with  $\Delta$  and  $\lambda$  arbitrarily, but fixed, see Theorems 10 and 11. Finally, H-SMUCE detects vanishing signal with the optimal minimax rate, even when the number of change-points is unbounded, only the constants are slightly worse. Remarkably, the only extra assumption we have to suppose is that signal and variance have to be constant on segments at least of order  $\log(n)/n$ , see Theorem 13. This reflects the additional difficulty to separate locally signal and noise levels in a multiscale fashion.

## 1.4. H-SMUCE in action

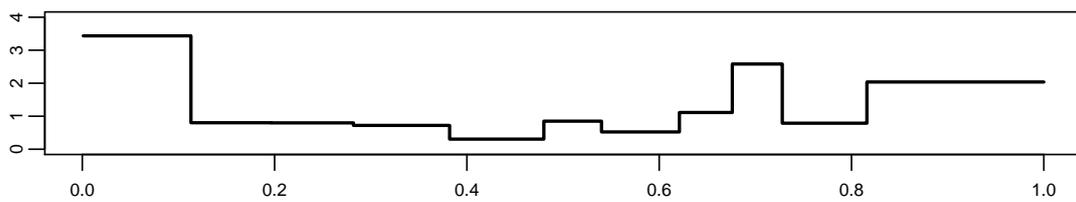
Figure 1.2 illustrates the performance of H-SMUCE in an example with  $n = 1000$  observations and  $K = 10$  change-points. We found that H-SMUCE misses for  $\alpha = 0.1$  one change-point as the choice  $\alpha = 0.1$  tunes H-SMUCE to provide the strong guarantee not to overestimate the number of change-points  $K$  with probability 0.9, see (1.8). But, for  $\alpha$  between 0.15 and 0.99, only displayed for  $\alpha = 0.5$ , the correct number of change-points is detected, while providing a weaker guarantee for not overestimating  $K$ . In addition, for  $\alpha$  between 0.15 and 0.99 each true change-point is covered by the associated confidence interval at level  $1 - \alpha$ . Notably, the reconstructions are remarkably stable in  $\alpha$ . In fact, combining Lemma 3 and (4.1) shows that the width of the confidence band is proportional to  $\sqrt{\log(1/\alpha)}$  which decreases only logarithmically for increasing  $\alpha$ .

We compare H-SMUCE with CBS (Venkatraman and Olshen, 2007), cumSeg (Muggeo and Adelfio, 2011) and LOOVF (Arlot and Celisse, 2011) in several simulation studies in Section 2.4. Thereby, H-SMUCE outperforms these methods as long as the constant segments are too short, since H-SMUCE is forced to be conservative on these scales to control the overall familywise error rate. We also examine robustness issues, see Section 2.4.3. In all of these simulations H-SMUCE performs very robustly and overestimates the number of change-points only rarely, in accordance with (1.8).

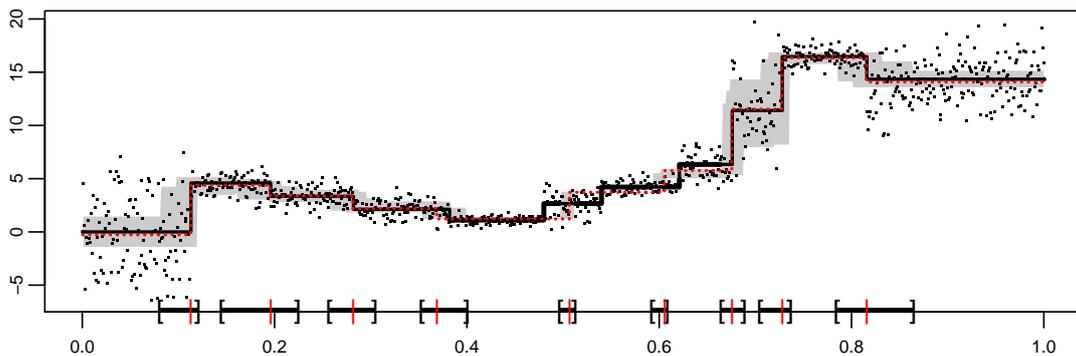
## 1.5. Multiscale estimation for ion channel recordings

We now come back to filtered ion channel recordings. Unfortunately, a straightforward extension of the multiscale regression estimator as described in Section 1.1 to filtered observations is not computational feasible, for a more detailed discussion see Section 3.3. Hence, we present in Section 3 three different methods which can be interpreted as approximations of a multiscale regression estimator for the ion channel models.

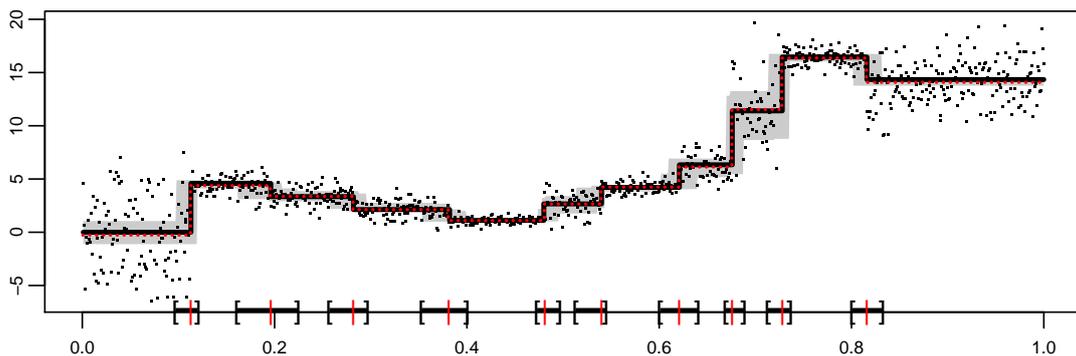
First of all, we extend the **J**ump-**S**egmentation by **M**ulti**R**esolution **F**ilter, J-SMURF,



(a) True standard deviation (black).



(b)  $\alpha = 0.1$ .



(c)  $\alpha = 0.5$ .

Figure 1.2.: (b), (c): Observations (black dots), true mean function (black line), confidence band (grey), confidence intervals for the change-point locations (brackets and thick line) and the estimate (red line) by H-SMUCE at given  $\alpha$  and with equal weights  $\beta_1, \dots, \beta_{d_n}$ , see (2.13) and (2.14).

which was proposed by Hotz et al. (2013), to heterogeneous noise by combining it with H-SMUCE. This method has good detection properties if events are long enough, but almost no power on smaller scales.

To also detect short events and estimate their parameters well we proposed in (Pein et al., 2017d) the **JUmp Local dEconvolution Segmentation** filter, **JULES**. This estimator obtains by multiscale estimation a piecewise constant approximation on the convolution which is then refined by postfiltering and deconvolution. The fast decaying filter kernel allows us to perform the deconvolution *locally*, i.e., to estimate only few parameters from only few observations at one time, which makes it computation feasible. Figure 1.3 shows

exemplarily an estimation of the signal underlying the observations in Figure 1.1.

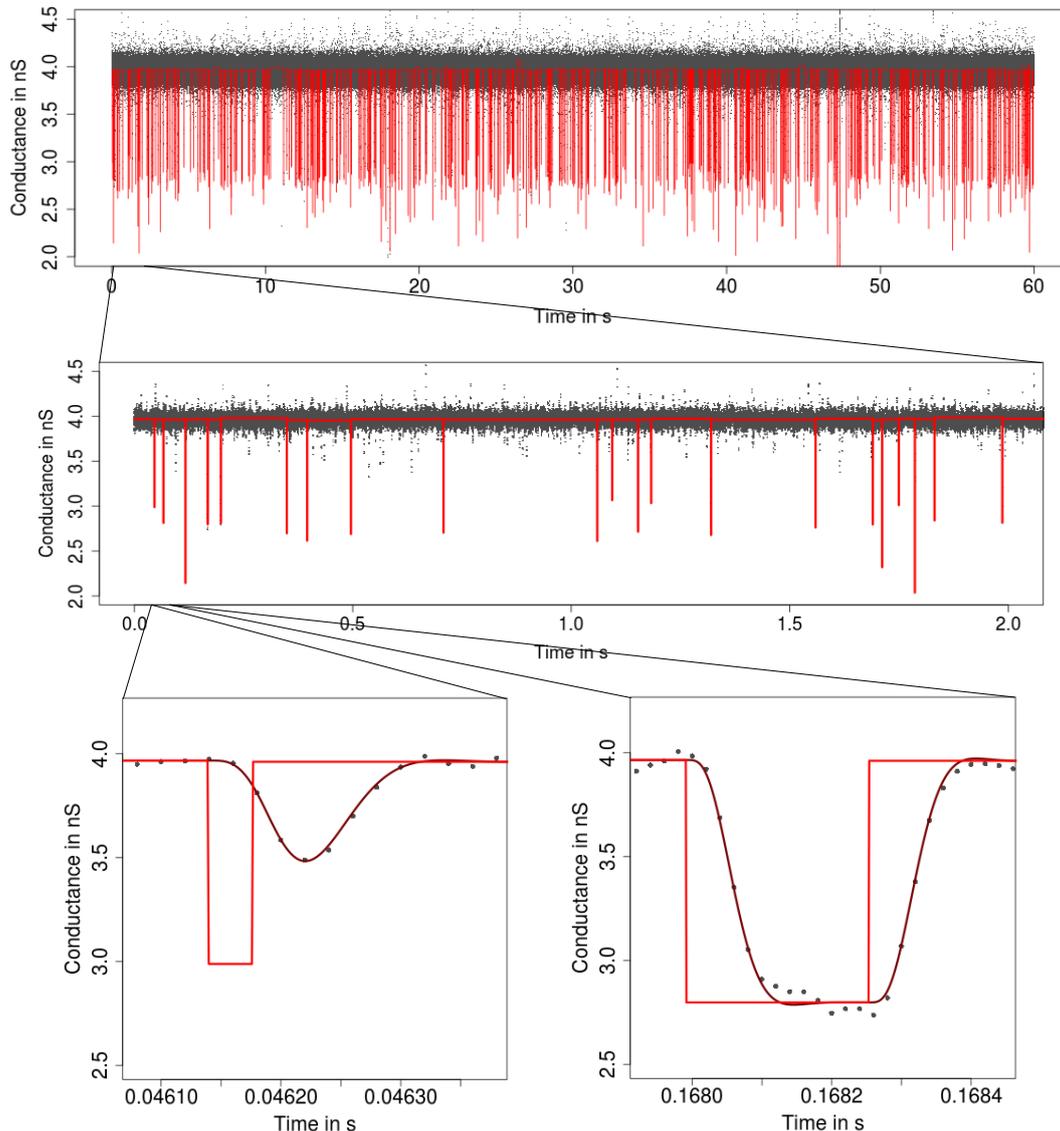


Figure 1.3.: Estimation by JULES (red) of the signal underlying the observations in Figure 1.1 and its convolution with the lowpass filter (darkred).

To also deal with heterogeneous noise we propose in this work a new method called **J-SMURF Improved by Local Tests And Deconvolution**, JILTAD. This approach estimates in a first step events on large scales by J-SMURF and then improves the estimation on small scales by tests that take into account the deconvolution explicitly as well as by deconvolution as above. An exemplary estimation in Figure 3.5 for the observations in Figure 3.3 shows that at presence of heterogeneous noise this method is indeed able to estimate events on small scales well.

The performance of these non-parametric estimation methods is assessed in a comprehensive simulation study and on gramicidin A recordings. They confirm that JULES and

JILTAD are indeed able to detect and estimate events that are short in time with very high precision, JULES only when the noise is homogeneous, but JILTAD also when the noise is heterogeneous.

## 1.6. Implementation and software

All presented methods are implemented in *R* (R Core Team, 2017). The multiscale estimators, in particular SMUCE and H-SMUCE, can be computed efficiently by pruned dynamic programs and implementations are provided by the *stepR* package (Pein et al., 2017a). A first algorithm was described in (Frick et al., 2014a) and implemented by Hotz and Sieling (2015). In (Pein et al., 2017c) we proposed a speed up for smaller intervals sets by computing firstly the confidence intervals and then the estimator by a dynamic program restricted to these intervals. The non-parametric estimation methods for the ion channel recordings are comparable multiscale regression estimators combined with deconvolution which can be computed locally by an iterative grid search. These methods are implemented in the *R* package *clampSeg* (Pein et al., 2017b). More details about the dynamic programs, the computation of the critical values and an overview about the provided *R* packages are given in Section 4 in which we also study the computation time of the algorithms theoretically and in simulations.

## 1.7. Application to PorB recordings

In Section 5 we present the application from (Bartsch et al., 2017), in which we compared how frequently ampicillin enters and the time duration it resides in the PorB wild type and a mutation called G103K as a function of the ampicillin concentration and the applied voltage. The very short durations and the huge amount of observations and events require an automatic analysis of these recordings with high precision on small temporal scales. By using JULES and JILTAD we showed that the average residence time of the ampicillin in the mutant is statistically significant longer than in the wild type and that the conductance of the mutant is smaller. These findings reason, together with other findings by molecular dynamic simulations, that ampicillin passes the mutant less likely, a potential source for antibiotic resistance. Understanding these interactions helps to develop new drugs against multidrug-resistant bacteria, a topic with heavily increasing importance in the past few years (Cosgrove and Carmeli, 2003; Rice, 2007; World Health Organization, 2014).

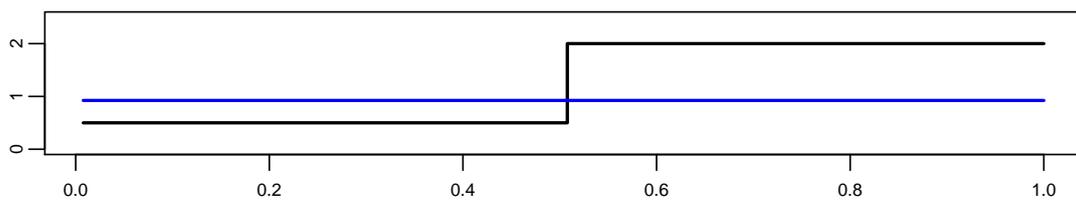
The work ends with a conclusion and an outlook in Section 6. All proofs are given in the Appendix A together with some auxiliary statements.

## 2. Heterogeneous Multiscale Change-Point Inference for Independent Observations

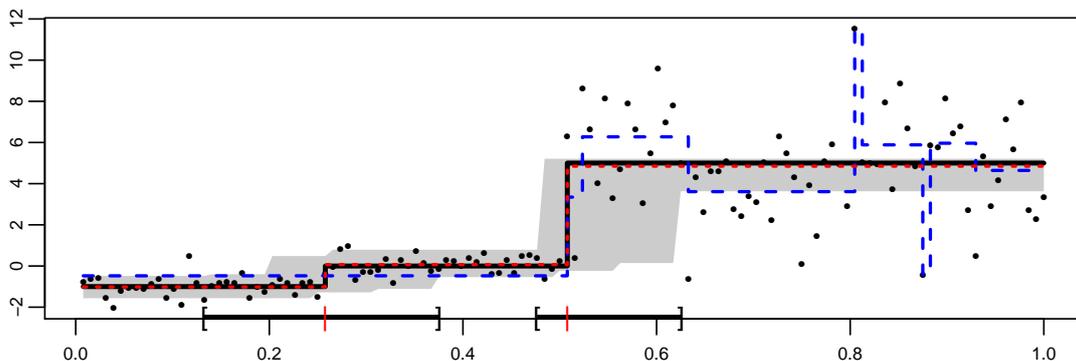
The methods for estimating the change-points in (1.1) and in related models are vast, see for instance (Yao, 1988; Donoho and Johnstone, 1994; Csörgo and Horváth, 1997; Bai and Perron, 1998; Braun et al., 2000; Birgé and Massart, 2001; Kolaczyk and Nowak, 2005; Boysen et al., 2009b; Harchaoui and Lévy-Leduc, 2010; Jeng et al., 2010; Killick et al., 2012; Rigollet and Tsybakov, 2012; Zhang and Siegmund, 2012; Fryzlewicz, 2014; Frick et al., 2014a; Du et al., 2016; Yau and Zhao, 2016; Li et al., 2016; Fang et al., 2016; Maidstone et al., 2017; Eichinger and Kirch, 2018) and the references in these works. However, a crucial condition in most of the afore-mentioned works is the assumption of homogeneous noise, i.e., that all errors have the same variance. In many applications, however, this assumption is violated and the variance varies over time,  $\sigma^2(i/n)$ , say. In addition to ion channel recordings, where open channel noise causes heterogeneous noise, this problem arises for instance in genetic, e.g., in the analysis of array CGH data, see (Muggeo and Adelfio, 2011; Arlot and Celisse, 2011). Further examples include economic applications, for instance the real interest rate is modeled by Bai and Perron (2003) as piecewise linear regression with covariates and heterogeneous noise. Hence, this setup is of its own interest and not limited to ion channel recordings.

To illustrate the effects of missing heterogeneity we show in Figure 2.1 a reconstruction by SMUCE (Frick et al., 2014a), implemented in the *R* package *stepR* (Pein et al., 2017a), assuming homogeneous noise.

The constant variance assumption of SMUCE leads to overestimation of the standard deviation, which is here pre-estimated by a global IQR type estimator based on differences similar to (3.11), in the first half and underestimation in the second half. Therefore, in Figure 2.1 SMUCE misses the first change-point and includes artificial change-points in the second half to compensate for the too small variance it is forced to use, see also (Zhou, 2014). Note, that this flaw is not a particular feature of SMUCE, it will occur for any sensible segmentation method which relies on a constant variance assumption. Hence, from Figure 2.1 the fundamental difficulty of the heterogeneous (multiscale) change-point regression problem becomes apparent: How to decide whether a change of fluctuations of the data result from highly frequent changes in the expectation  $\mu$  or merely from an increase of the noise level? Apparently, if changes can occur on any scale this is a notoriously difficult issue and proper separation of signal and noise cannot be performed without extra information.



(a) True (black) and estimated (blue) standard deviation.



(b) Simulated observations (black dots) together with the true signal (black line), the confidence band (grey), the confidence intervals for the change-point locations (brackets) as well as the estimates by our new method H-SMUCE (red dotted line) and by SMUCE (blue dashed line), both with  $\alpha = 0.1$ .

Figure 2.1.: Illustration of missing heterogeneity.

Indeed, the basis of our approach is that it is often a reasonable assumption to exclude changes of the variance in constant segments of  $\mu$ , see the examples above and Section 3.1.2 for the heterogeneous ion channel model. Under this relatively weak assumption, we show in this work that estimation of  $\mu$  for heterogeneous data becomes indeed feasible. In addition, we also aim for a method which is robust when changes in the variance occur at locations where the signal is constant, as we believe that this cannot be excluded in many practical cases. To this end, we proposed in (Pein et al., 2017c) a new **H**eterogeneous **S**imultaneous **M**ultiscale **C**hange-point **E**stimator, H-SMUCE, which recovers the signal under heterogeneous noise over a broad range of scales, controls the familywise error rate to overestimate the number of change points, allows for confidence statements, obeys certain statistical optimality properties and can be efficiently computed. At the same time it is robust against heterogeneous noise on constant signal segments and as a by-product also against more heavily tailed errors.

## 2.1. Heterogeneous change-point model

To be more specific, we assume the *heterogeneous* Gaussian change-point model

$$Y_i = \mu(i/n) + \sigma(i/n)\epsilon_i, \quad i = 1, \dots, n, \quad (2.1)$$

where now the variance  $\sigma^2$  is also given by an unknown piecewise constant function and  $\epsilon_1, \dots, \epsilon_n$  are assumed to be independent, standard Gaussian errors. For the following theoretical results we assume that the variance function  $\sigma^2$  can have change-points only at the same locations as the expectation function  $\mu$ . In other words,  $(\mu, \sigma^2)$  is a pair of unknown piecewise constant functions in

$$\mathcal{S} := \left\{ (\mu, \sigma^2) : [0, 1] \mapsto \mathbb{R}^2, \mu = \sum_{k=0}^K m_k \mathbb{1}_{[\tau_k, \tau_{k+1})}, \sigma^2 = \sum_{k=0}^K s_k^2 \mathbb{1}_{[\tau_k, \tau_{k+1})}, K \in \mathbb{N} \right\}, \quad (2.2)$$

with unknown change-point locations  $\tau_0 = 0 < \tau_1 < \dots < \tau_K < 1 = \tau_{K+1}$  for some unknown number of change-points  $K \in \mathbb{N}$  and also unknown function values  $m_k \in \mathbb{R}$  and  $s_k^2 \in \mathbb{R}_+$  of  $\mu$  and  $\sigma^2$ . For technical reasons, we define  $\mu(1)$  and  $\sigma^2(1)$  by continuous extension of  $\mu$  and  $\sigma^2$ , respectively. For identifiability of  $\mu$  we assume  $m_k \neq m_{k+1} \forall k = 0, \dots, K$  and exclude isolated changes in the signal by assuming that  $\mu$  is a right continuous function. It is important to stress that in (2.2) we allow the variance to *potentially* have changes at the locations of the changes of the signal, but the variance  $\sigma^2$  need not necessarily change when  $\mu$  changes, as we do not assume  $s_k^2 \neq s_{k+1}^2$ . In particular, homogeneous observations are still part of the model. Vice versa, we assume that within a constant segment of  $\mu$  it may *not* happen that the variance changes, i.e., the local signal to noise ratio is assumed to be constant on  $[\tau_k, \tau_{k+1})$  for all  $k = 0, \dots, K$ . We argue that this is a reasonable assumption in many applications, in particular for ion channel recordings, see Section 3.1.2, since a change-point represents typically a change of the condition of the underlying state. Moreover, for example, in many engineering applications a locally constant signal to noise ratio is assumed (Guillaume et al., 1990), which motivates our modeling as well. However, we stress that the restriction to model (2.2) is only required for our theory. For the practical application we show in simulations in Section 2.4.3 that H-SMUCE is robust against a violation of this assumption, i.e., when a change in variance may occur without a change in the signal, and hence still works well in a general heterogeneous change-point model with arbitrary variance changes.

For a better understanding it is illustrative to distinguish our setting from the case when it is known *before hand* that changes in the variance occur necessarily with changes in the signal. This increases potentially the detection power as under this assumption variance changes can be used for finding signal changes, as well. The information gain due to the variance changes for this case has been recently quantified by Enikeeva et al. (2016) in terms of the minimax detection boundary for single vanishing signal bumps of size  $\delta_n \searrow 0$ . More precisely, if the base line variance is  $\sigma_0^2$  and the variance at the bump is  $\sigma_0^2 + \sigma_n^2$  then the constant in the minimax detection boundary is  $b = \sqrt{2}\sigma_0\sqrt{2/(2+c^2)}$  for  $c = \sigma_0^{-1} \lim_{n \rightarrow \infty} \sigma_n/\delta_n$ , see (Enikeeva et al., 2016, Theorems 3.1-3.3). For the particular case of homogeneous variance, i.e.,  $\sigma_n^2 = 0$ , we obtain  $b = \sqrt{2}\sigma_0$  and the factor  $\sqrt{2/(2+c^2)} = 1$  becomes maximal, see also (Dümbgen and Walther, 2008; Frick et al., 2014a). This reflects that no additional information on the location of a change can

be gained from the variance in the homogeneous case. Comparing this to the heterogeneous case we see that when the variance change is known to be large enough, i.e.,  $\sigma_0^{-1} \lim_{n \rightarrow \infty} \sigma_n / \delta_n > 0$ , additional information for the signal change can be gained from the variance change, as then  $b < \sqrt{2}\sigma_0$ , provided it is known that signal and variance change *simultaneously*.

In contrast, in the present setting the variance need not necessarily change when the signal changes, hence the worst case of no variance change from above is contained in our model, which lower bounds the detection boundary. The situation is further complicated due to the fact that missing knowledge of a variance change can potentially even have an adverse effect because in model (2.2) detection power will be potentially decreased further as the nuisance parameter  $\sigma^2(\cdot)$  hinders estimation of change-points of  $\mu$ . For this situation the optimal minimax *constants* are unknown to us, but from the fact that the model with a constant variance is a submodel of our model (2.2) it immediately follows that the minimax constant for a single bump has to be at least  $\sqrt{2}\sigma_0$ . This will allow us to show that H-SMUCE attains the same optimal minimax detection *rate* as for the homogeneous case and  $4\sigma_0$  instead of  $\sqrt{2}\sigma_0$  as the constant appearing in the minimax detection boundary. Remarkably, the only extra assumption we have to suppose is that signal and variance have to be constant on segments at least of order  $\log(n)/n$ , see Theorem 13. This reflects the additional difficulty to separate locally signal and noise levels in a multiscale fashion. In other words, when we assume that the number of i.i.d. neighboring observations (no change in signal and variance) in each segment is at least of order  $\log(n)$ , separation of signal and noise is done by H-SMUCE in an optimal way (possibly up to a constant).

## 2.2. Methodology

Assuming the heterogeneous multiple Gaussian change-point model as described in Section 2.1, we propose the **H**eterogeneous **S**imultaneous **M**ultiscale **C**hange-point **E**stimator, H-SMUCE. Like the general multiscale regression estimator in (1.2), this estimator is defined as the restricted maximum likelihood estimator

$$\hat{\mu} := \operatorname{argmin}_{\mu \in C(\mathbf{Y}, \mathbf{q})} L(\mathbf{Y}, \mu) := \operatorname{argmin}_{\mu \in C(\mathbf{Y}, \mathbf{q})} \sum_{i=1}^n \frac{(Y_i - \mu(i/n))^2}{\hat{\sigma}_i^2}, \quad (2.3)$$

with  $\hat{\sigma}_i^2 := (v - u + 1)^{-1} \sum_{j \in [u/n, v/n]} (Y_j - \mu(j/n))^2$  and  $u, v$  the smallest and largest integers, respectively, such that  $\mu$  is constant on the interval  $[u/n, v/n]$  and  $i/n$  is contained in this interval. Here, the set of all candidate functions, which is the set of all piecewise constant functions

$$\mathcal{M} := \left\{ \mu : [0, 1] \mapsto \mathbb{R}, \mu = \sum_{k=0}^K m_k \mathbb{1}_{[\tau_k, \tau_{k+1})}, K \in \mathbb{N} \right\}, \quad (2.4)$$

is restricted to the multiscale constraint

$$C(\mathbf{Y}, \mathbf{q}) := \{\operatorname{argmin}_{\mu \in \mathcal{M}} |\mu|_0 \text{ s.t. } T_n(\mathbf{Y}, \mu, \mathbf{q}) \leq 0\}. \quad (2.5)$$

The vector of critical values  $\mathbf{q} = (q_1, \dots, q_n)$  will be defined in Section 2.2.1. In the *multiscale test* statistic

$$T_n(\mathbf{Y}, \mu, \mathbf{q}) := \max_{\left[\frac{i}{n}, \frac{j}{n}\right] \in \mathcal{D}(\mu)} T_i^j(\mathbf{Y}, \mu([i/n, j/n])) - q_{j-i+1} \quad (2.6)$$

we choose for the interval set  $\mathcal{I}$  the dyadic partition  $\mathcal{D}$ , to be defined below, and for the local test statistic  $T_i^j$  the likelihood ratio test statistic

$$T_i^j(\mathbf{Y}, m_{ij}) := (j - i + 1) \frac{(\bar{Y}_{ij} - m_{ij})^2}{2\hat{s}_{ij}^2}, \quad (2.7)$$

with  $\bar{Y}_{ij} := (j-i+1)^{-1} \sum_{l=i}^j Y_l$  and local variance estimate  $\hat{s}_{ij}^2 := (j-i)^{-1} \sum_{l=i}^j (Y_l - \bar{Y}_{ij})^2$ . The *dyadic partition*  $\mathcal{D}$  is defined as

$$\mathcal{D} := \bigcup_{k=1}^{d_n} \mathcal{D}_k, \quad (2.8)$$

where  $d_n := \lfloor \log_2(n) \rfloor$  is the number of different scales and

$$\mathcal{D}_k := \left\{ \left[ \frac{1 + (l-1)2^k}{n}, \frac{l2^k}{n} \right], l = 1, \dots, \left\lfloor \frac{n}{2^k} \right\rfloor \right\} \quad (2.9)$$

defines the set of intervals from the dyadic partition with length  $2^k n^{-1}$ . This choice allows fast computation and simplifies the asymptotic analysis, since it contains at most  $n - 1$  intervals. Nevertheless, our methodology can easily be adapted to other intervals systems, see Remark 2.

### 2.2.1. Scale dependent critical values

For the definition of H-SMUCE it remains to determine the vector of critical values  $\mathbf{q}$  in (2.5). Since only the entries with indices  $2^1, \dots, 2^{d_n}$  matter, we write, with slight abuse of notation,  $\mathbf{q} = (q_1, \dots, q_{d_n})$  until the end of Section 2. Then, we require that the multiscale test on the r.h.s. in (2.5) is a level  $\alpha$  test, i.e.,

$$\sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{P}_{(\mu, \sigma^2)} \left( \max_{\left[\frac{i}{n}, \frac{j}{n}\right] \in \mathcal{D}(\mu)} \left[ T_i^j(\mathbf{Y}, \mu([i/n, j/n])) - q_{\log_2(j-i+1)} \right] > 0 \right) \leq \alpha, \quad (2.10)$$

where  $\mathbb{P}_{(\mu, \sigma^2)}$  stands for the model (2.1), (2.2) with expectation-variance pair  $(\mu, \sigma^2)$ . To this end, we consider a vector of standard Gaussian observations  $\mathbf{Z} = (Z_1, \dots, Z_n)$  instead of  $\mathbf{Y} = (Y_1, \dots, Y_n)$ , since the supremum in (2.10) is attained at  $\mu \equiv 0$  and  $\sigma^2 \equiv 1$ , see

the proof of Theorem 5. And, we define the statistics  $T_1, \dots, T_{d_n}$  with  $\mathcal{D}_k$  in (2.9) as

$$T_k := \max_{[i/n, j/n] \in \mathcal{D}_k} T_i^j(\mathbf{Z}, 0) \text{ for } k = 1, \dots, d_n. \quad (2.11)$$

Then, the critical values  $q_1, \dots, q_{d_n}$  fulfill (2.10) if

$$\mathbb{P} \left( \max_{k=1, \dots, d_n} [T_k - q_k] > 0 \right) = 1 - F(q_1, \dots, q_{d_n}) = \alpha, \quad (2.12)$$

with  $F$  the cumulative distribution function of  $(T_1, \dots, T_{d_n})$ .

As the critical values  $q_1, \dots, q_{d_n}$  are not uniquely determined by (2.12) they can be chosen to render the multiscale test particularly powerful for certain scales. To this end, we introduce weights

$$\beta_1, \dots, \beta_{d_n} \geq 0, \text{ with } \sum_{k=1}^{d_n} \beta_k = 1, \quad (2.13)$$

where  $\beta_k = 0$  means to omit the  $k$ -th scale, i.e.,  $q_k = \infty$ . Finally, we define  $q_1, \dots, q_{d_n}$  implicitly through

$$\frac{1 - F_1(q_1)}{\beta_1} = \dots = \frac{1 - F_{d_n}(q_{d_n})}{\beta_{d_n}}, \quad (2.14)$$

with  $F_k$  the cumulative distribution function of  $T_k$ . If  $\beta_k = 0$  we omit the corresponding terms in the system of equations in (2.14). The weights determine the fractions between the probabilities that a test on a certain scale rejects, and hence regulate the allocation of the level  $\alpha$  among the single scales. In summary, the choice of the vector of critical values  $\mathbf{q} = (q_1, \dots, q_{d_n})$  boils down to choosing the significance level  $\alpha$  and the weights  $\beta_1, \dots, \beta_{d_n}$ . We discuss possible choices in Section 2.3.4 more carefully. If no prior information on scales is available, a default option is always to set all weights equal, i.e.,  $\beta_1 = \dots = \beta_{d_n} = 1/d_n$ . The next result shows that the vector of critical values satisfying (2.12)-(2.14) is always well-defined.

**Lemma 1** (Existence and uniqueness). *For any  $\alpha \in (0, 1)$  and for any weights  $\beta_1, \dots, \beta_{d_n}$ , s.t. (2.13) holds, there exists a unique vector of critical values  $\mathbf{q} = (q_1, \dots, q_{d_n}) \in \mathbb{R}_+^{d_n}$  which fulfills the equations (2.12) and (2.14).*

An explicit computation of the vector  $\mathbf{q}$  (or  $F$ ) appears to be very hard, since the statistics  $T_1, \dots, T_{d_n}$  are dependent, although the dependence structure is explicitly known. Alternatively, it would be helpful to have an approximation for the distribution (and hence its quantiles) of the maximum in (2.10), which, however, appears to be rather difficult, as well. Therefore, we determine in Section 4.2.2 the vector  $\mathbf{q}$  by Monte-Carlo simulations. Note that the distribution does not depend on the specific element  $(\mu, \sigma^2) \in \mathcal{S}$ , and hence the critical values can be computed in a universal manner. We stress that the determination of the scale dependent critical values is not restricted to our setting and can also be applied to multiscale testing in other contexts. Different to scale penalization and similar to the block criterion in (Rufibach and Walther, 2010), no model dependent derivations are

required and the critical values are adapted to the exact finite sample distribution of the local test statistics. However, our approach allows additionally a flexible scale calibration by the choice of the weights, see Section 2.3.4, and arbitrary interval sets can be used as the following remark points out.

*Remark 2.* This approach can easily be adjusted to an arbitrary interval set  $\mathcal{I}$  and to all multiscale tests such that Lemma 1 is still satisfied. This requires that under the null hypothesis the distribution of the local test statistics do not depend on the unknown parameter as well as monotonicity and continuity of the cumulated distribution functions. To this end, we replace the vector  $(T_1, \dots, T_{d_n})$  by the vector  $(T_1, \dots, T_n)$ , empty scales should be omitted, with

$$T_k := \max_{\substack{[i/n, j/n] \in \mathcal{I}, \\ j-i+1=k}} T_i^j(\mathbf{Z}, 0). \quad (2.15)$$

Again it remains to choose the significance level  $\alpha \in (0, 1)$  and the weights  $\beta_1, \dots, \beta_n$  to determine the critical values. Note, however, that the critical values and its bounds in Lemma 3 and therefore the results in Section 2.3 (besides Theorems 5 and 6) will depend on the specific system  $\mathcal{I}$  and the specific local test statistics  $T_i^j$  and have to be computed for each choice separately.

Although depending on the underlying signal, employing a larger interval set than  $\mathcal{D}$  lead in general to a better detection power, but at the price of a larger computation time. Hence, in practice, a trade-off between computational and statistical efficiency may guide this choice as well. For a discussion of the computation time see Section 4.5. Typical choices are the system of all intervals (of order  $\mathcal{O}(n^2)$ , most efficient, but computationally expensive), the system of all intervals containing a dyadic number of observations ( $\mathcal{O}(n \log(n))$ , intermediate efficiency and computation time). Interesting choices might be also approximating sets like  $\mathcal{J}_{\text{app}}$  that were introduced in (Walther, 2010) and are larger than the dyadic partition, but achieve the minimax boundary in the context of density estimation.

## 2.3. Theory

In this section we collect our theoretical results. We start with finite bounds for the critical values. These will allow to bound  $\mathbb{P}(\hat{K} \neq K)$ . With these bounds we obtain confidence statements for the signal  $\mu$  and its main characteristics. Finally, we investigate asymptotic detection rates of H-SMUCE for vanishing signals.

### 2.3.1. Finite bounds for over- and underestimation

In the following we require upper bounds for the critical values, since the definition of the critical values by the equations (2.12)-(2.14) is implicit.

**Lemma 3** (Bound on critical values). *Let  $\mathbf{q} = (q_1, \dots, q_{d_n})$  be the vector of critical values*

defined by (2.12)-(2.14), then for every  $k \in \{2, \dots, d_n\}$  such that

$$2^{-k} \log \left( \frac{n}{2^k \alpha \beta_k} \right) \leq \frac{1}{2} \quad (2.16)$$

we have

$$q_k \leq 8 \log \left( \frac{n}{2^k \alpha \beta_k} \right). \quad (2.17)$$

*Remark 4.* The log term of the bound (2.17) can be split into a scale dependent penalty term  $\log(n2^{-k})$  which is of the same order as the penalties in the homogeneous case in (Dümbgen and Spokoiny, 2001; Frick et al., 2014a), and into the term  $\log((\alpha\beta_k)^{-1})$  which incorporates the significance level  $\alpha$  and the weight  $\beta_k$ .

The following theorem shows that the significance level  $\alpha$  controls the probability to overestimate the number of change-points.

**Theorem 5** (Overestimation control I). *Assume the heterogeneous Gaussian change-point model (2.1). For a signal  $\mu \in \mathcal{M}$ , let  $\mathcal{J}(\mu)$  be the set of its change-points and  $K := |\mathcal{J}(\mu)|$  be its number of change-points. Let further  $\hat{K}$  be the estimated number of change-points by H-SMUCE, i.e.,*

$$\hat{K} := \min \left\{ |\mathcal{J}(\mu)| : \mu \in \mathcal{M} \text{ with } \max_{\substack{[i, j] \in \mathcal{D}(\mu) \\ [i/n, j/n]}} \left[ T_i^j(\mathbf{Y}, \mu([i/n, j/n])) - q_{ij} \right] \leq 0 \right\}. \quad (2.18)$$

Then, for any vector of critical values  $\mathbf{q}$  with significance level  $\alpha \in (0, 1)$  and weights  $\beta_1, \dots, \beta_{d_n}$  in (2.12)-(2.14), uniformly over  $\mathcal{S}$  in (2.2) it holds

$$\sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K} > K \right) \leq \alpha.$$

The theorem gives us a direct interpretation of the parameter  $\alpha$  as the probability to overestimate the number of change-points. This even holds locally, i.e., on every union of adjoining segments of the estimator H-SMUCE with probability  $1 - \alpha$  there are at least as many change-points as detected. Moreover, we strengthen the result by showing that the probability to estimate additional changes decays exponentially fast and hence the expected overestimation is small.

**Theorem 6** (Overestimation control II). *Under the assumptions of Theorem 5, we have*

$$\sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K} > K + 2k \right) \leq \alpha^{k+1}, \quad \forall k \in \mathbb{N}_0.$$

Moreover,

$$\sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{E}_{(\mu, \sigma^2)} \left[ (\hat{K} - K)_+ \right] \leq \frac{2\alpha}{1 - \alpha}.$$

To control the probability  $\mathbb{P}(\hat{K} \neq K)$  we need additionally an upper bound for the probability to *underestimate*  $K$ . Unlike to the *overestimation* bounds in the Theorems 5 and 6

the probability to underestimate cannot be bounded uniformly over  $\mathcal{S}$ , since size and scale of changes could be arbitrarily small. This is made more precise in Theorem 13 which gives the detection boundary in terms of the smallest (standardized) jump size  $\Delta$  and the smallest scale  $\lambda$ . The next theorem provides an exponential bound uniformly over the subset

$$\mathcal{S}_{\Delta,\lambda} := \left\{ (\mu, \sigma^2) \in \mathcal{S} : \Delta \leq \inf_{1 \leq k \leq K} \frac{|m_k - m_{k-1}|}{\max(s_{k-1}, s_k)} \text{ and } \lambda \leq \inf_{0 \leq k \leq K} (\tau_{k+1} - \tau_k) \right\}, \quad (2.19)$$

with  $\Delta, \lambda > 0$  arbitrary, but fixed.

**Theorem 7** (Underestimation control). *Let  $\mathcal{S}_{\Delta,\lambda}$  be as in (2.19) with  $\Delta, \lambda > 0$  arbitrary, but fixed, and  $k_n := \lfloor \log_2(n\lambda/4) \rfloor$ . We define*

$$\eta := \left[ 1 - 3 \exp \left( -\frac{1}{48} \left( \sqrt{\frac{n\lambda\Delta^2}{32}} - \sqrt{16 \log \left( \frac{8}{\lambda\alpha\beta_{k_n}} \right)} \right)_+^2 \right) \right]_+^2.$$

*Under the assumptions of Theorem 5 and if  $n\lambda \geq 32$  and*

$$(n\lambda)^{-1} \log \left( \frac{8}{\lambda\alpha\beta_{k_n}} \right) \leq \frac{1}{512}$$

*are satisfied, then uniformly in  $\mathcal{S}_{\Delta,\lambda}$*

$$\mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K} < K \right) \leq 1 - \eta^K \text{ and } \mathbb{E}_{(\mu, \sigma^2)} \left[ \left( K - \hat{K} \right)_+ \right] \leq K(1 - \eta). \quad (2.20)$$

Roughly speaking, H-SMUCE detects any change-point of the signal  $\mu$  under the assumptions of Theorem 7 at least with probability  $\eta$ . A sharper version with different probabilities  $\eta_1, \dots, \eta_K$  is given in Theorem 19. Such a result clarifies the dependence on the different weights, but is technically more elaborate. Combining Theorems 5, 6 and 7 gives upper bounds for the probability  $\mathbb{P}(\hat{K} \neq K)$  and the expectation  $\mathbb{E}[|\hat{K} - K|]$  that H-SMUCE misspecifies the number of change-points.

*Remark 8* (Vanishing errors). For a fixed signal (fixed  $\Delta$  and  $\lambda$  are sufficient) both errors vanish asymptotically if  $\alpha = \alpha_n \rightarrow 0$  is chosen such that  $\log(\alpha_n\beta_{k_n,n})/n \rightarrow 0$ , with triangular scheme  $\beta_{1,n}, \dots, \beta_{d_n,n}$  for the weights in (2.14). We can achieve a rate arbitrary close to the exponential rate by the choice  $\alpha_n = \exp(-n/r_n)$ , with  $r_n \rightarrow \infty$  arbitrarily slow. The condition on the sequence  $\beta_{k_n,n}$  allows a variety of possible choices of the weights, too. For instance, the choice  $\beta_{1,n} = \dots = \beta_{d_n,n} = 1/d_n$ , which weights all scales equally, fulfils this condition.

A direct consequence is the *strong model consistency* of H-SMUCE.

**Theorem 9** (Strong model consistency). *Assume the setting of Theorem 5 and let  $(\hat{K}_n)_n$  be the sequence of estimated numbers of change-points by H-SMUCE, where  $\hat{K}_n$  is as  $\hat{K}$  with significance level  $\alpha_n$  and corresponding weights  $\beta_{1,n}, \dots, \beta_{d_n,n}$ . Moreover, let  $\mathcal{S}_{\Delta,\lambda}$*

be as in (2.19) with  $\Delta, \lambda > 0$  arbitrary, but fixed, and  $k_n := \lfloor \log_2(n\lambda/4) \rfloor$ . Let  $\rho > 0$  be arbitrary, but fixed. If

$$\lim_{n \rightarrow \infty} \frac{n^{1+\rho}}{\alpha_n} = 0 \text{ and } \lim_{n \rightarrow \infty} \frac{\log(\alpha_n \beta_{k_n, n})}{n} = 0 \quad (2.21)$$

holds, then  $\hat{K}_n \rightarrow K$ , almost surely and uniformly in  $\mathcal{S}_{\Delta, \lambda}$ .

Again, there is a wide range of sequences  $\alpha_n$  and  $\beta_{k_n, n}$  to satisfy (2.21). Moreover, we still have (weak) model consistency, if  $\alpha_n \rightarrow 0$  and the second condition of (2.21) holds.

### 2.3.2. Confidence sets

In this section we obtain confidence sets for the signal  $\mu$  and for the locations of the change-points. First, we show that  $C(\mathbf{Y}, \mathbf{q})$  in (2.5) is a confidence set for the unknown signal  $\mu$ .

**Theorem 10** (Confidence set). *Assume the setting of Theorem 5 and let  $\mathcal{S}_{\Delta, \lambda}$  be as in (2.19) with  $\Delta, \lambda > 0$  arbitrary, but fixed, and  $k_n := \lfloor \log_2(n\lambda/4) \rfloor$ . Let  $C(\mathbf{Y}, \mathbf{q}_n)$  be as in (2.5) and  $\mathbf{q}_n$  be a vector of critical values determined by significance level  $\alpha$  and weights  $\beta_{1, n}, \dots, \beta_{d_n, n}$ , with  $\lim_{n \rightarrow \infty} n^{-1} \log(\beta_{k_n, n}) = 0$ . Then,*

$$\lim_{n \rightarrow \infty} \inf_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)}(\mu \in C(\mathbf{Y}, \mathbf{q}_n)) \geq 1 - \alpha. \quad (2.22)$$

This shows that the asymptotic coverage of  $C(\mathbf{Y}, \mathbf{q}_n)$  is at least  $1 - \alpha$ . Note that also the length of the vector  $\mathbf{Y}$  increases in  $n$ , but we suppressed it to simplify notation. Lemma 20 gives an exponential inequality similar to (2.20) which shows that  $C(\mathbf{Y}, \mathbf{q})$  is also a non-asymptotic confidence set, but at a different significance level. We further derive from this set confidence intervals for the change-point locations.

**Theorem 11** (Change-point locations). *Assume the setting of Theorem 10, where  $\alpha$  is replaced by a sequence  $\alpha_n \rightarrow 0$ . Let  $c_n := r_n/n \leq \lambda/2$  and  $k_n := \lfloor \log_2(nc_n/2) \rfloor$  s.t.*

$$\liminf_{n \rightarrow \infty} \frac{r_n}{\log(n)} > \frac{216}{\min(\Delta^2, 1)} \text{ and } \lim_{n \rightarrow \infty} \frac{\log(\alpha_n \beta_{k_n, n})}{r_n} = 0. \quad (2.23)$$

Then,

$$\lim_{n \rightarrow \infty} \sup_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} \left( \sup_{\hat{\mu} \in C(\mathbf{Y}, \mathbf{q}_n)} \max_{k=1, \dots, K} c_n^{-1} |\tau_k - \hat{\tau}_k| > 1 \right) = 0. \quad (2.24)$$

Here, the rate  $c_n$  is equal to the sampling rate  $1/n$  up to the (logarithmic) rate  $r_n$  depending on the tuning parameters  $\alpha_n \beta_{k_n, n}$ . For example, if  $\alpha_n \beta_{k_n, n} \asymp n^{-\gamma}$ ,  $\gamma \geq 0$ ,  $r_n/\log(n) \rightarrow \infty$  is sufficient to satisfy (2.23). A non-asymptotic statement is given in Lemma 21. For visualization of the confidence statements it is useful to further derive a confidence band  $B(\mathbf{q}_n)$  for the signal as in (Frick et al., 2014a, Corollary 3 and the explanation around). It can be shown that also the collection  $I(\mathbf{q}_n) = \{\hat{K}_n, B(\mathbf{q}_n), [L_k, R_k]_{k=1, \dots, \hat{K}_n}\}$ ,

with  $[L_k, R_k]$  confidence intervals for the change-point locations according to Theorem 11, satisfies (2.22). Recall Figures 1.2 and 2.1 for an illustration. It is also possible to strengthen the statements of this section to sequences of vanishing signals with  $\Delta_n \rightarrow 0$  and  $\lambda_n \rightarrow 0$  slow enough, but we omit such results.

### 2.3.3. Asymptotic detection rates for vanishing signals

So far we always considered a constant signal, in this section we focus on the detection of vanishing changes when the number of observations increases. We start with the detection of a single vanishing bump against a noisy background.

**Theorem 12** (Single vanishing bump). *Assume the heterogeneous Gaussian change-point model (2.1) with sequences of bump signals  $\mu_n(t) := m_0 + \delta_n \mathbb{1}_{I_n}(t)$  and  $\sigma_n(t) := \mathbb{1}_{I_n^C}(t) + s_n \mathbb{1}_{I_n}(t)$ , where  $\delta_n \neq 0$  is a sequence of change-point sizes,  $s_n > 0$  a sequence of standard deviations on  $I_n \in \mathcal{D}$ , which is a sequence of intervals with  $|I_n| \rightarrow 0$ . Let  $k_n := \lfloor \log_2(n|I_n|) \rfloor$  and  $\Delta_n := |\delta_n|/s_n$  be the sequence of the signal to noise ratios. Let further  $(\hat{K}_n)_n$  be the sequence of the estimated numbers of change-points by H-SMUCE (2.18), with significance levels  $\alpha_n$  and weights  $\beta_{1,n}, \dots, \beta_{d_n,n}$ . We further assume*

$$\sqrt{n|I_n|}\Delta_n \geq (4 + \epsilon_n)\sqrt{-\log(|I_n|)}, \quad (2.25)$$

with possibly  $\epsilon_n \rightarrow 0$ , but such that  $\epsilon_n\sqrt{-\log(|I_n|)} \rightarrow \infty$  and

$$\limsup_{n \rightarrow \infty} \frac{\sqrt{-\log(\alpha_n \beta_{k_n,n})}}{\epsilon_n \sqrt{-\log(|I_n|)}} < \frac{1}{4},$$

$$\liminf_{n \rightarrow \infty} \frac{n|I_n|}{\log(n)} > 64 \text{ and } \lim_{n \rightarrow \infty} \frac{\log(\alpha_n \beta_{k_n,n})}{n|I_n|} = 0, \quad (2.26)$$

$$\lim_{n \rightarrow \infty} s_n \frac{\sqrt{|I_n^C|}}{\sqrt{|I_n|}} = \infty \text{ and} \quad (2.27)$$

$$\liminf_{n \rightarrow \infty} \frac{\log(\beta_{k_n,n})}{\log(\beta_{\min,n})} > 0, \text{ with } \beta_{\min,n} := \min\{\beta_{1,n}, \dots, \beta_{d_n,n}\}. \quad (2.28)$$

Then,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{(\mu_n, \sigma_n^2)}(\hat{K}_n > 0) = 1. \quad (2.29)$$

Conditions (2.25) and (2.26) are the main assumptions of the theorem to detect the vanishing signal on  $I_n$ . We will discuss them together with the conditions of Theorem 13. We also need the weak technical conditions (2.27) and (2.28) on the size of  $|I_n^C|$  and the minimal weight  $\beta_{\min,n}$  to ensure that the detection power on the complement  $I_n^C$  is large enough, too. Condition (2.28) is for instance fulfilled by uniform weights  $\beta_{1,n} = \dots = \beta_{d_n,n} = 1/d_n$ , but many other choices are possible, too. We further assumed  $I_n \in \mathcal{D}$ , otherwise we have to replace  $I_n$  by the largest subinterval which is an element of the dyadic partition. Such an interval exists always and has at least length  $n^{-1}2^{\lfloor \log_2(n|I_n|/2) \rfloor} > |I_n|/4$ . Therefore,

omitting the condition  $I_n \in \mathcal{D}$  would not change the rate. It is possible to strengthen (2.29) further to  $\lim_{n \rightarrow \infty} \mathbb{P}_{(\mu_n, \sigma_n^2)}(\hat{K}_n \geq K) = 1$  if we increase all constants a little bit. We now move to the detection of a signal with several vanishing change-points.

**Theorem 13** (Multiple vanishing change-points). *Assume the heterogeneous Gaussian change-point model (2.1). Let  $(K_n)_n := (|\mathcal{J}(\mu_n)|)_n$  be the sequence of true number of change-points. Let further  $(\hat{K}_n)_n$ ,  $\alpha_n$  and  $\beta_{1,n}, \dots, \beta_{d_n,n}$  be as in Theorem 12. Let  $\mathcal{S}_{\Delta_n, \lambda_n} \subset \mathcal{S}$  be a sequence of submodels as in (2.19) and  $k_n := \lfloor \log_2(n\lambda_n/4) \rfloor$ . We further assume*

$$\liminf_{n \rightarrow \infty} \frac{n\lambda_n}{\log(n)} > 512 \text{ and } \lim_{n \rightarrow \infty} \frac{\log(\alpha_n \beta_{k_n, n})}{n\lambda_n} = 0 \quad (2.30)$$

as well as

- (1) for large scales, i.e.,  $\liminf_{n > 0} \lambda_n > 0$ , the limit  $n\lambda_n \Delta_n^2 \log(1/(\alpha_n \beta_{k_n, n}))^{-1} \rightarrow \infty$ ,
- (2) for small scales, i.e.,  $\lambda_n \rightarrow 0$ , the inequality

$$\sqrt{n\lambda_n} \Delta_n \geq \left( \sqrt{512} + C + \epsilon_n \right) \sqrt{-\log(\lambda_n)} \quad (2.31)$$

with possibly  $\epsilon_n \rightarrow 0$ , but such that  $\epsilon_n \sqrt{-\log(\lambda_n)} \rightarrow \infty$  and

$$\limsup_{n \rightarrow \infty} \frac{\sqrt{\log(8/(\alpha_n \beta_{k_n, n}))}}{\epsilon_n \sqrt{-\log(\lambda_n)}} < \frac{1}{\sqrt{512}},$$

with  $C = 0$  for  $K_n$  bounded and  $C = 16\sqrt{6}$  for  $K_n$  unbounded.

Then,

$$\lim_{n \rightarrow \infty} \sup_{(\mu_n, \sigma_n^2) \in \mathcal{S}_{\Delta_n, \lambda_n}} \mathbb{P}_{(\mu_n, \sigma_n^2)}(\hat{K}_n < K_n) = 0.$$

Theorems 12 and 13 state conditions on the tuning parameters  $\alpha_n$  and  $\beta_{k_n, n}$  as well as on the length of the minimal scale  $|I_n| =: \lambda_n$  (to simplify notations we only write  $\lambda_n$  in the following) and the standardized jump size  $\Delta_n$  to detect the vanishing signals uniformly over  $\mathcal{S}_{\Delta_n, \lambda_n}$ . If, in addition,  $\lim_{n \rightarrow \infty} \alpha_n = 0$  holds, then we control also the probability to overestimate the number of change-points and therefore the estimation of the number of change-points is still consistent in the case of a vanishing signal. The main condition in both theorems is that  $\sqrt{n\lambda_n} \Delta_n$  has to be at least of order  $\sqrt{-\log(\lambda_n)}$ , see (2.25) and (2.31). This is optimal in the sense that no signal with a smaller rate can be detected asymptotically with probability one, see (Dümbgen and Spokoiny, 2001; Chan and Walther, 2013; Frick et al., 2014a) for the case of homogeneous observations, and note that this is a sub-model of our model. But different to the homogeneous case we need, in addition, that  $\lambda_n$  is at least of order  $\log(n)/n$ , see (2.26) and (2.30). Such a restriction appears reasonable, since for the additional variance estimation only the number of observation on the segment is relevant and not the size of the change. Finally, we observe that the constants encountered in the lower detection bound for H-SMUCE

in (2.25) and (2.31) increase with the difficulty of the estimation problem, where the difficulty is represented by the number of vanishing segments. All of these constants are a little bit larger as the analogue constants for SMUCE in (Frick et al., 2014a, Theorem 5 and 6) reflecting the additional difficulty encountered by the heterogeneous noise. More precisely, we have 4 instead of the optimal  $\sqrt{2}$  for one vanishing segment,  $\sqrt{512}$  instead of 4 for a bounded number of vanishing segments and  $\sqrt{512} + 16\sqrt{6}$  instead of 12 for an unbounded number of vanishing segments. Note again, that the optimal constants for the heterogeneous case are unknown to us.

### 2.3.4. Choice of the tuning parameters

In this section we discuss the choice of the tuning parameters  $\alpha$  and  $\beta_1, \dots, \beta_{d_n}$ .

**Choice of  $\alpha$**  The choice depends on the application. If a strict overestimation control of the number of change-points  $K$  is desirable,  $\alpha$  should be chosen small, e.g., 0.05 or 0.1, recall Theorems 5 and 6. This might come at the expense of missing change-points but with large probability not detecting too many, recall Figure 1.2 and see also the simulations in Section 2.4. If change-point screening is the primarily goal, i.e., we aim to avoid missing of change-points,  $\alpha$  should be increased, e.g.,  $\alpha = 0.5$  or even higher, since Theorem 7 shows that the error probability to underestimate the number of change-points decreases with increasing  $\alpha$ . If model selection, i.e.,  $\hat{K} = K$ , is the major aim, an intermediate level that balances the over- and underestimation error should be chosen, e.g.  $\alpha$  between 0.1 and 0.5. Both errors vanish super polynomially for the asymptotic choice  $\alpha = \alpha_n \in \exp(-o(n))$ , see Remark 8. A finite sample approach is to weight these error probabilities  $\gamma\mathbb{P}(\hat{K} > K) + (1 - \gamma)\mathbb{P}(\hat{K} < K)$ , with  $\gamma \in (0, 1)$ , and to choose  $\alpha$  such that its upper bound

$$\gamma\alpha + (1 - \gamma) \left( 1 - \left[ 1 - 3 \exp \left( -\frac{1}{48} \left( \sqrt{\frac{n\lambda\Delta^2}{32}} - \sqrt{16 \log \left( \frac{8}{\lambda\alpha\beta_k} \right)} \right)_+^2 \right) \right]_+^{2K} \right)$$

is minimized. This also allows to incorporate prior information on  $(\lambda, \Delta)$ . Alternatively, the bound on the expectation  $\mathbb{E}[|\hat{K} - K|]$  by combining Theorems 6 and 7 can be minimized to take the size of the misestimation into account. Despite of all possibilities to choose the "best"  $\alpha$  for a given application, comparing estimates at different  $\alpha$  can be helpful to trace the "stability of evidence" of the estimated change-points at different significance levels. Of course, the interpretation of such a "significance screening" does not allow for a frequentist interpretation of a significance level anymore as  $\alpha$  has to be fixed in advance, see for instance (Schervish, 1996). Nevertheless, it might give for instance some indication whether to perform further experiments. Despite of this, for a fixed  $\alpha$  the confidence statements of H-SMUCE can also be used to support findings by other estimators.

**Choice of  $\beta_1, \dots, \beta_{d_n}$**  As a default choice we recommend equal weights  $\beta_1 = \dots = \beta_{d_n} = 1/d_n$ . This choice fulfils, together with many other choices, the conditions of Theorems 9 and 10. Unlike as for the significance level  $\alpha$  only the bound for the underestimation of the number of change-points depends on these weights. Note, that this gives the user the possibility to incorporate prior information on the scales without violating the overestimation control in Theorems 5 and 6. If for instance changes are expected to occur only on small segments then the detection power on these scales can be increased if the first weights are chosen large and the other ones small (or even zero). In contrast, if the general signal to noise ratio is expected to be very small, then it is nearly impossible to detect changes on small scales and larger scales should be weighted more to detect at least the changes on these scales. A quantitative influence of the weights on the detection power can be seen in the underestimation bound in Theorem 19 which is a refinement of Theorem 7. We also investigate such choices quantitatively in simulations in Section 2.4.2.

## 2.4. Simulations

In this section we compare H-SMUCE in simulations with CBS (Venkatraman and Olshen, 2007), cumSeg (Muggeo and Adelfio, 2011) and LOOVF (Arlot and Celisse, 2011) as they are also designed to be robust against heterogeneous noise. Moreover, we include SMUCE (Frick et al., 2014a) in simulations with a constant variance as a benchmark to examine how much the detection power of H-SMUCE decreases in this case, which may be regarded as the price for adaptation to heterogeneous noise. In most of the simulations we fix the weights  $\beta_1, \dots, \beta_{d_n} = 1/d_n$  and vary the significance level  $\alpha$ . A simulation with tuned weights can be found in Section 2.4.2. The results for H-SMUCE are obtained by the  $R$  function *stepFit.hsmuce* in the *HSMUCE* package (Pein, 2016), but up to small numerical errors the same results will be obtained by the *stepR* package using the function *stepFit*. For circular binary segmentation (CBS) we call the function *segmentByCBS*, <http://cran.r-project.org/web/packages/PSCBS/>, version 0.40.4, 2014-02-04, with the standard parameters. For the cross-validation method LOOVF we use the Matlab function *proc.LOOVF*, <http://www.di.ens.fr/~arlot/code/CHPTCV.htm>, version 1.0, 2010-10-27, with the parameter choice of the demo file. For cumSeg we call the method *jumpoints*, <http://cran.r-project.org/web/packages/cumSeg/>, version 1.1, 2011-10-14, with the parameter  $k$  large enough such that the estimation is not influenced by this choice. For SMUCE we call the function *smuceR*, <http://cran.r-project.org/web/packages/stepR>, version 1.0-3, 2015-06-18, with the standard parameters, in particular the interval set of all intervals is used if  $n \leq 1000$ .

To avoid specific interactions between the signal and the dyadic partition we generate in each repetition a random pair  $(\mu_R, \sigma_R^2) \in \mathcal{S}$ . All random variables are independent from each other.

- (a) We fix the number of observations  $n$ , the number of change-points  $K$ , a constant  $C$  and a minimum value for the smallest scale  $\lambda_{\min}$ .

- (b) We draw the locations of the change-points  $\tau_0 := 0 < \tau_1 < \dots < \tau_K < 1 =: \tau_{K+1}$  uniformly distributed with the restriction that  $\lambda := \min_{k=0, \dots, K} |\tau_{k+1} - \tau_k| \geq \lambda_{\min}$ .
- (c) We choose the function values  $s_0, \dots, s_K$  of the standard deviation function  $\sigma_R$  by  $s_k := 2^{U_k}$ , where  $U_0, \dots, U_K$  are uniform distributed on  $[-2, 2]$ .
- (d) We determine the function values  $m_0, \dots, m_K$  of the signal  $\mu_R$  such that

$$|m_k - m_{k-1}| = \sqrt{\frac{C}{n} \min \left( \frac{\tau_{k+1} - \tau_k}{s_k^2}, \frac{\tau_k - \tau_{k-1}}{s_{k-1}^2} \right)^{-1}} \quad \forall k = 1, \dots, K. \quad (2.32)$$

Thereby, we start with  $m_0 = 0$  and choose randomly with probability  $1/2$  whether the expectation increases or decreases.

By (2.32) we provide a situation where all change-points are similarly hard to find, recall the minimax detection boundary from Section 2.3.3. An example has been displayed in Figure 1.2 in the introduction, where H-SMUCE misses at  $\alpha = 0.1$  one change-point and detects for  $\alpha$  between 0.15 and 0.99 (only displayed for  $\alpha = 0.5$ ) the correct number of change-points. In Figure 2.2 we see that CBS (Venkatraman and Olshen, 2007) finds also all change-points, but detects false positives. Less good is the performance of cumSeg (Muggeo and Adelfio, 2011) and LOOVF (Arlot and Celisse, 2011) which both miss several changes and LOOVF adds also a false positive. We compare these methods now more extensively. All simulations are repeated 10 000 times.

In the following we report the difference between the estimated  $\hat{K}$  and the true number  $K$  of change-points as well as the mean of the absolute value of this difference. Additionally, we use the false positive sensitive location error

$$\text{FPSLE} = \frac{n}{2\hat{K}} \sum_{k=1}^{\hat{K}+1} |\tau_{l_{k-1}} - \hat{\tau}_{k-1}| + |\tau_{l_k} - \hat{\tau}_k|,$$

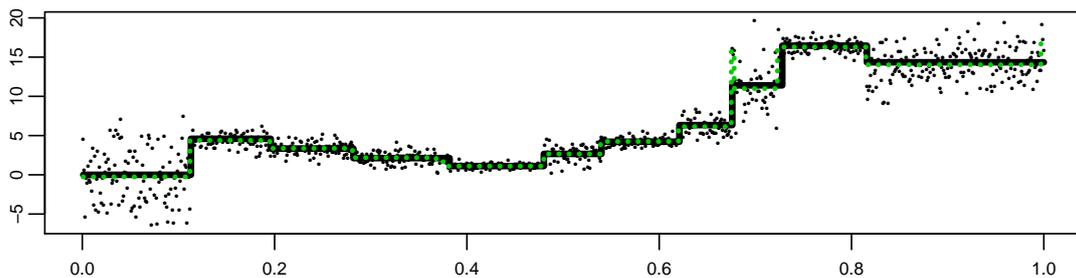
with  $l_k \in \{1, \dots, K+1\}$  such that  $(\hat{\tau}_{k-1} + \hat{\tau}_k)/2 \in (\tau_{l_{k-1}}, \tau_{l_k}]$ , i.e., the left and right neighbouring change-points to the middle point of  $(\hat{\tau}_{k-1}, \hat{\tau}_k]$ , and the false negative sensitive location error

$$\text{FNSLE} = \frac{n}{2\hat{K}} \sum_{k=1}^{K+1} |\tau_{k-1} - \hat{\tau}_{l_{k-1}}| + |\tau_k - \hat{\tau}_{l_k}|,$$

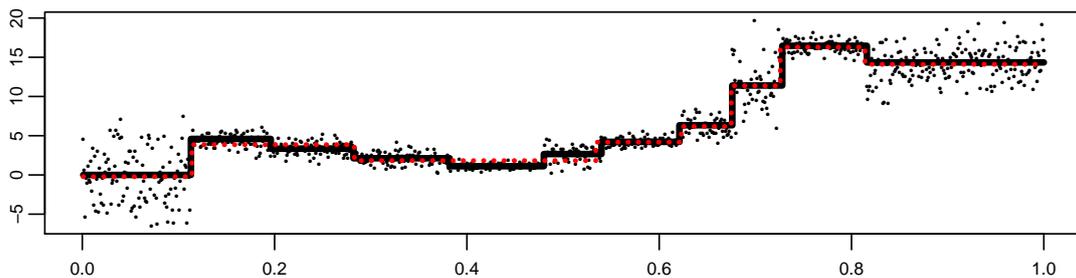
with  $l_k \in \{1, \dots, \hat{K} + 1\}$  such that  $(\tau_{k-1} + \tau_k)/2 \in (\hat{\tau}_{l_{k-1}}, \hat{\tau}_{l_k}]$ , see (Futschik et al., 2014, Section 3.1), to rate the estimation of the locations of the change-points. We also show the mean integrated squared (absolute) error MISE (MIAE) for all methods.

### 2.4.1. Simulation results

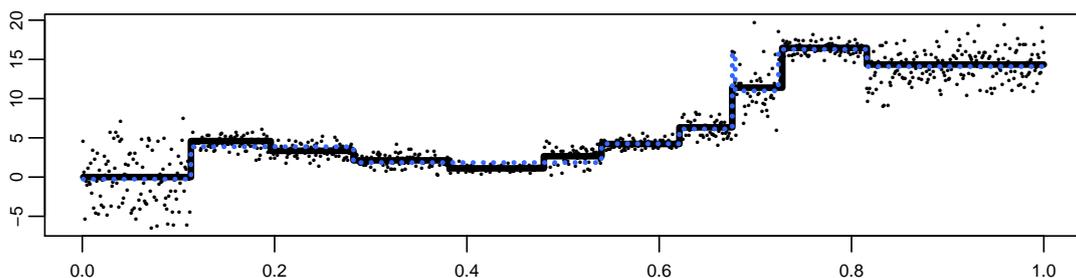
In this section we discuss the results of the simulations assuming model (2.1) and (2.2). In Table 2.1 we consider a constant variance and in Table 2.2 heterogeneous errors. We



(a) CBS.



(b) cumSeg.



(c) LOOVF.

Figure 2.2.: Observations (black points) and true signal (black line) together with estimates by CBS, cumSeg and LOOVF for the data from Figure 1.2. All parameters are chosen as described in Section 2.4.

excluded LOOVF from simulations for larger  $n$  due to its large computation time, for more details on the run times see (Pein et al., 2017c, Figure 8 in the Supplement).

All simulations confirm the overestimation control  $\alpha$  for H-SMUCE from Theorem 5 and the exponential decay of the overestimation in Theorem 6. For a constant signal, corresponding to  $K = 0$  in Table 2.1, H-SMUCE overestimates the number of change-points even slightly less than SMUCE, whereas CBS and cumSeg overestimate hardly ever. In the case of a constant variance we found that the detection power of H-SMUCE is only slightly worse than SMUCE for  $K = 2$ , although SMUCE used instead of the dyadic partition  $\mathcal{D}$  the system of all intervals. The difference is larger for  $K = 10$  and in this case also CBS and cumSeg performs better than H-SMUCE, since the detection power of H-SMUCE depends strongly on the lengths of the constant segments. Moreover,  $\lambda_{\min}$

Table 2.1.: Simulations with constant variance and  $C = 200$ . The columns from left to right give the setting, the method, the proportions of  $\hat{K} - K$  and the means of the corresponding error criteria.

Setting	Method	$\leq -2$	$-1$	$0$	$+1$	$\geq +2$	$ \hat{K} - K $	FPSLE	FNSLE	MISE	MIAE
n = 1000, K = 0, $\mu = \mu_R \equiv 0$ , $\sigma = \sigma_R$ $\equiv \text{const}$	H-SMUCE, $\alpha = 0.1$	-	-	0.965	0.035	0.000	0.035	17.75	4.73	0.0035	0.0365
	H-SMUCE, $\alpha = 0.3$	-	-	0.867	0.128	0.005	0.138	68.95	18.42	0.0045	0.0401
	H-SMUCE, $\alpha = 0.5$	-	-	0.719	0.256	0.025	0.307	153.45	41.25	0.0061	0.0454
	SMUCE, $\alpha = 0.1$	-	-	0.965	0.034	0.001	0.036	17.90	5.03	0.0039	0.0371
	SMUCE, $\alpha = 0.3$	-	-	0.832	0.160	0.008	0.177	88.45	24.80	0.0059	0.0435
	SMUCE, $\alpha = 0.5$	-	-	0.667	0.298	0.035	0.370	184.90	50.94	0.0082	0.0499
	CBS	-	-	0.991	0.000	0.009	0.018	8.90	1.26	0.0037	0.0351
cumSeg	-	-	0.999	0.001	0.000	0.001	0.30	0.06	0.0029	0.0345	
n = 1000, K = 2, $\lambda_{\min} = 30$ , $\mu = \mu_R$ , $\sigma \equiv 1$	H-SMUCE, $\alpha = 0.1$	0.010	0.174	0.802	0.014	0.000	0.208	26.32	72.66	0.0132	0.0613
	H-SMUCE, $\alpha = 0.3$	0.004	0.108	0.819	0.067	0.002	0.187	38.10	52.90	0.0114	0.0571
	H-SMUCE, $\alpha = 0.5$	0.002	0.070	0.768	0.150	0.010	0.244	64.14	48.50	0.0111	0.0573
	SMUCE, $\alpha = 0.1$	0.003	0.074	0.912	0.011	0.000	0.092	16.96	34.03	0.0092	0.0513
	SMUCE, $\alpha = 0.3$	0.001	0.040	0.892	0.065	0.002	0.112	32.24	27.39	0.0090	0.0513
	SMUCE, $\alpha = 0.5$	0.001	0.025	0.806	0.155	0.013	0.209	63.30	32.33	0.0095	0.0536
	CBS	0.005	0.060	0.821	0.082	0.033	0.221	37.55	37.57	0.0111	0.0527
cumSeg	0.025	0.116	0.749	0.099	0.011	0.289	65.32	82.63	0.0364	0.0738	
n = 1000, K = 2, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma \equiv 1$	H-SMUCE, $\alpha = 0.1$	0.009	0.160	0.815	0.015	0.000	0.194	27.14	68.91	0.0127	0.0611
	H-SMUCE, $\alpha = 0.3$	0.004	0.098	0.829	0.067	0.001	0.176	37.77	49.63	0.0111	0.0572
	H-SMUCE, $\alpha = 0.5$	0.002	0.063	0.774	0.152	0.009	0.237	63.46	46.06	0.0109	0.0573
	SMUCE, $\alpha = 0.1$	0.003	0.068	0.919	0.009	0.000	0.084	16.82	31.94	0.0091	0.0515
	SMUCE, $\alpha = 0.3$	0.001	0.035	0.899	0.063	0.002	0.104	31.19	25.81	0.0090	0.0515
	SMUCE, $\alpha = 0.5$	0.001	0.020	0.819	0.147	0.013	0.195	59.86	30.23	0.0095	0.0537
	CBS	0.005	0.058	0.824	0.083	0.031	0.215	37.50	36.27	0.0112	0.0532
cumSeg	0.023	0.110	0.769	0.090	0.008	0.262	59.74	79.25	0.0336	0.0741	
n = 1000, K = 10, $\lambda_{\min} = 30$ , $\mu = \mu_R$ , $\sigma \equiv 1$	H-SMUCE, $\alpha = 0.1$	0.508	0.330	0.161	0.001	0.000	1.634	54.37	172.66	0.1112	0.1842
	H-SMUCE, $\alpha = 0.3$	0.354	0.377	0.263	0.006	0.000	1.233	44.53	127.81	0.0817	0.1561
	H-SMUCE, $\alpha = 0.5$	0.253	0.384	0.346	0.017	0.000	0.987	40.88	102.88	0.0679	0.1419
	SMUCE, $\alpha = 0.1$	0.163	0.352	0.485	0.001	0.000	0.721	29.14	77.49	0.0424	0.1193
	SMUCE, $\alpha = 0.3$	0.093	0.301	0.598	0.007	0.000	0.513	24.23	56.17	0.0366	0.1099
	SMUCE, $\alpha = 0.5$	0.062	0.258	0.657	0.022	0.001	0.415	23.34	46.37	0.0342	0.1060
	CBS	0.033	0.129	0.531	0.204	0.102	0.644	42.69	45.08	0.0417	0.1078
cumSeg	0.163	0.216	0.403	0.165	0.053	0.904	65.16	105.59	0.1107	0.1492	
n = 1000, K = 10, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma \equiv 1$	H-SMUCE, $\alpha = 0.1$	0.445	0.356	0.198	0.001	0.000	1.474	59.32	162.03	0.0913	0.1801
	H-SMUCE, $\alpha = 0.3$	0.303	0.384	0.307	0.005	0.000	1.104	47.34	120.10	0.0682	0.1532
	H-SMUCE, $\alpha = 0.5$	0.213	0.379	0.390	0.018	0.001	0.881	41.98	96.70	0.0577	0.1398
	SMUCE, $\alpha = 0.1$	0.155	0.351	0.494	0.000	0.000	0.697	32.51	77.29	0.0426	0.1235
	SMUCE, $\alpha = 0.3$	0.085	0.299	0.612	0.004	0.000	0.485	26.14	55.78	0.0368	0.1131
	SMUCE, $\alpha = 0.5$	0.054	0.252	0.680	0.014	0.000	0.381	23.81	45.39	0.0344	0.1086
	CBS	0.027	0.135	0.524	0.203	0.111	0.653	45.64	44.88	0.0425	0.1116
cumSeg	0.165	0.217	0.389	0.179	0.050	0.904	63.73	104.37	0.1037	0.1522	

plays a similar role as the number of change-points  $K$ , since the average constant segments length decreases if  $\lambda_{\min}$  decreases or  $K$  increases. Worse results for smaller lengths are due to the familywise error control  $\alpha$  of H-SMUCE as it guarantees a strict control of overestimating the number of change-points.

Similar results can be observed for  $n = 100$  with heterogeneous errors. CBS performs better than cumSeg and LOOVF, and in particular better than in the single change-point setting. CBS outperforms H-SMUCE for  $K = 5$ , although H-SMUCE has a much smaller tendency to overestimate the number of change-points, whereas in particular CBS and LOOVF tend to overestimation. This can also be seen for the MISE and MIAE as these measures are much more affected by underestimation than by overestimation. These findings are also supported by the FPSLE and the FNSLE, the FPSLE is heavily affected by overestimation, whereas the FNSLE is larger in case of underestimation.

Table 2.2.: Simulations with heterogeneous errors and  $C = 200$ . The columns from left to right give the setting, the method, the proportions of  $\hat{K} - K$  and the means of the corresponding error criteria.

Setting	Method	$\leq -2$	-1	0	+1	$\geq +2$	$ \hat{K} - K $	FPSLE	FNSLE	MISE	MIAE
n = 100, K = 2, $\lambda_{\min} = 15$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.000	0.125	0.873	0.002	0.000	0.128	1.51	4.07	0.8182	0.3308
	H-SMUCE, $\alpha = 0.3$	0.000	0.042	0.945	0.013	0.000	0.055	1.04	1.70	0.4217	0.2482
	H-SMUCE, $\alpha = 0.5$	0.000	0.016	0.940	0.043	0.000	0.060	1.63	1.26	0.2776	0.2291
	CBS	0.000	0.001	0.925	0.058	0.016	0.092	2.03	0.79	0.2220	0.2143
	cumSeg	0.000	0.066	0.720	0.167	0.047	0.343	6.50	4.39	0.4898	0.3053
	LOOVF	0.000	0.031	0.700	0.163	0.106	0.683	12.83	3.36	0.3167	0.2639
n = 100, K = 5, $\lambda_{\min} = 15$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.608	0.364	0.028	0.000	0.000	1.610	13.51	32.33	9.5104	1.8626
	H-SMUCE, $\alpha = 0.3$	0.212	0.577	0.211	0.000	0.000	1.003	8.63	19.80	6.5362	1.3263
	H-SMUCE, $\alpha = 0.5$	0.061	0.466	0.473	0.001	0.000	0.588	5.27	11.65	3.9992	0.9047
	CBS	0.001	0.008	0.884	0.089	0.018	0.137	1.65	1.02	0.4539	0.3130
	cumSeg	0.098	0.230	0.544	0.117	0.012	0.588	6.93	12.13	1.2454	0.5441
	LOOVF	0.031	0.112	0.520	0.152	0.184	1.648	14.61	6.92	0.5887	0.4042
n = 1000, K = 2, $\lambda_{\min} = 30$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.000	0.007	0.974	0.018	0.000	0.026	8.42	5.83	0.0195	0.0617
	H-SMUCE, $\alpha = 0.3$	0.000	0.001	0.921	0.075	0.002	0.080	24.72	9.57	0.0193	0.0636
	H-SMUCE, $\alpha = 0.5$	0.000	0.000	0.827	0.162	0.012	0.185	53.23	17.09	0.0204	0.0668
	CBS	0.005	0.019	0.774	0.146	0.056	0.298	52.95	21.17	0.0347	0.0711
	cumSeg	0.022	0.161	0.683	0.103	0.030	0.387	64.04	92.66	0.0765	0.1112
	LOOVF	0.000	0.002	0.982	0.017	0.000	0.018	7.25	4.35	0.0182	0.0630
n = 1000, K = 2, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.000	0.000	0.926	0.071	0.002	0.076	22.64	8.49	0.0196	0.0657
	H-SMUCE, $\alpha = 0.3$	0.000	0.000	0.830	0.160	0.010	0.181	50.02	16.22	0.0214	0.0692
	H-SMUCE, $\alpha = 0.5$	0.003	0.011	0.776	0.153	0.057	0.296	53.69	15.85	0.0355	0.0730
	CBS	0.016	0.155	0.699	0.098	0.031	0.370	60.63	84.69	0.0739	0.1132
	cumSeg	0.000	0.002	0.982	0.017	0.000	0.018	7.25	4.35	0.0182	0.0630
	LOOVF	0.000	0.000	0.926	0.071	0.002	0.076	22.64	8.49	0.0196	0.0657
n = 1000, K = 10, $\lambda_{\min} = 30$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.123	0.429	0.446	0.002	0.000	0.686	22.83	55.06	0.4045	0.2402
	H-SMUCE, $\alpha = 0.3$	0.016	0.199	0.770	0.015	0.000	0.245	11.98	21.12	0.1863	0.1618
	H-SMUCE, $\alpha = 0.5$	0.002	0.088	0.863	0.045	0.001	0.140	11.84	12.71	0.1220	0.1404
	CBS	0.002	0.008	0.463	0.316	0.211	0.843	47.26	15.20	0.1274	0.1435
	cumSeg	0.439	0.243	0.187	0.085	0.046	1.674	94.91	228.44	0.3120	0.2806
	LOOVF	0.000	0.002	0.982	0.017	0.000	0.018	7.25	4.35	0.0182	0.0630
n = 1000, K = 10, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.025	0.262	0.711	0.002	0.000	0.315	16.94	32.39	0.2102	0.1866
	H-SMUCE, $\alpha = 0.3$	0.002	0.058	0.925	0.015	0.000	0.076	8.46	10.58	0.1009	0.1372
	H-SMUCE, $\alpha = 0.5$	0.000	0.017	0.940	0.043	0.001	0.061	9.03	7.72	0.0860	0.1307
	CBS	0.001	0.007	0.451	0.319	0.222	0.868	47.81	15.10	0.1293	0.1463
	cumSeg	0.433	0.254	0.197	0.082	0.035	1.601	97.00	223.47	0.2771	0.2794
	LOOVF	0.000	0.004	0.983	0.013	0.000	0.017	50.65	30.94	0.0016	0.0183
n = 10000, K = 2, $\lambda_{\min} = 30$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.000	0.002	0.936	0.061	0.001	0.065	188.73	63.72	0.0016	0.0188
	H-SMUCE, $\alpha = 0.3$	0.000	0.001	0.865	0.128	0.006	0.142	407.41	125.46	0.0016	0.0197
	H-SMUCE, $\alpha = 0.5$	0.012	0.036	0.532	0.200	0.220	0.886	1548.96	373.22	0.0057	0.0235
	CBS	0.054	0.245	0.600	0.084	0.017	0.477	682.64	1457.08	0.0090	0.0379
	cumSeg	0.000	0.001	0.984	0.015	0.000	0.016	53.23	24.89	0.0014	0.0182
	LOOVF	0.000	0.000	0.941	0.057	0.002	0.060	181.06	59.83	0.0014	0.0188
n = 10000, K = 2, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.000	0.000	0.870	0.124	0.007	0.137	394.16	115.62	0.0016	0.0197
	H-SMUCE, $\alpha = 0.3$	0.000	0.000	0.870	0.124	0.007	0.137	394.16	115.62	0.0016	0.0197
	H-SMUCE, $\alpha = 0.5$	0.012	0.035	0.521	0.208	0.225	0.917	1601.54	366.42	0.0058	0.0238
	CBS	0.052	0.241	0.603	0.087	0.016	0.473	673.81	1430.47	0.0084	0.0377
	cumSeg	0.023	0.231	0.741	0.005	0.000	0.282	58.42	165.72	0.0178	0.0431
	LOOVF	0.006	0.123	0.844	0.027	0.000	0.162	68.27	98.25	0.0122	0.0385
n = 10000, K = 10, $\lambda_{\min} = 30$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.006	0.123	0.844	0.027	0.000	0.162	68.27	98.25	0.0122	0.0385
	H-SMUCE, $\alpha = 0.3$	0.003	0.079	0.854	0.064	0.002	0.151	108.19	87.63	0.0103	0.0377
	H-SMUCE, $\alpha = 0.5$	0.024	0.043	0.180	0.222	0.531	2.088	1286.59	525.95	0.0198	0.0475
	CBS	0.619	0.169	0.130	0.059	0.024	2.345	1000.55	3122.28	0.0433	0.0917
	cumSeg	0.009	0.165	0.819	0.007	0.000	0.190	59.11	124.05	0.0132	0.0418
	LOOVF	0.001	0.064	0.905	0.029	0.001	0.097	67.32	65.54	0.0089	0.0375
n = 10000, K = 10, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.000	0.029	0.900	0.067	0.003	0.102	103.42	60.04	0.0078	0.0368
	H-SMUCE, $\alpha = 0.3$	0.019	0.034	0.162	0.228	0.557	2.203	1317.31	467.47	0.0198	0.0475
	H-SMUCE, $\alpha = 0.5$	0.607	0.188	0.131	0.051	0.023	2.277	997.64	3105.88	0.0405	0.0925
	CBS	0.609	0.284	0.107	0.001	0.000	1.908	155.65	504.02	0.1016	0.1031
	cumSeg	0.278	0.399	0.318	0.006	0.000	1.044	94.53	263.30	0.0640	0.0789
	LOOVF	0.140	0.371	0.470	0.019	0.000	0.696	84.07	182.54	0.0483	0.0703
n = 10000, K = 25, $\lambda_{\min} = 30$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.015	0.024	0.069	0.128	0.765	3.348	921.91	409.98	0.0411	0.0723
	H-SMUCE, $\alpha = 0.3$	0.934	0.036	0.018	0.009	0.003	6.028	1043.82	3488.43	0.1159	0.1540
	H-SMUCE, $\alpha = 0.5$	0.396	0.383	0.220	0.001	0.000	1.334	146.74	387.66	0.0699	0.0945
	CBS	0.103	0.359	0.528	0.010	0.000	0.591	85.33	175.03	0.0390	0.0715
	cumSeg	0.038	0.241	0.690	0.030	0.001	0.352	78.74	114.01	0.0291	0.0647
	LOOVF	0.010	0.017	0.055	0.120	0.799	3.529	934.29	346.33	0.0405	0.0726
n = 10000, K = 25, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	CBS	0.934	0.036	0.019	0.008	0.003	5.849	1053.35	3462.62	0.1022	0.1547
	cumSeg	0.934	0.036	0.019	0.008	0.003	5.849	1053.35	3462.62	0.1022	0.1547

In all simulations with heterogeneous errors and 1 000 observations H-SMUCE outperforms the other methods, for 10 000 observations this becomes even more pronounced. In comparison to the simulation with 100 observations the tendency of CBS to overestimate the number of change-points becomes then also more prominent. Finally, in further simulations (not displayed) we found that the detection power of all methods decreases for smaller  $C$  in (2.32), but all results remain qualitatively the same. All in all, we found that H-SMUCE performs well as sample size becomes larger, in particular if the constant segments are not too short as indicated by assumption (2.30) in Theorem 13.

## 2.4.2. Prior information on scales

To demonstrate the effect of incorporating prior knowledge about those scales where change-points are likely to happen we consider again the observations from Table 2.2 with  $n = 10\,000$ ,  $K = 10$  and  $\lambda_{\min} = 50$ . To this end, we use the adapted weights, where we eliminate the smallest three scales  $k = 1, 2, 3$ , since all constant segments contain at least 50 observations and therefore these small scales are not needed for detection. Moreover, we choose  $\tilde{\beta}_4 = 1/4$ ,  $\tilde{\beta}_5 = 1/4$ ,  $\tilde{\beta}_6 = 1/6$ ,  $\tilde{\beta}_7 = 1/6$ ,  $\tilde{\beta}_8 = 1/12$ ,  $\tilde{\beta}_9 = 1/12$  in decreasing order, since change-points on smaller scales are more likely and harder to detect. For the same reasons we eliminate the four largest scales  $k = 10, 11, 12, 13$ , too.

Table 2.3.:  $n = 10\,000$  observations,  $K = 10$  change-points,  $C = 200$  and  $\lambda_{\min} = 50$  from Table 2.2. H-SMUCE uses the weights  $\tilde{\beta}_4, \dots, \tilde{\beta}_9$ . The columns from left to right give the method, the proportions of  $\hat{K} - K$  and the means of the corresponding error criteria.

Method	$\leq -2$	-1	0	+1	$\geq +2$	$ \hat{K} - K $	FPSLE	FNSLE	MISE	MIAE
H-SMUCE, $\alpha = 0.1$	0.005	0.117	0.876	0.002	0.000	0.130	50.82	113.50	0.0107	0.0406
H-SMUCE, $\alpha = 0.3$	0.000	0.032	0.952	0.016	0.000	0.049	48.39	49.84	0.0075	0.0368
H-SMUCE, $\alpha = 0.5$	0.000	0.013	0.940	0.045	0.001	0.061	78.86	48.19	0.0072	0.0368

A comparison of Table 2.2 and 2.3 shows that the modified weights increase the detection power of H-SMUCE for all significance levels, so we encourage the user to adapt the weights if prior information on the scales where changes occur is available.

## 2.4.3. Robustness against model violations

We begin by investigating how robust the methods are against a violation of the assumption that the standard deviation changes only at the same locations as the mean changes. Such violations occur in many applications and it is often impossible to model these changes precisely. We consider continuous changes as well as abrupt changes. Figure 2.3 shows the used standard deviation functions in Table 2.4 to examine robustness against variance changes on constant segments. We consider with the sinus-shaped standard deviation  $\sigma_1(t) := 1 + 0.5 \sin(20\pi t)$  continuous changes, with the piecewise linear standard deviation  $\sigma_2(t) := 0.5 + \sum_{i=0}^9 (10t - i) \mathbb{1}_{(0.1i, 0.1(i+1)]}(t)$  continuous and abrupt changes at the same time and with the piecewise constant standard deviation  $\sigma_3(t) := \sum_{i=1}^{n/200} 0.5 \mathbb{1}_{(200(i-1)/n, 200(i-1)/n+100/n]}(t) + \mathbb{1}_{(200(i-1)/n+100/n, 200i/n]}(t)$  abrupt changes.

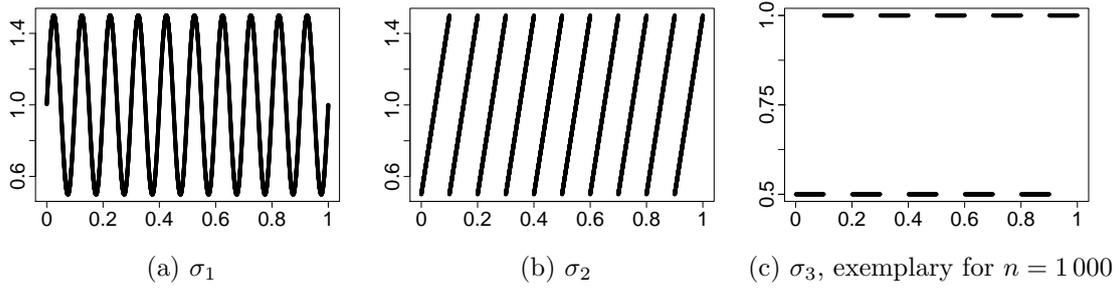


Figure 2.3.: The standard deviation functions used in Table 2.4.

In Table 2.4 we see that H-SMUCE and CBS perform very robust against heterogeneous noise on the constant segments, whereas, remarkably, the detection power of cumSeg is even improved. Moreover, in additional simulations (not displayed) with less observations we found that LOOVF is very robust, too.

Furthermore, we investigate robustness against heavy tails of the error distribution. In Table 2.5 we consider  $t_3$ -distributed errors which are scaled such that the expectation and the standard deviation are the same as in Section 2.4.1.

As expected SMUCE is not robust against heavy tails, whereas H-SMUCE provides reasonable results. In comparison to Gaussian errors H-SMUCE is not influenced for  $K = 0$ ,

Table 2.4.: Simulations with standard deviations  $\sigma_1(\cdot)$ - $\sigma_3(\cdot)$  from Figure 2.3 and  $C = 200$ . The columns from left to right give the setting, the method, the proportions of  $\hat{K} - K$  and the means of the corresponding error criteria.

Setting	Method	$\leq -2$	$-1$	$0$	$+1$	$\geq +2$	$ \hat{K} - K $	FPSLE	FNSLE	MISE	MIAE
n = 1000, K = 0, $\mu = \mu_R \equiv 0$ , $\sigma = \sigma_1$	H-SMUCE, $\alpha = 0.1$	-	-	0.968	0.032	0.000	0.033	16.30	4.12	0.0013	0.0277
	H-SMUCE, $\alpha = 0.3$	-	-	0.876	0.118	0.005	0.129	64.60	15.91	0.0018	0.0306
	H-SMUCE, $\alpha = 0.5$	-	-	0.734	0.239	0.027	0.293	146.75	36.45	0.0023	0.0338
	CBS	-	-	0.916	0.001	0.083	0.186	93.25	11.21	0.0045	0.0288
	cumSeg	-	-	1.000	0.000	0.000	0.000	0.20	0.04	0.0011	0.0264
n = 1000, K = 0, $\mu = \mu_R \equiv 0$ , $\sigma = \sigma_2$	H-SMUCE, $\alpha = 0.1$	-	-	0.968	0.031	0.001	0.032	16.10	4.12	0.0013	0.0278
	H-SMUCE, $\alpha = 0.3$	-	-	0.876	0.118	0.005	0.129	64.55	15.73	0.0017	0.0306
	H-SMUCE, $\alpha = 0.5$	-	-	0.734	0.241	0.024	0.292	145.80	35.28	0.0022	0.0340
	CBS	-	-	0.937	0.004	0.060	0.135	67.70	8.96	0.0034	0.0281
	cumSeg	-	-	0.999	0.001	0.000	0.001	0.40	0.12	0.0011	0.0264
n = 1000, K = 0, $\mu = \mu_R \equiv 0$ , $\sigma = \sigma_3$	H-SMUCE, $\alpha = 0.1$	-	-	0.969	0.030	0.001	0.032	15.75	3.91	0.0007	0.0210
	H-SMUCE, $\alpha = 0.3$	-	-	0.875	0.119	0.006	0.130	65.10	16.31	0.0009	0.0227
	H-SMUCE, $\alpha = 0.5$	-	-	0.737	0.236	0.026	0.290	145.15	36.09	0.0012	0.0250
	CBS	-	-	0.937	0.002	0.061	0.134	67.10	8.64	0.0019	0.0213
	cumSeg	-	-	0.999	0.001	0.000	0.001	0.35	0.10	0.0006	0.0199
n = 10000, K = 10, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma = \sigma_1$	H-SMUCE, $\alpha = 0.1$	0.013	0.185	0.796	0.005	0.000	0.218	661.16	755.83	0.0212	0.0684
	H-SMUCE, $\alpha = 0.3$	0.003	0.076	0.890	0.031	0.001	0.113	543.91	548.21	0.0167	0.0585
	H-SMUCE, $\alpha = 0.5$	0.001	0.041	0.886	0.069	0.003	0.117	513.55	468.37	0.0147	0.0542
	CBS	0.000	0.001	0.191	0.155	0.653	2.636	1590.35	276.51	0.0092	0.0358
	cumSeg	0.206	0.118	0.413	0.193	0.070	0.984	790.10	1054.73	0.0146	0.0502
n = 10000, K = 10, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma = \sigma_2$	H-SMUCE, $\alpha = 0.1$	0.014	0.205	0.776	0.006	0.000	0.238	421.19	513.32	0.0156	0.0556
	H-SMUCE, $\alpha = 0.3$	0.001	0.077	0.894	0.027	0.001	0.108	348.50	358.14	0.0119	0.0475
	H-SMUCE, $\alpha = 0.5$	0.000	0.038	0.897	0.062	0.002	0.105	344.93	311.35	0.0106	0.0446
	CBS	0.000	0.000	0.215	0.174	0.611	2.362	1454.85	247.26	0.0085	0.0346
	cumSeg	0.114	0.102	0.467	0.236	0.082	0.795	756.12	720.95	0.0136	0.0478
n = 10000, K = 10, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma = \sigma_3$	H-SMUCE, $\alpha = 0.1$	0.019	0.233	0.744	0.004	0.000	0.276	161.27	251.06	0.0053	0.0301
	H-SMUCE, $\alpha = 0.3$	0.002	0.069	0.904	0.025	0.000	0.099	137.86	136.79	0.0036	0.0254
	H-SMUCE, $\alpha = 0.5$	0.000	0.029	0.906	0.062	0.003	0.096	170.29	128.56	0.0033	0.0248
	CBS	0.000	0.000	0.246	0.173	0.582	2.189	1134.85	214.71	0.0047	0.0263
	cumSeg	0.054	0.051	0.516	0.279	0.101	0.669	749.33	499.10	0.0070	0.0346

Table 2.5.: Simulations with  $t_3$  distributed errors and  $C = 200$ . The columns from left to right give the setting, the method, the proportions of  $\hat{K} - K$  and the means of the corresponding error criteria.

Setting	Method	$\leq -2$	-1	0	+1	$\geq +2$	$ \hat{K} - K $	FPSLE	FNSLE	MISE	MIAE
n = 1000, K = 0, $\mu = \mu_R \equiv 0$ , $\sigma = \sigma_R$ ,	H-SMUCE, $\alpha = 0.1$	-	-	0.982	0.018	0.000	0.018	9.05	2.53	0.0031	0.0347
	H-SMUCE, $\alpha = 0.3$	-	-	0.927	0.071	0.001	0.074	37.05	10.31	0.0034	0.0361
	H-SMUCE, $\alpha = 0.5$	-	-	0.824	0.167	0.009	0.185	92.40	26.35	0.0043	0.0392
	SMUCE, $\alpha = 0.1$	-	-	0.001	0.001	0.999	11.859	5929.70	369.55	0.8710	0.1491
	SMUCE, $\alpha = 0.3$	-	-	0.000	0.000	1.000	14.803	7401.65	397.77	0.9338	0.1674
	SMUCE, $\alpha = 0.5$	-	-	0.000	0.000	1.000	16.862	8431.00	411.30	0.9730	0.1787
	CBS	-	-	0.991	0.000	0.009	0.018	9.05	1.13	0.0058	0.0340
cumSeg	-	-	0.955	0.001	0.044	0.188	93.90	11.98	0.0682	0.0375	
n = 1000, K = 2, $\lambda_{\min} = 30$ , $\mu = \mu_R$ , $\sigma \equiv 1$ ,	H-SMUCE, $\alpha = 0.1$	0.008	0.136	0.848	0.007	0.000	0.160	25.70	62.95	0.0120	0.0578
	H-SMUCE, $\alpha = 0.3$	0.003	0.086	0.876	0.035	0.000	0.127	29.74	44.61	0.0103	0.0537
	H-SMUCE, $\alpha = 0.5$	0.001	0.055	0.851	0.090	0.003	0.152	44.21	38.62	0.0097	0.0524
	SMUCE, $\alpha = 0.1$	0.000	0.000	0.001	0.001	0.998	11.104	2683.40	250.21	0.3046	0.1232
	SMUCE, $\alpha = 0.3$	0.000	0.000	0.000	0.000	1.000	13.984	3361.80	283.17	0.3264	0.1340
	SMUCE, $\alpha = 0.5$	0.000	0.000	0.000	0.000	1.000	15.991	3836.28	302.43	0.3400	0.1419
	CBS	0.053	0.161	0.726	0.043	0.018	0.346	46.69	119.74	0.0241	0.0712
cumSeg	0.025	0.097	0.722	0.093	0.063	0.456	108.11	81.86	0.0557	0.0707	
n = 10000, K = 10, $\lambda_{\min} = 50$ , $\mu = \mu_R$ , $\sigma = \sigma_R$	H-SMUCE, $\alpha = 0.1$	0.002	0.079	0.916	0.004	0.000	0.086	93.09	119.69	0.0130	0.0425
	H-SMUCE, $\alpha = 0.3$	0.000	0.025	0.957	0.017	0.000	0.043	86.32	81.78	0.0105	0.0397
	H-SMUCE, $\alpha = 0.5$	0.000	0.012	0.950	0.038	0.000	0.050	99.93	76.24	0.0097	0.0389
	CBS	0.467	0.148	0.167	0.107	0.111	2.516	1356.25	6254.20	0.0877	0.1308
	cumSeg	0.586	0.192	0.136	0.055	0.032	2.242	997.13	3005.71	0.0433	0.0906

underestimation is more distinct in the constant variance scenario and detection power is even increased in the scenario with heterogeneous errors. In comparison, CBS is not influenced for  $K = 0$ , too, underestimates and overestimates in the constant variance scenario and is slightly worse with a tendency to underestimation in the scenario with heterogeneous errors, whereas cumSeg overestimates rarely, but heavily for  $K = 0$ , underestimates and overestimates in the constant variance scenario and is robust in the last scenario.

In summary, H-SMUCE appear to be robust against a wide range of variance changes on constant segments and to be only slightly affected by larger tails than Gaussian, in particular no tendency to overestimation was visible in our simulations. This may be explained by the fact that the local likelihood tests of H-SMUCE are quite robust against heterogeneous noise, see for instance (Bakirov and Szekely, 2006; Ibragimov and Müller, 2010), and against non-normal errors, see (Lehmann and Romano, 2005) and the references therein. Unlike the number of change-points, the locations are sometimes miss-estimated, since the restricted maximum likelihood estimator is influenced by changes of the variance. Instead, more robust estimators, for instance local median and MAD estimators, could be used.

Real data examples were shown in Figure 3.4c and in (Pein et al., 2017c, Section 5) by applying H-SMUCE to subsampled PorB recordings. They confirm the results of the simulations. H-SMUCE works well on larger scales and does not tend to include false positives, while it does not have a good detection power on smaller scales due to the strong familywise error control. Hence, in the next section we avoid subsampling of the ion channel recordings and propose more sophisticated approaches to deal with the filtering. At the price that no systematic theory is provided, since similar results to them in Section 2.3 would be very technical or even impossible.

## 3. Ion Channel Recordings

In this section we extend the multiscale regression estimator (1.2) to models with filtered observations. We start by modeling the recordings in Section 3.1, firstly with homogeneous noise and secondly with heterogeneous noise. Afterwards, we discuss in Section 3.2 the desired properties of an estimator for the signal underlying the ion channel recordings. In Section 3.3 we describe briefly the computational difficulties arising by the filtering. Nonetheless, we present three different methods for estimating the underlying signal: J-SMURF in Section 3.4, JULES in Section 3.5 and JILTAD in Section 3.7. The used deconvolution techniques are explained in Section 3.6. After a discussion of how these estimations can be used for analyzing ion channel recordings in Section 3.8, we compare these methods with existing non-parametric estimation methods in an extensive simulation study in Section 3.9. This section ends with an application of these methods to Gramicidin A recordings to study their behavior on real data, too. For the sake of clarity, we explain the general ideas for homogeneous noise and afterwards adaptations for heterogeneous noise if required. And note that in the following we will use statistical instead of electrophysiology terms. We say for instance non-parametric estimation instead of model-free idealization.

### 3.1. Model

The homogeneous ion channel model was already introduced in (Hotz et al., 2013; Pein et al., 2017d). Nevertheless, we start in Section 3.1.1 with a precise definition of this model and illustrate afterwards the influence of the convolution on the signal for a better intuition. Finally, in Section 3.1.2 we extend this model to heterogeneous noise.

#### 3.1.1. Homogeneous ion channel model

Gating events of an ion channel occur usually on much smaller times scales, 1 ns–100 ns, than the sampled observations are recorded, 1 kHz–100 kHz rates. Hence, channel recordings have the appearance of abrupt random changes (Hamill et al., 1981). Consequently, the conductance of a channel is modeled by a piecewise constant signal

$$\mu(t) = \sum_{k=0}^K m_k \mathbb{1}_{[\tau_k, \tau_{k+1})}(t), \quad (3.1)$$

where  $t$  from now on denotes physical time. The (unknown) conductance values are denoted by  $m_0, \dots, m_K$ , its (unknown) number of conductivity changes by  $K$  and the (unknown) change-point locations by  $-\infty =: \tau_0 < 0 < \tau_1 < \dots < \tau_K < \tau_{K+1} := \tau_{\text{end}}$ . Here, we denote the time when recording ends by  $\tau_{\text{end}}$ . Moreover, we only consider change-points after recording started at time 0 and assume for simplicity that the signal before recordings started is constant and equal to the first function value. When truncating the filter we will see that this is a less severe restriction than it seems for the moment, since only a very short time period before will be relevant. Nonetheless, for an analysis of how long the channel dwells in the same conductance level we exclude always the first and last segment. We stress that this class of signals

$$\mathcal{M} := \left\{ \mu : [-\infty, \tau_{\text{end}}] \mapsto \mathbb{R}, \mu(t) = \sum_{k=0}^K m_k \mathbb{1}_{[\tau_k, \tau_{k+1})}(t), K \in \mathbb{N} \right\}, \quad (3.2)$$

with slight abuse of notation by redefining  $\mathcal{M}$  from (2.4), is very flexible as potentially an arbitrarily number of change-points and function values can be imposed, see Figure 1.3 for an estimation in this class. In comparison to hidden Markov models, we do not assume any prior knowledge on the function values and on the distribution of change-point locations. Allowing a new function value for every segment enables a reasonable estimation even in the presence of artifacts, for instance in ion channel recordings base line fluctuations are common. Nonetheless, an extension of our methods to the assumption of a known, small alphabet of possible functions values is briefly discussed in Section 6.2.2.

The very small conductance of a single channel, typically in the range of picosiemens up to few nanosiemens, requires sophisticated electronic recordings devices, including one or several amplifiers (Molecular Devices, 2008). To stay in the transmission range of the amplifier, high frequent noise components, e.g., caused by shot noise, are attenuated by convolving the recordings with an analogue lowpass filter. Typically, a four, six or eight pole lowpass Bessel filter is integrated in the hardware of the technical measurement device. Finally, the recorded currents are digitized equidistantly with sample rate  $f_s$  and divided by the applied constant voltage. Thus, the recorded observations  $Y_1, \dots, Y_n$  are the measured conductivity at equidistant time points  $t_i = i/f_s$ ,  $i = 1, \dots, n$ , with sampling rate  $f_s = n/\tau_{\text{end}}$ . And we assume that this time series results from convolving the signal  $\mu$  perturbed by standard Gaussian white noise  $\eta$  with an analogue lowpass filter, with kernel  $F$ , also called impulse response, and digitization at sampling rate  $f_s$ . Expressed as a formula,

$$Y_i = (F * (\mu + \sigma_0 \eta))(i/f_s) = (F * \mu)(i/f_s) + \epsilon_i, \quad i = 1, \dots, n, \quad (3.3)$$

with noise level  $\sigma_0 > 0$  and convolution operator  $(f * g)(t) = \int_{-\infty}^{\infty} f(t-s)g(s)ds$ . Here,  $n$  denotes the total number of observations, typically several hundred thousands up to few millions. And the white noise  $\eta$  is specified such that  $W_I := \int_I \eta(t)dt$  is a Gaussian distributed random variable with expectation zero and variance  $|I|$  for every interval  $I \subset \mathbb{R}$

of finite length and  $\mathbb{E}[W_I W_J] = |I \cap J|$  for all intervals  $I, J \subset \mathbb{R}$ . For a precise definition and a proof of existence see (Giné and Nickl, 2016, Example 2.1.11). Hence, the errors  $\epsilon_1, \dots, \epsilon_n$  are Gaussian and centered,  $\mathbb{E}[\epsilon_i] = 0$ , and have covariance  $\text{Cov}[\epsilon_i, \epsilon_{i+j}] = \sigma_0^2 (F * F)(j/f_s)$ . All data sets we analyze later are filtered with a 4-pole lowpass Bessel filter with normalized cutoff frequency of 0.1, which equals the cutoff frequency in time divided by the sampling rate. We denote the kernel by  $F$  and its antiderivative by  $\mathcal{F}$ , also called step function. In Figure 3.1 we show these two functions together with the resulting autocorrelation function of this filter for sampling rate 1, i.e., with time domain equal to the sampling points. Moreover, Figure 3.2 shows the convolution of the filter with a signal containing a single change-point and a signal containing an isolated peak, computed by a truncation of the filter as explained now.

Typically used lowpass filters have kernels with support the whole positive real numbers, also called filters with infinite impulse response, but a quickly decaying kernel  $F$ . For instance the kernels of Bessel filters decay exponentially fast (Proakis and Manolakis, 1996). Thus, we truncate the kernel  $F$  at some point, say  $L/f_s$ , which provides further simplification of our model. The truncated kernel is denoted by  $F_L$  and rescaled such that  $\int_{-\infty}^{\infty} F_L = \int_0^{L/f_s} F_L = 1$ , see (Hotz et al., 2013). Its antiderivative is denoted by  $\mathcal{F}_L$ . As a working rule, we choose  $L$  such that the autocorrelation function  $(F * F)(\cdot)/(F * F)(0)$  of the analogue lowpass filter is below  $10^{-3}$  at  $L/f_s$  and afterwards. For a 4-pole Bessel filter with normalized cutoff frequency 0.1 this choice leads to  $L = 11$  and is confirmed visually by comparing the measured observations with the convolution of the estimated signal with the truncated kernel, see for instance the lower panels in Figures 1.3, 3.22 and 3.23. We also truncate the autocorrelation function at  $L/f_s$ , resulting in an  $L$ -dependent process. All in all, model (3.3) reduces to

$$Y_i = (F_L * \mu)(i/f_s) + \epsilon_i, \quad i = 1, \dots, n, \quad (3.4)$$

where  $\epsilon_1, \dots, \epsilon_n$  are centered, Gaussian distributed random variables with covariance

$$\text{Cov}[\epsilon_i, \epsilon_{i+j}] = \begin{cases} \sigma_0^2 (F * F)(j/f_s), & \text{for } |j| = 0, \dots, L, \\ 0, & \text{for } |j| > L. \end{cases} \quad (3.5)$$

Note that we truncate the autocorrelation function itself as this leads to more accurate values for lags close to  $L$  instead of convolving the truncated kernel with itself. Untruncated and truncated kernel, their antiderivatives and untruncated and truncated autocorrelation functions are compared in Figure 3.1. All filter functions are implemented in the R function *lowpassFilter* in the package *clampSeg* (Pein et al., 2017b). We stress that all integrals are computed analytically and exactly.

Figure 3.1 confirms that for all three functions the errors by truncation are very small, almost invisible. Moreover, the convolution of a piecewise constant function (3.1) with the

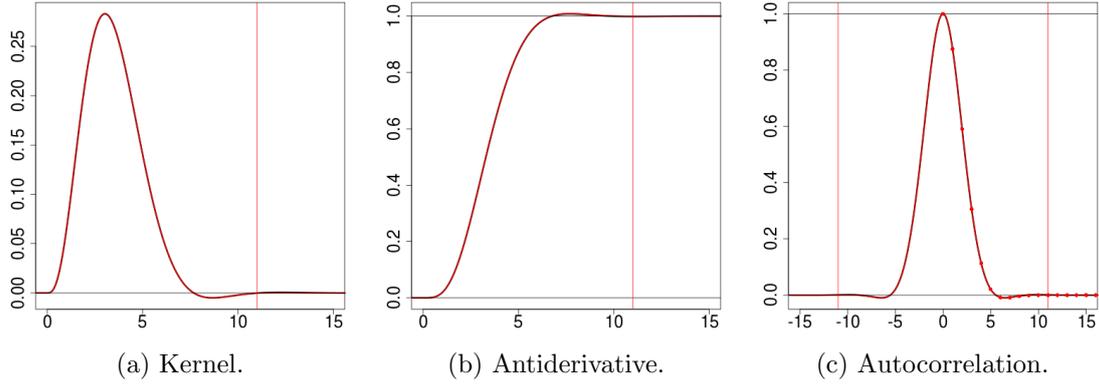


Figure 3.1.: Untruncated (black line) and truncated (red line) kernel function, antiderivative and autocorrelation function of a four-pole Bessel lowpass filter with normalized cutoff frequency 0.1 and sampling rate 1. The filter length  $L = 11$  is indicated by vertical red lines and the autocorrelation values at positive integer time lags by red points.

truncated kernel is given by

$$(F_L * \mu)(t) = \sum_{k=0}^K m_k \int_{t-\tau_{k+1}}^{t-\tau_k} F_L(s) ds = \sum_{k=0}^K m_k [\mathcal{F}_L(t - \tau_k) - \mathcal{F}_L(t - \tau_{k+1})]. \quad (3.6)$$

Note that most of the differences are zero, since  $\mathcal{F}_L(t) = 0$  if  $t \leq 0$  and  $\mathcal{F}_L(t) = 1$  if  $t \geq L/f_s$ . Hence, the convolution can be computed exactly and efficiently. For a better understanding of how the filter acts on the signal and hence on the expectation of the observations we show in Figure 3.2 the convolution for signals with a single change-point and with a single peak, a short one as well as one longer than the filter length. If the signal contains only one change-point,

$$\mu(t) = \begin{cases} m_0 & \text{if } t < \tau_1, \\ m_1 & \text{if } t \geq \tau_1, \end{cases} \quad (3.7)$$

the convolution is given by

$$\begin{aligned} (F_L * \mu)(t) &= m_0 [1 - \mathcal{F}_L(t - \tau_1)] + m_1 \mathcal{F}_L(t - \tau_1) \\ &= \begin{cases} m_0 & \text{if } t \leq \tau_1, \\ m_0 [1 - \mathcal{F}_L(t - \tau_1)] + m_1 \mathcal{F}_L(t - \tau_1) & \text{if } \tau_1 < t < \tau_1 + L/f_s, \\ m_1 & \text{if } t \geq \tau_1 + L/f_s. \end{cases} \end{aligned} \quad (3.8)$$

For a peak,

$$\mu(t) = \begin{cases} m_0 & \text{if } t < \tau_1, \\ m_1 & \text{if } \tau_1 \leq t < \tau_2, \\ m_2 & \text{if } t \geq \tau_2, \end{cases} \quad (3.9)$$

the convolution reads as

$$(F_L * \mu)(t) = m_0[1 - \mathcal{F}_L(t - \tau_1)] + m_1[\mathcal{F}_L(t - \tau_1) - \mathcal{F}_L(t - \tau_2)] + m_2\mathcal{F}_L(t - \tau_2). \quad (3.10)$$

This is equal to  $m_0$  if  $t \leq \tau_1$ ,  $m_1$  if  $t \in [\tau_1 + L/f_s, \tau_2]$  and  $m_2$  if  $t \geq \tau_2 + L/f_s$ .

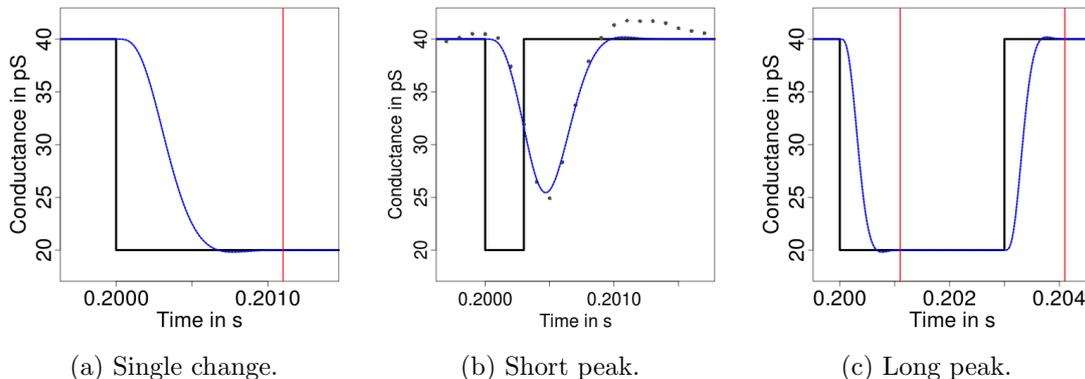


Figure 3.2.: Signals (black line) containing one change-point, an short peak and a longer peak and their convolutions (blue line) with a four-pole lowpass Bessel filter with normalized cutoff frequency of 0.1 and sampling rate  $10^4$ . Vertical red lines indicate the change-point locations plus the filter length  $L/f_s$ . The functions values, the change-point locations and the filter are representative of the later simulations and Gramicidin A recordings.

We highlight two conclusions which will be very important in the following sections. First of all, the convolution is constant and identical to the signal  $\mu$  on the interval  $[\tau_k + L/f_s, \tau_{k+1}]$ , which might be empty, but changes continuously on  $(\tau_k, \tau_k + L/f_s)$ . Secondly, in Figure 3.2b the convolution does not reach the function value  $m_1 = 20$  pS. This is the general case for peaks shorter than the filter length  $L/f_s$ . Hence, deconvolution techniques are necessary to estimate all function values correctly.

In a real data example the constant variance  $\sigma_0^2$  is also unknown, but can be pre-estimated  $\sqrt{n}$ -consistently by difference based type estimators, see (Tecuapetla-Gómez and Munk, 2017) and the references therein. We follow the suggestion of Hotz et al. (2013) and use

$$\hat{\sigma}_0 = \frac{\text{IQR}(Y_{L+1} - Y_1, \dots, Y_n - Y_{n-L})}{2\Phi^{-1}(0.75)\sqrt{2(F * F)(0)}}, \quad (3.11)$$

with  $\Phi^{-1}$  the quantile function of the standard normal distribution. This estimator is consistent and usually slightly overestimates the variance for finitely many observations depending on how many and how huge the changes in the signal are. An implementation is given by the R function *sdrobnorm* in the *stepR* package (Pein et al., 2017a). Hence, in the following sections the global variance  $\sigma_0^2$  is assumed to be known.

### 3.1.2. Heterogeneous ion channel model

We now model additional *open channel noise* which occur in some measurements. This phenomenon describes measurements with larger noise on segments with a larger conduc-

tance. Its name refers to the fact that a larger conductance results from an open pore. It is caused by interference of the channel proteins with the membrane (Sigworth, 1985, 1986; Sigworth et al., 1987; Heinemann and Sigworth, 1988, 1990, 1991) and occurs for some larger channels for some membrane systems, but it is hard to predict in advance whether it will occur or not. For instance, in the PorB measurements it did not occur when black lipid membranes were used, see Figure 1.1, but occurred when solvent-free membranes were used, see Figure 3.3.

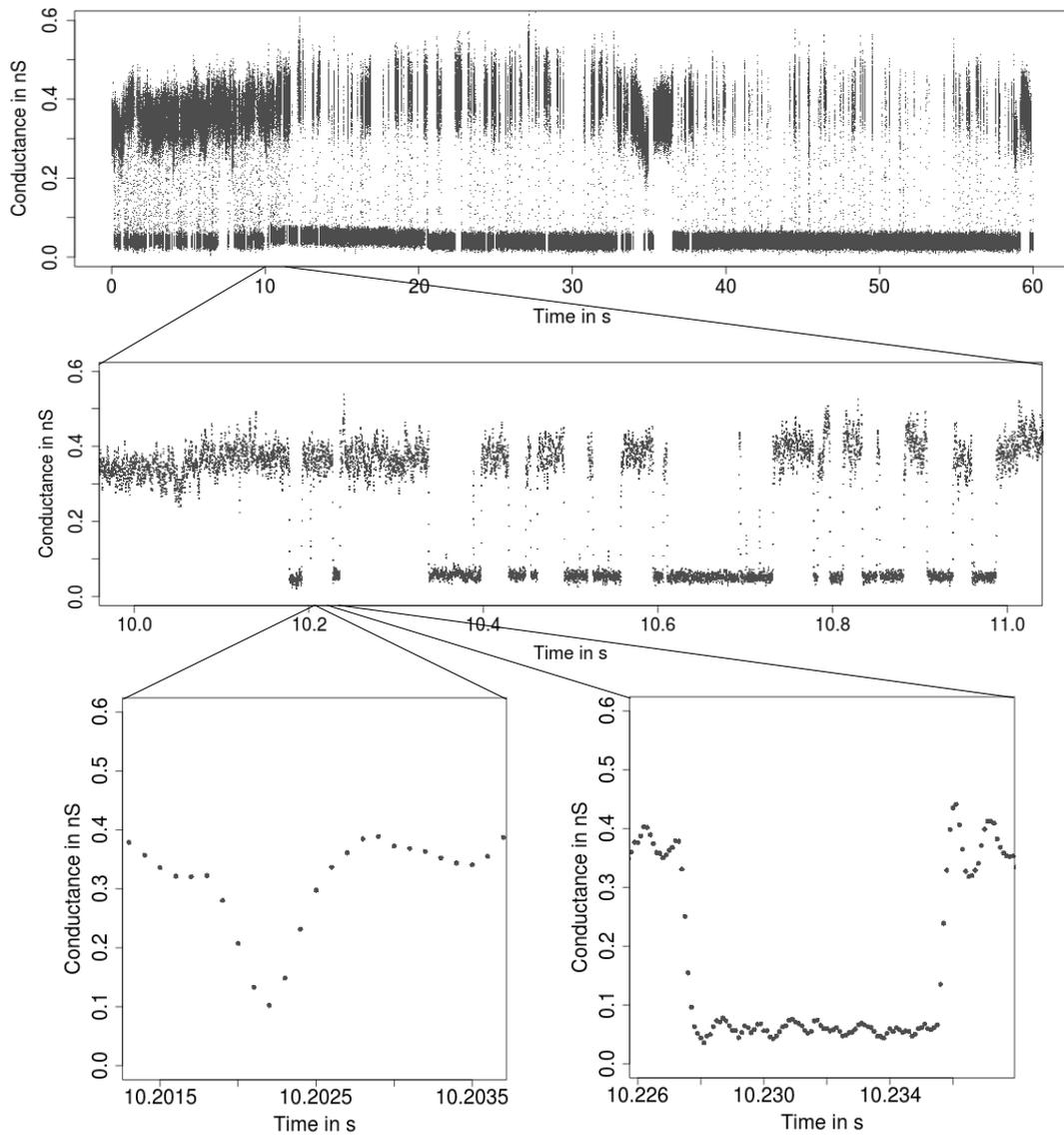


Figure 3.3.: Observations (grey points) of a representative conductance time series for PorB mutant G103K recorded by the patch clamp technique using solvent-free bilayers at 20 mV.

The membrane influences the channel heavily and, hence, the use of a different membrane system often affects the noise levels, but also the gating dynamics and even the conductance levels. For more details how the use of solvent-free and black lipid membranes affected

the recordings with PorB see Section 5. Thus, using a different membrane system is not always possible. Moreover, extracting the channel protein, inserting it in a membrane and establishing the recordings is usually a work of several months up to years. Hence, heterogeneous noise is not always avoidable and has to be included in the model.

Since the noise level is larger when the conductance is larger, we assume as in Section 2 that at the locations where the signal changes also the variance is allowed to change, but does not have to. In more detail, we still assume a signal of form (3.1). But in (3.3) the white noise is now scaled by a piecewise constant standard deviation function with potential change-points at the same locations as the signal

$$\sigma(t) = \sum_{k=0}^K s_k \mathbb{1}_{[\tau_k, \tau_{k+1})}(t), \quad (3.12)$$

$s_0, \dots, s_K > 0$ , instead by a global standard deviation  $\sigma_0$ . In other words, we assume a signal-variance pair  $(\mu, \sigma^2)$  from the set

$$\mathcal{S} := \left\{ (\mu, \sigma^2) : [-\infty, \tau_{\text{end}}] \mapsto \mathbb{R}^2, \mu = \sum_{k=0}^K m_k \mathbb{1}_{[\tau_k, \tau_{k+1})}, \sigma^2 = \sum_{k=0}^K s_k^2 \mathbb{1}_{[\tau_k, \tau_{k+1})}, K \in \mathbb{N} \right\}. \quad (3.13)$$

Note that this redefines  $\mathcal{S}$  from (2.2). We truncate the kernel and autocorrelation function of the lowpass filter in the same way as we did for the homogeneous model. Hence, the resulting observations still have expectation  $(F_L * \mu)(i/f_s)$ , but the covariance structure is now

$$\text{Cov}[Y_i, Y_{i+j}] = \begin{cases} \sum_{k=0}^K s_k^2 [\mathcal{A}(i/f_s - \tau_k, j/f_s) - \mathcal{A}(i/f_s - \tau_{k+1}, j/f_s)] & \text{for } |j| = 0, \dots, L, \\ 0 & \text{for } |j| > L, \end{cases} \quad (3.14)$$

with

$$\mathcal{A}(t, l) := \int_0^t F(s)F(s + j/f_s)ds. \quad (3.15)$$

A proof of (3.14) is given by Proposition 22. An analytic solution of  $\mathcal{A}(t, l)$  is implemented in the *R* function *lowpassFilter*. Hence, also (3.14) can be computed exactly and efficiently. The major aim will be now to estimate the unknown signal  $\mu$  taking into account the convolution, the specific structure of  $\mu$  in (3.1), and if the noise is heterogeneous the structure of  $\sigma$  in (3.12). Before we will present the estimation methods, we discuss in the next section their desired properties.

## 3.2. Requirements on the estimation

The estimation should be objective. This means no subjective choices by the experimenter should be necessary during the estimation process, only few tuning parameters might be chosen *prior* to the estimation. Although, this point might read self-evidently for many statisticians, a subjective and time-consuming analysis by visual inspection is still com-

mon in electrophysiology studies. Often, the event times are chosen manually or in a semi-automatic way, as offered by standard recording software like pCLAMP 10 (Molecular Devices). To maximize objectivity the amount of false positive detections should be controlled statistically.

To be universally applicable, assumptions on the signal should be kept as mild as possible. In particular, contrary to hidden Markov models, no parametric model should be assumed for the gating dynamics. We stress that hidden Markov models are with justification well established for analyzing ion channel recordings, see (Ball and Rice, 1992; Venkataraman et al., 2000; Qin et al., 2000; de Gunst et al., 2001; Siekmann et al., 2011) among many others. Their strong assumptions allow detection of events with a power that can not be reached by non-parametric approaches, in particular not on small scales. Moreover, their compact results, typically only the conductance levels and a transition or a Q-matrix have to be estimated, can be interpreted well and often directly connected to biochemical processes.

In the same spirit are approaches based on the conductance (current) distribution (Yellen, 1984; Heinemann and Sigworth, 1991; Schroeder, 2015). They do not provide an estimation of the underlying signal, but estimate the parameter of the underlying Markov model (or a similar assumed parametric model). Thereto, the smoothing effect of the filter is used as a feature, since the recorded observations between the underlying conductance levels provide information about the rates of a Markov model. These approaches can resolve the gating dynamics on very small temporal scales, in particular still when events follow each other quickly, which is in such a temporal resolution not possible by hidden Markov or non-parametric estimation. On the other hand, the number of unknown parameters has to be small to obtain good results.

However, hidden Markov models and fitting the conductance distribution require to pre-determine a model for the gating dynamics, in particular to fix the number of states. Although some model selection approaches exist, this is not an easy task. The Markovian property is also not always valid, cf. (Fuliński et al., 1998; Mercik and Weron, 2001; Goychuk et al., 2005; Shelley et al., 2010). Moreover, in the presence of artifacts the assumption of few conductance levels that occur repeatedly leads to heavy misestimations or require elaborate data cleaning, while in many of these situations non-parametric methods provide still a reasonable estimation. We suggest to compare both approaches whenever possible to verify results. We did this for instance for the PorB recordings with ampicillin in (Bartsch et al., 2017). Additionally, non-parametric methods are often much faster, an important point we discuss below. Finally, even when a Markov approach is preferred, non-parametric methods, in addition to model selection approaches, can be used before to select and verify the specific Markov model, in particular the number of states and possible transitions, and to explore and potentially remove artifacts. Furthermore, fitting the conductance distribution requires often an initial guess for the parameters which can be provided by non-parametric methods.

As mentioned before the biggest disadvantage of non-parametric approaches is that they

miss events that are short in time. However, for estimating the rates of a Markov model this is not always an issue, since missing events can be taken into account, see Section 3.8. For instance, the exact distribution of the observed times when events below a resolution limit are missed is calculated by Hawkes et al. (1990) and estimators for the rates of a Markov model that are corrected for missing events are provided by Qin et al. (1996); Colquhoun et al. (1996); Epstein et al. (2016). All in all, hidden Markov models, fitting the conductance distribution, and non-parametric estimation complement each other well and should not be seen as competing approaches.

While the literature of hidden Markov models and fitting the conductance distribution is comprehensive, in particular also from a theoretical perspective, we see a demand for better working non-parametric approaches. For an overview about existing methods and their limitations see the following paragraphs, the simulations in Section 3.9 and their application to Gramicidin A measurements in Section 3.10. Hence, we will focus in this work on non-parametric approaches.

Since a typical recorded time series consists of several hundred thousands up to few million observations and a full study requires often the analysis of more than hundred time series, the estimation should be fast, desirably almost linear in the number of observations. For instance in the PorB analysis in Section 5, where each time series contained three million observations and in each time series on average more than thousands events were found, in total more than a billion observations had to be analyzed. This also shows again the demand for fully automatic estimation methods.

While events with a large conductance difference (in comparison to the noise level) on longer time scales are easy to detect by almost every approach, even a fixed threshold is often enough, two common phenomena called *subconductivity* and *flickering* require more evolved estimation methods. Subconductivity means changes between two conductance values close to each other, while flickering describes events on very small time scales, often only few observations long. Detection of flickering events is especially challenging due to the filtering. For an illustration how the signal, and hence the expectation of the observations, is smoothed by filtering see Figure 3.2 and the explanations around. Finding and analyzing such phenomena is especially important as both have typically their own dynamics and can result from different molecular processes than the usual gating on larger scales.

Hence, a good detection power is required on small, but also on large time scales. To this end, we will follow in this section a multiscale approach as well. A first multiscale approach for ion channel recordings, called J-SMURF, was proposed by Hotz et al. (2013). It detects events on larger time scales well and controls at the same time the probability to overestimate the number of change-points, but has almost no detection power on small scales. For more details see Section 3.4, where we recall and extend this approach to observations with heterogeneous noise, and the simulations and applications in Sections 3.9 and 3.10.

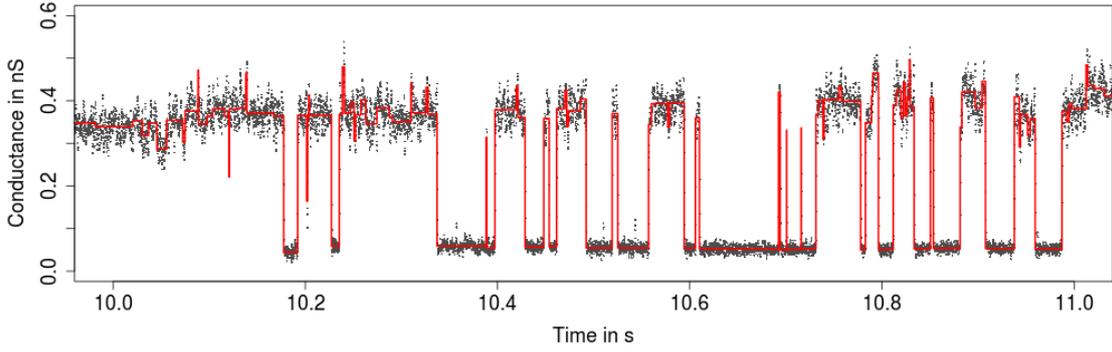
In comparison, many existing methods such as amplitude thresholding (Colquhoun, 1987;

Sakmann and Neher, 1995) and slope thresholding (Basseville and Benveniste, 1983; VanDongen, 1996) require additional filtering or manually chosen thresholds and, hence, threaten to miss events on certain scales as the filter length predetermines the scales on which events can be detected. On the other hand, many of these methods do not take into account the filter and, hence, include a severe amount of false positives. We compare our multiscale methods in Section 3 exemplarily with TRANSIT from VanDongen (1996) and briefly with an estimation based on the minimal description length, MDL, proposed by Gnanasambandam et al. (2017). All in all, we conclude that our multiscale method JULES, proposed in (Pein et al., 2017d) and described in Section 3.5, is to our best knowledge the first non-parametric estimation method for ion channel recordings that detects events on very small time scales reliably, where the filter complicates estimation severely, while still being able to detect events on large scales well.

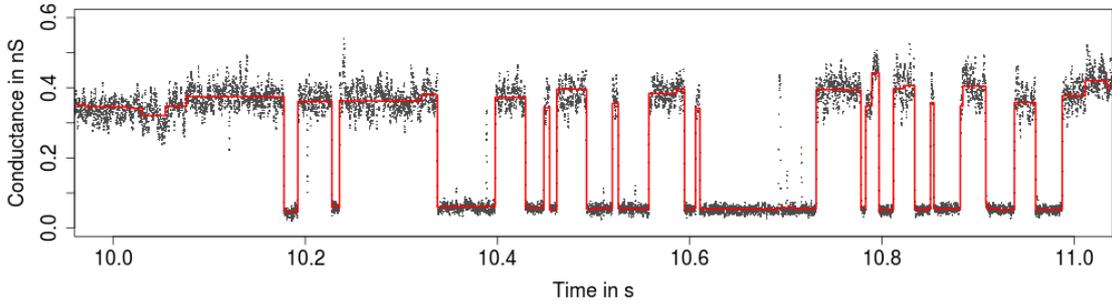
Moreover, even if events are detected, precise estimation of the change-point locations and the function values is difficult, in particular on small scales, due to the convolution of the signal with the lowpass filter, see Figure 3.2. The only semi-automatic approach available to fit such events is the *SCAN* software from Colquhoun and Sigworth (1995) which allows *time-course fitting*. This means to estimate change-point locations and function values by least squares fitting based on an approximation of the Bessel filter kernel by a Gauss kernel, but interventions by the experimenter are required. Moreover, a function value can only be estimated by least squares fitting if the event is long enough, otherwise it is guessed by previously estimated values. Unfortunately, downloading their implementation requires individual permission, and our request was not answered. We remark that the last update was ten years ago. For a fully-automatic and precise estimation we present in Section 3.6 the deconvolution approach which we proposed in (Pein et al., 2017d). The deconvolution is performed *locally*, i.e., only few parameters are estimated from a small amount of observations at a time, which allows fast computation.

In addition, as discussed in Section 3.1.2 methods that deal with *open channel noise* are required. However, to the best of our knowledge no non-parametric approach assumes such a model. Hence, either the heterogeneous noise or the filtering has to be ignored, or the observations have to be subsampled to mitigate the filter effects. These approaches are illustrated in Figure 3.4 for the observations in the middle panel of Figure 3.3 using the estimators JULES (Pein et al., 2017d), which ignores the heterogeneous noise, and H-SMUCE (Pein et al., 2017c), which ignores the filtering.

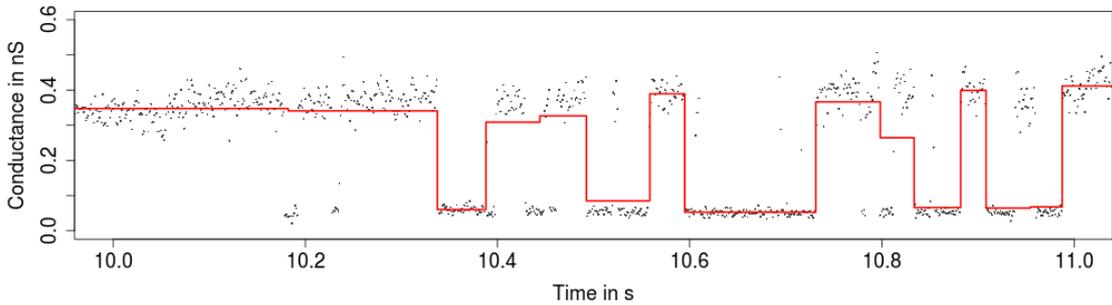
JULES detects many small events when conductance and variance are large, but not when they are small. These findings are most likely noise or base line fluctuations, which are also scaled by the open channel noise. These detections make the estimation less reliable. Indeed, the rates of a simulated hidden Markov model with parameters similar to them underlying the observations in Figure 3.3 could not be recovered by estimations using JULES, see Section 3.9.7. This effect will be more severe when the variance heterogeneity is larger. The effect of missing heterogeneous noise is also illustrated in Figure 2.1 for independent observations. In comparison, H-SMUCE misses short events, see for instance



(a) Estimation by JULES, it takes into account the filtering, but not the variance heterogeneity.



(b) Estimation by H-SMUCE, it takes into account the variance heterogeneity, but not the filtering.



(c) Estimation by H-SMUCE using subsampled observations by a factor of eleven.

Figure 3.4.: Illustration of missing variance heterogeneity or the filtering for the observations in the middle panel of Figure 3.3.

the missed peaks around 10.2s, 10.4s or 10.7s. This is even more severe when the observations are subsampled. Hence, we conclude that none of these approaches leads to satisfactory results. Thus, there is a demand for methods that deal in addition to filtered data and events on various time scales also with heterogeneous noise, while the correlation structure is still explicitly given by the known filter, see Section 3.1.2. For this task we propose in Section 3.7 a new method called JILTAD.

### 3.3. Multiscale regression for filtered observations and its problems

A straightforward extension of H-SMUCE or in general of the multiscale regression estimator as described in Section 1.1 to filtered observations is not computational feasible. A first issue already arises when specifying the hypotheses. For unfiltered observations we test on all intervals on which the candidate function is constant whether its function value is the expectation of the observations on the interval. For filtered observations the expectation of the observations is not anymore given by the function value itself, but rather by the convolution of the signal with the lowpass filter at the corresponding design point. Thus, the expectation of the observations  $Y_i, \dots, Y_j$  is determined by the signal on  $[(i - L)/f_s, j/f_s]$ . The other way around, information about the underlying signal on an interval  $[i/f_s, j/f_s]$  is provided by the observations  $Y_{i+1}, \dots, Y_{j+L-1}$ . Hence, when the candidate signal is constant on an interval  $[i/f_s, j/f_s]$  to use all the available information for testing the hypothesis that the signal has function value  $m_{ij}$  on that interval versus the alternative of a different function value requires knowledge of the candidate function on  $[(i - L + 1)/f_s, (j + L - 1)/f_s]$ . This is no issue for obtaining a multiscale test, but does not allow to compute the corresponding multiscale estimator (1.2) by a dynamic program as proposed in (Frick et al., 2014a) and outlined in Section 4.1. That's because a crucial step is the computation of the acceptance region of each single local test, but this region depends now on the candidate function outside of the interval on which we test, too. Similar issues arise for minimizing the cost functional, since (4.4) is not satisfied due to the dependency of the observations. However, when dynamic programming is not applicable, we doubt that it is possible to compute the estimator efficiently such that it can be applied to hundred thousands and more observations. Hence, approximations are necessary to obtain a computable estimator.

In what follows we present with J-SMURF, JULES and JILTAD three estimators which uses multiscale regression combined with deconvolution. We remark that their analysis is less theoretically and relies on simulations and real data applications, since the assumed models are complicated and a theoretical analysis like in Section 2 for H-SMUCE would be very technical. Instead we focus on estimators that can be computed efficiently and work well for the given measurements.

### 3.4. J-SMURF

In this section we briefly present the **J**ump-**S**egmentation by **M**ulti**R**esolution **F**ilter, J-SMURF, from Hotz et al. (2013), together with some small modifications we made. Afterwards, we combine this approach with H-SMUCE from Section 2 to extend it to heterogeneous noise.

### 3.4.1. J-SMURF for homogeneous noise

As mentioned after Figure 3.2, the signal and the convolution of the signal with the lowpass filter differ only at the beginning of each segment. Hence, if the signal is constant on an interval  $[i/f_s, j/f_s]$  with function value  $m_{ij}$  and the first  $L$  observations  $Y_i, \dots, Y_{i+L-1}$  are ignored, all other observations have constant expectation given by function value  $m_{ij}$ . Hence, in J-SMURF tests are only performed on intervals longer than  $L/f_s$ , and the beginning of each interval is ignored.

Instead of using a test statistic similar to the one for independent observations

$$T_i^j(\mathbf{Y}, m_{ij}) := (j - i - L + 1) \frac{(\bar{Y}_{i+L,j} - m_{ij})^2}{2\sigma_0^2}, \quad (3.16)$$

as in Hotz et al. (2013), we use the *partial sum* test statistic

$$T_i^j(\mathbf{Y}, m_{ij}) := (j - i - L + 1)^2 \frac{(\bar{Y}_{i+L,j} - m_{ij})^2}{2\text{Var} \left[ \sum_{l=i+L}^j Y_l \right]}, \quad (3.17)$$

with  $\text{Var} \left[ \sum_{l=i+L}^j Y_l \right] = \sigma_0^2 (j - i + 1 - L) \mathcal{A}(\infty, 0) + 2 \sum_{l=1}^L (j - i + 1 - L - l)_+ \mathcal{A}(\infty, l/f_s)$  and  $x_+ = \max(x, 0)$ . This statistic adapts better to the correlation structure. Although, corresponding tests at the same significance level are identical as the relative difference is only a constant, such a scaling matters for a multiscale test if the quantiles are chosen by scale penalization. Note that in (Hotz et al., 2013) the local test statistic is written as  $\sqrt{2T_i^j(\mathbf{Y}, m_{ij})}$ , but also the penalization and definition of  $\mathbf{q}$  are changed accordingly. Even more powerful, in particular on intervals only slightly larger than the filter length, is the likelihood ratio test statistic for the given filtered model, i.e.,

$$T_i^j(\mathbf{Y}, m_{ij}) := \frac{\left( (\mathbf{Y}_{i+L,j} - \mathbf{m}_{ij})^t \Sigma_{i+L,j}^{-1} \mathbf{1} \right)^2}{2 \mathbf{1}^t \Sigma_{i+L,j}^{-1} \mathbf{1}}, \quad (3.18)$$

with  $\mathbf{Y}_{i+L,j} = (Y_{i+L}, \dots, Y_j)^t$  the vector of observations,  $\mathbf{m}_{ij}$  and  $\mathbf{1}$  vectors of the same length with all entries equal to  $m_{ij}$  and 1, respectively,  $\Sigma_{i+L,j}$  the variance-covariance matrix of the observations  $Y_{i+L}, \dots, Y_j$  given by (3.5) and  $(\cdot)^t$  the transposition. However, its computation lasts much longer. A slightly worse detection power should not be a big concern, since J-SMURF is mainly used when all events occur on large scales and here the difference is negligible. Moreover, we will use J-SMURF in Section 3.7 as a first step with improvements on smaller scales later on, but also therefore a slightly worse detection power on small scales has no significant influence.

Tests are performed over all intervals on which the candidate function  $\mu$  is constant and that contains a dyadic number of observations equal to or larger than  $L$ . Note that Hotz et al. (2013) performed tests on a different interval set. They tested over all intervals on which the candidate function is constant and for which  $j - i + 1 - L$  is a dyadic number.

Although depending on the true signal, their choice leads in general to a better detection power on scales slightly larger than the filter length, but its computation lasts much longer. The very long computation times of J-SMURF was one major criticism in (Gnanasambandam et al., 2017). To be fair, both implementations differ in other points, too, and a fast implementation of their set might be possible as well, but our approach was easier to integrate in the dynamic programming framework of the *stepR* package (Pein et al., 2017a).

The vector of critical values  $\mathbf{q} = (q_1, \dots, q_n)$  is obtained by scale penalization as described in Section 1.1 such that  $T_n$  is a level  $\alpha$  test, but  $(j-i+1)$  in (1.6) is replaced by  $(j-i+1-L)$  to account for the smaller amount of observations used in the test statistic.

Finally, J-SMURF is obtained by minimization of  $L(\mathbf{Y}, \mu) = \sum_{l=1}^n (Y_l - \mu(l/f_s))^2$  restricted to all solutions in the optimization set  $C(\mathbf{Y}, \mathbf{q})$  as defined in (1.3). This restricted least squares estimator ignores the convolution and hence the estimated change-point locations are typically a little bit shifted to the right. For more details see the similar discussion for JULES in Section 3.6.1. To correct for this, Hotz et al. (2013) suggested to move all estimated change-point locations by a constant factor  $t_0$ , defined by  $\mathcal{F}_L(t_0) = 1/2$ , to the left. Alternatively, the deconvolution approach from Section 3.6 can be used for J-SMURF as well. In the simulations we will use the approach from Hotz et al. (2013). For JILTAD we will use a deconvolution approach as in Section 3.6.

J-SMURF has good detection properties when events are long enough, at least longer than the filter length  $L/f_s$ , since no test is performed on scales below. For more details see the simulation study and application in (Hotz et al., 2013) in which J-SMURF was used to prove the existence of subgating by an acylated Gramicidin A derivative, see (Hotz et al., 2013, Figure 8). Note that the recordings in Section 3.10 are of native Gramicidin A, which does not show subgating, but flickering, i.e., events on very small time scales.

### 3.4.2. J-SMURF for heterogeneous noise

With the work in Section 2, an extension of J-SMURF to heterogeneous noise is rather straightforward. The simplest choice for a local test statistic is the H-SMUCE test statistic without taking into account the first  $L$  observations, i.e.,

$$T_i^j(\mathbf{Y}, m_{ij}) := (j-i+1-L) \frac{(\bar{Y}_{i+L,j} - m_{ij})^2}{2\hat{\sigma}_{i+L,j}^2}, \quad (3.19)$$

with  $\bar{Y}_{i+L,j} = (j-i+1-L)^{-1} \sum_{l=i+L}^j Y_l$  and  $\hat{\sigma}_{i+L,j} = (j-i-L)^{-1} \sum_{l=i+L}^j (Y_l - \bar{Y}_{i+L,j})^2$ . This statistic allows the construction of a multiscale test, since the vector of critical values can be chosen by weights as in Section 2.2.1, because under the hypothesis the distribution of the local test statistic is independent of the unknown parameters. Alternatively, the dependency can be taken into account explicitly. Firstly, by correcting the bias of the mean value and of the variance estimate similar as above for the partial sum test. One might call it a *studentized partial sum* test statistic, although it is not t-distributed due

to the dependency. However, both statistics lead to the same multiscale test, since the two statistics differ only by a constant scaling factor and we obtain the vector of critical values by weights, where such a scaling has no influence. Secondly, the likelihood ratio test statistic can be used. We found again in simulations (not displayed) that the likelihood ratio test statistic is slightly more powerful on smaller scales, but much slower to compute. Hence, we use the test statistic (3.19).

We define the final estimator as H-SMUCE in (2.3), in particular with the same functional to minimize. Also this approach misestimates the change-point locations a little bit as it does not take into account the convolution of the signal with the lowpass filter.

### 3.5. JULES

The major drawback of J-SMURF is that it has (almost) no power on small scales, since no tests are performed on scales below the filter length  $L/f_s$ . To tackle this issue we present in this section our multiscale approach from (Pein et al., 2017d, Section III.A). We disregard for the moment the underlying convolution of the signal with the lowpass filter to obtain a computable multiscale regression estimator that takes into account scales below the filter length. More precisely, we develop an estimator for the model

$$Y_i = \mu(i/f_s) + \epsilon_i, \quad i = 1, \dots, n. \quad (3.20)$$

Here,  $\mu$  is still a function of type (3.1). Thus, similar to the independent model we have a piecewise constant function without convolution perturbed by random Gaussian errors  $\epsilon_1, \dots, \epsilon_n$ , but they are  $L$ -dependent with known covariance structure given by the filter. In other words, with respect to model (3.4), in a first step we aim for a piecewise constant approximation of  $F_L * \mu$  which will be refined in Section 3.5.2 via postfiltering to an estimation of  $\mu$ . Therefore, the key observation is again that the signal and its convolution only differ at the first  $L$  design points of each segment. The final estimator of  $\mu$ , called **J**Ump **L**ocal **d**Econvolution **S**egmentation filter, **JULES**, is obtained in Section 3.6 by combining this approach with local deconvolution. A summary of these steps is given in Algorithm 1. Figure 1.3 shows exemplarily an estimation of the signal underlying the observations in Figure 1.1.

#### 3.5.1. Piecewise constant approximation

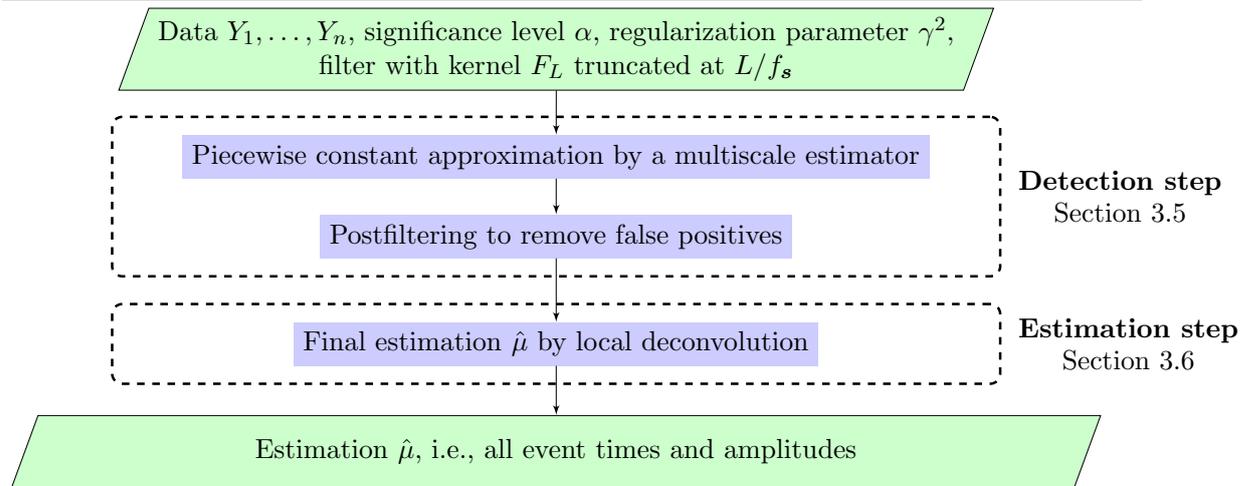
Assuming model (3.20), we use a multiscale estimator (1.2) with the same functional  $L(\mathbf{Y}, \mu)$  as for J-SMURF, i.e.,  $L(\mathbf{Y}, \mu) = \sum_{l=1}^n (Y_l - \mu(l/f_s))^2$ , but with local test statistic

$$T_i^j(\mathbf{Y}, m_{ij}) := (j - i + 1)^2 \frac{(\bar{Y}_{i,j} - m_{ij})^2}{2\text{Var} \left[ \sum_{l=i}^j Y_l \right]}, \quad (3.21)$$

---

**Algorithm 1** Steps of JULES.
 

---



with  $\text{Var} \left[ \sum_{l=i}^j Y_l \right] = \sigma_0^2 (j - i + 1)(F * F)(0) + 2\sigma_0^2 \sum_{l=1}^L (j - i + 1 - l)_+(F * F)(l/f_s)$  and  $x_+ = \max(x, 0)$ . We use again a partial sum test, since it provides the best compromise between detection power and computation time, as previously discussed for J-SMURF.

### 3.5.2. Postfiltering

As noticed in Figure 3.2, the signal  $\mu$  and the convolution of the signal with the truncated filter  $F_L * \mu$  only differ at the beginning of each segment, i.e., when  $\mu$  changes abruptly at  $\tau_i$  from  $m_{i-1}$  to  $m_i$ ,  $F_L * \mu$  changes on  $(\tau_i, \tau_i + L/f_s)$  continuously from  $m_{i-1}$  to  $m_i$  and is afterwards constant and identical to the signal  $\mu$  until the next change at  $\tau_{i+1}$  occur. Consequently, the prior piecewise constant approximation might have multiple change-points in  $[\tau_i, \tau_i + L/f_s]$  instead one change-point at  $\tau$ , but up to estimation errors the change-point at  $\tau$  does not cause change-points somewhere else, see (Boysen et al., 2009a; Li et al., 2017) for a mathematically rigorous statement. To correct for this we merge a segment with all subsequent segments for which the distance between the two starting points is less than  $L/f_s$  and if all changes are in the same direction. More precisely, we set the starting point of the merged segments to the starting point of the first segment and its function value to the value on the last segment. This ignores the small *overshoot* of the Bessel filter, i.e., kernel values below zero, which potentially causes additional changes in the opposite direction. We found that this has no significant effect for the data we analyzed, since the overshoot is very small in comparison to the noise level, see Figure 3.1, and could not be detected. Hence, simulations show that our aim to bound the probability to overestimate the number of change-points from above by  $\alpha$  is indeed satisfied. However, a rigorous proof seems to be difficult and requires in comparison to Theorem 5 more assumptions on the observations. For instance the atypical assumption that the signal to noise ratio is low enough such that the overshoot of the Bessel filter is not detected.

### 3.5.3. Heterogeneous noise

Unfortunately, the approach is not feasible for heterogeneous noise. Since the expectation, given by the convolution, is not constant on  $(\tau_i, \tau_i + L/f_s)$ , any local variance estimator that assumes a constant expectation overestimates the variance severely. Hence, tests using a test statistic similar to (2.7) have (almost) no power on such an interval. Thus, such an approach is not able to detect events below the filter length and detects roughly the same events as J-SMURF assuming heterogeneous noise. All in all, in the presence of heterogeneous noise it is mandatory to take into account the convolution explicitly to detect short events. We will propose such an approach in Section 3.7.

## 3.6. Local deconvolution

In this section we present the parameter estimation step of JULES from (Pein et al., 2017d, Section III B). In this step the final estimation is obtained by estimating the precise change-point locations  $\hat{\tau}_1, \dots, \hat{\tau}_{\hat{K}}$  and function values  $\hat{m}_0, \dots, \hat{m}_{\hat{K}}$  by local deconvolution. As noted after Figure 3.2, a deconvolution approach is crucial for estimating events on scales below the filter length with high accuracy. Nonetheless, this approach can be applied to J-SMURF as well to correct the shifted change-point locations. To this end, we split the constant segments in short and long ones. Basically, long segments are those in which the function value can be estimated accurately without taking the lowpass filter explicitly into account, details are given below. Pre-estimating the function value on long segments allows to deconvolve the signal *locally*, i.e., to estimate at one time only few parameters from only few observations, which makes the deconvolution computationally feasible. These parameters are estimated by the maximum likelihood estimator, which is computed by an iterative grid search, see below for algorithmic details.

### 3.6.1. Long segments

We refer to a segment  $[\hat{\tau}_k, \hat{\tau}_{k+1})$  as long if we can estimate its function value  $\hat{m}_k$  accurately without taking into account the lowpass filter explicitly. Recall from Figure 3.2 and the explanations around that if the signal  $\mu$  changes at  $\tau_k$ , its convolution  $F_L * \mu$  changes continuously on  $(\tau_k, \tau_k + L/f_s)$  and is constant on  $[\tau_k + L/f_s, \tau_{k+1}]$  and equal to  $m_k$ . Hence, if the change-point  $\tau_k$  is detected, its estimated location will be most likely in the interval  $[\tau_k, \tau_k + L/f_s]$ . The other way around, a detected change-point  $\hat{\tau}_k$  implies a change-point in the interval  $[\hat{\tau}_k - L/f_s, \hat{\tau}_k]$  and, hence, a constant convolution on  $[\hat{\tau}_k + L/f_s, \hat{\tau}_{k+1} - L/f_s]$ . If the observations in this interval are enough to determine the function value, say ten or more, by the median, i.e.,  $\hat{m}_k = \text{median}(Y_{f_s \hat{\tau}_k + L}, \dots, Y_{f_s \hat{\tau}_{k+1} - L})$ , we denote the segment as long and otherwise as short.

### 3.6.2. Maximum likelihood estimation

Having the function values on long segments estimated enables us to perform the deconvolution locally, i.e., at one time only few observations have to be taken into account and only few parameters have to be estimated. More precisely, let  $[\hat{\tau}_{i-1}, \hat{\tau}_i)$  and  $[\hat{\tau}_j, \hat{\tau}_{j+1})$ ,  $i \leq j$ , be two consecutive long segments (no long segment inside, but potentially some short segments). Then, we aim to improve the estimated change-point locations  $\hat{\tau}_i, \dots, \hat{\tau}_j$  and the estimated function values  $\hat{m}_i, \dots, \hat{m}_{j-1}$ . Note that the function values  $\hat{m}_{i-1}$  and  $\hat{m}_j$  are already estimated well by the median as described above. Recall that a detected change-point  $\hat{\tau}$  implies a change-point in  $[\hat{\tau} - L/f_s, \hat{\tau}]$  and that the function value  $\hat{m}_i$  affects the convolution on  $(\hat{\tau}_i, \hat{\tau}_{i+1} + L/f_s)$ . Therefore, we only have to take into account the observations  $Y_{f_s \hat{\tau}_i - L + 1}, \dots, Y_{f_s \hat{\tau}_j + L - 1}$  and maximize its likelihood in the parameters to improve. This is equivalent to minimize

$$(\mathbf{Y}_{i,j} - \mathbf{m}_{i,j})^t \Sigma_{i,j}^{-1} (\mathbf{Y}_{i,j} - \mathbf{m}_{i,j}), \quad (3.22)$$

with  $\mathbf{Y}_{i,j} := (Y_{f_s \hat{\tau}_i - L + 1}, \dots, Y_{f_s \hat{\tau}_j + L - 1})^t$  the vector of the observations,

$$\mathbf{m}_{i,j} = \left( (F_L * \mu) \left( \frac{f_s \hat{\tau}_i - L + 1}{f_s} \right), \dots, (F_L * \mu) \left( \frac{f_s \hat{\tau}_j + L - 1}{f_s} \right) \right)^t$$

the vector of the expected conductance levels  $F_L * \mu$  for a signal  $\mu$  and  $\Sigma_{i,j}$  the known correlation matrix of the observation vector given by (3.5). For the signal  $\mu$  the number of change-points and the function values  $\hat{m}_{i-1}$  and  $\hat{m}_j$  are fixed by the prior estimation and the change-point locations are restricted to the intervals  $[\hat{\tau}_k - L/f_s, \hat{\tau}_k]$ ,  $k = i, \dots, j$ . Minimization is performed by an iterative grid search, see below.

### 3.6.3. Regularization

For the Bessel filter used in the later analyzed recordings we have to regularize the correlation matrix  $\Sigma_{i,j}$  in (3.22), since the matrix is ill-conditioned, i.e., has a condition number around  $10^{-4}$ . Hence, small errors in the model, for instance in ion channel recordings often base line fluctuations occur, are amplified and may result in a large misestimation. A standard approach to cope with this is *Tikhonov* /  $L_2$  regularization, i.e., we replace  $\Sigma_{i,j}$  by  $\Sigma_{i,j} + \gamma^2 I$ , with  $I$  the identity matrix and  $\gamma^2 \geq 0$  a regularization parameter, see (Engl et al., 2000, Chapter 5). Such a regularization allows quick computation of (3.22) by the Cholesky decomposition, which can be stored, and by solving a triangular system of equations using the banded structure of the matrix  $\Sigma_{i,j} + \gamma^2 I$ . This is implemented in C++ by interfacing the FORTRAN routines *dpbtf2* and *dtbsv* from the linear algebra package *LAPACK*. Note that the correlation structure is determined by the filter and hence the parameter  $\gamma^2$  can be chosen universally for all recordings on the same system. For our data we choose the regularization parameter such that the convolution fits well to the recordings, which is nicely confirmed by visual inspection as exemplary shown in the

lower panels of Figures 1.3 and 3.22. Note, that this precalibration step has to be done only on a small data excerpt in advance. We found that the results are robust in  $\gamma^2$ , and simply setting  $\gamma^2 = 1$  led to satisfactory results in all cases. For automatic choices, e.g., by cross-validation, and for other regularizations, e.g., truncated SVD, see (Vogel, 2002) and the references therein.

### 3.6.4. Iterative grid search

We now describe how we compute the (local) maximum likelihood estimator by an iterative grid search. In general, the computation of the estimator is difficult due to the non-convexity of the optimization problem to minimize (3.22). Grid search means that we fix for each change-point  $\hat{\tau}_i, \dots, \hat{\tau}_j$  a set of possible locations, the grid, and compute (3.22) for all combinations and take the solution with the minimal value. Note that for given change-point locations close formulas for the optimal levels  $\hat{m}_i, \dots, \hat{m}_{j-1}$  exist, which allows fast computation of (3.22). In theory such an optimization can be done for arbitrarily fine grids. However, to keep it computationally feasible, we start with the observations grid. More precisely, we use  $\{\hat{\tau}_k - L/f_s, \dots, \hat{\tau}_k\}$  for the change-point  $\hat{\tau}_k, k = i, \dots, j$ , since a detected change-point  $\hat{\tau}_k$  implies a change-point in the interval  $[\hat{\tau}_k - L/f_s, \hat{\tau}_k]$ . Afterwards, a refinement can be done by repeating the grid search iteratively with finer and finer grids around the change-point locations found in the previous step. We found that this refinement for estimating the locations with arbitrary precision works very well in practice. We found it sufficient to iterate such a refinement twice, each with a ten times finer grid between the neighboring grid points of the candidate location found in the previous step. This approach is computationally feasible for a single change-point between two long segments, i.e., only one change-point location has to be estimated, and for a peak, i.e., a short segment between two long segments, here we estimate the function value of the short segment and the two change-point locations of the short segment between the two long segments. In rare situations more than one short segment between two long segments occur. The same approach could then be done for  $k$  consecutive short segments by looking for  $k + 1$  change-points between two long segments and estimating all  $k$  function values. However, such an optimization is time consuming and not necessary for our data. Hence, we abandon and simply return for these parameters the estimation from Section 3.5. For the analysis we drop these segments which leads to potential miss of events if they are too close together. Figure 3.12 shows that for the analyzed Gramicidin A recordings this only occurs if the distance between two events is less than 3.2 ms, roughly 3 times the filter length. This is much smaller than the estimated average distance between two events of 0.3 s and hence the whole effect was indeed negligible. Also for the PorB recordings we analyze in Section 5 the minimal required distance is much shorter than the average distance.

### 3.7. JILTAD

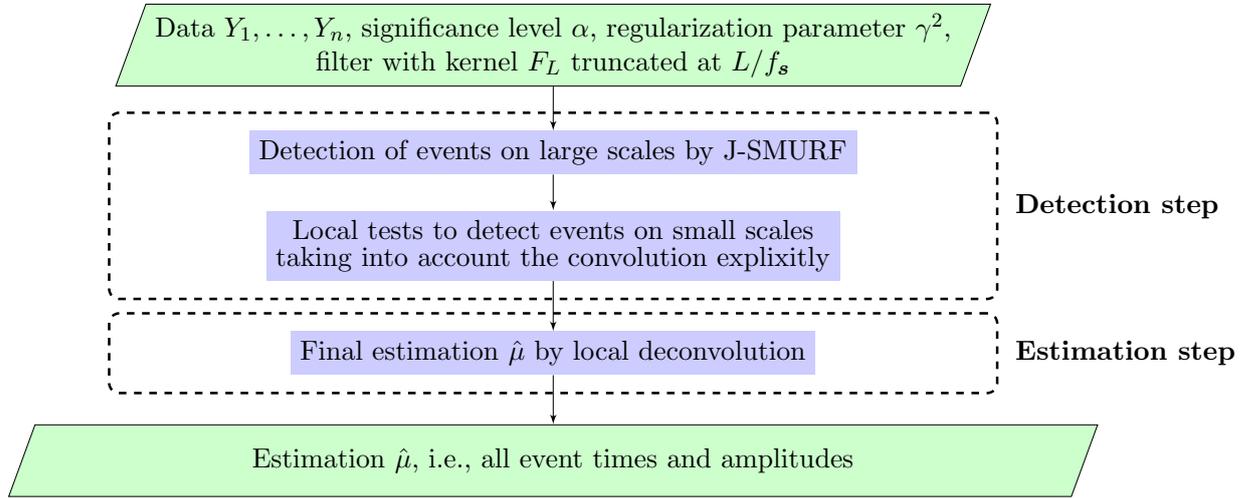
JULES estimates events on small scales well if the noise is homogeneous, but an extension of this approach to heterogeneous noise is impossible, see Section 3.5.3. Moreover, it does not take into account the deconvolution for detection, which potentially reduces the detection power. We will indeed see in simulations in Section 3.9.3 that the detection power of the method we propose now is larger than the one of JULES.

To overcome these issues we propose in this work a new method called **J-SMURF Improved by Local Tests And Deconvolution**, JILTAD, that takes into account the convolution explicitly for detecting events. As discussed in Section 3.3, this requires knowledge of the signal in a neighborhood of the interval on which we test, which makes it difficult to compute the corresponding multiscale estimator efficiently. Similar to the deconvolution we overcome this burden by focusing firstly on large scales and then improving the estimation on small scales. Events on large scales are detected by J-SMURF as described before without taking into account the convolution explicitly. Then, in a second step we test locally whether additional events on small scales have to be incorporated. For this the convolution is taken into account explicitly by assuming that signal and noise left and right of the interval on which we test are given by the previous estimation by J-SMURF. The final change-point locations and function values are again estimated by local deconvolution. Exact details how to perform and combine these steps are described in the following. We call this method **J-SMURF Improved by Local Tests And Deconvolution**, JILTAD. A summary about these steps is given in Algorithm 2.

---

**Algorithm 2** Steps of JILTAD.

---



An exemplary estimation in Figure 3.5 for the observations in Figure 3.3 shows that at presence of heterogeneous noise this method is indeed able to estimate events on small scales well, while at the same time it does not include additional artificial change-points in segments with larger variance.

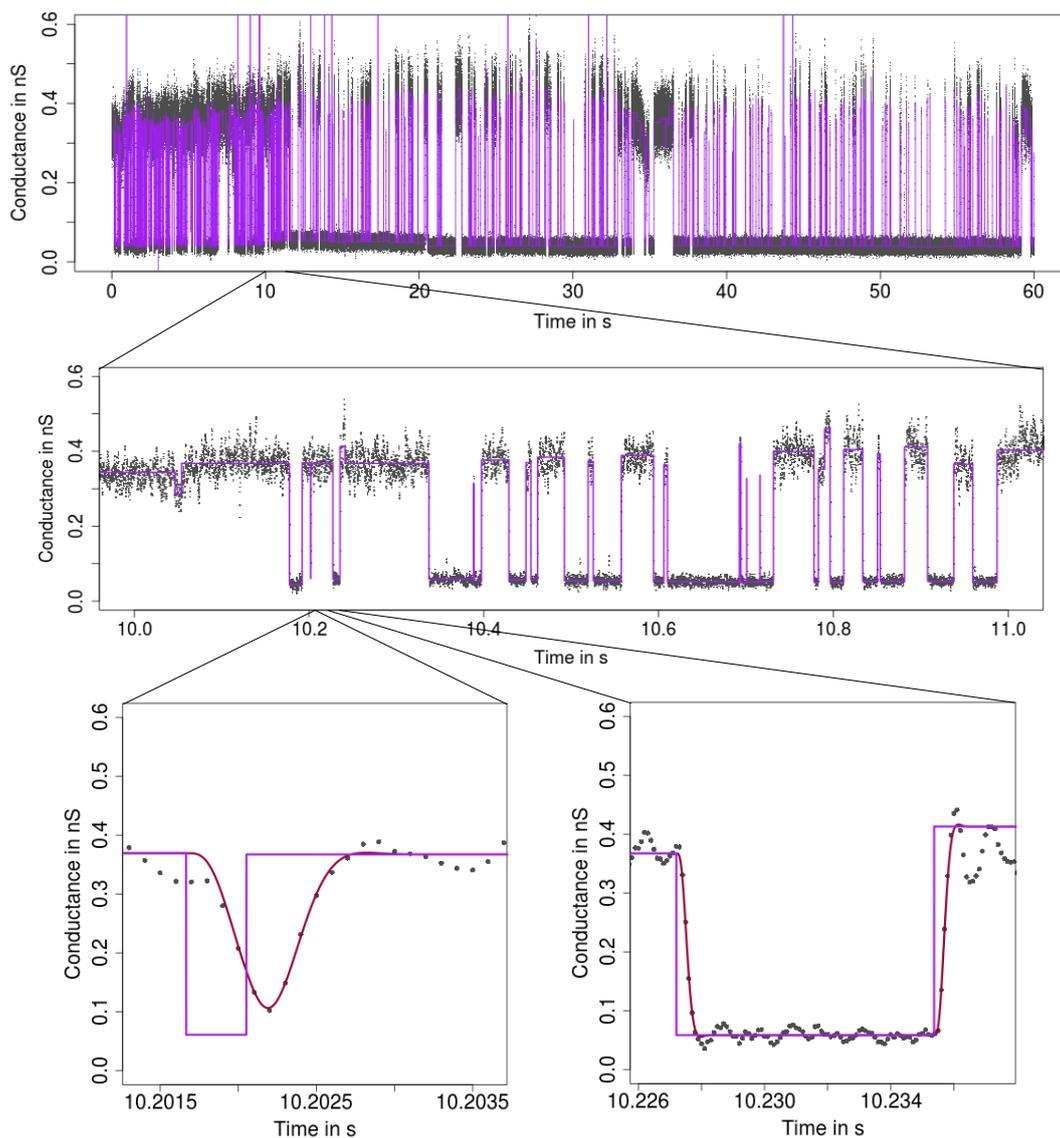


Figure 3.5.: Estimation by JILTAD (purple) of the signal underlying the observations in Figure 3.3 and its convolution with the lowpass filter (red-purple).

### 3.7.1. J-SMURF step

We apply J-SMURF as described in Section 3.4 with significance level  $\alpha_1 < \alpha$ . We do not correct in this step for the shifted locations. This will be done by a deconvolution approach as described in Section 3.7.2.

### 3.7.2. Local testing

To find events on small scales we test on all intervals containing  $l = 1, \dots, l_{\max}$  observations whether the estimation by J-SMURF is the underlying signal or whether an additional event on the tested interval is significantly better. Here,  $l_{\max}$  should be chosen such that events on all larger scales are already detected by J-SMURF (or have such a small jump

size that they are also not detectable by these tests). By default, we choose  $l_{\max} = 20$  for homogeneous observations and  $l_{\max} = 65$  in the presence of heterogeneous noise.

As explained in Section 3.3, only the observations  $Y_{i+1}, \dots, Y_{j+L-1}$  provide information about the test problem and the signal on  $[(i - L + 1)/f_s, (j + L - 1)/f_s]$  affects the expectation of these observations. Hence, for a local test on an interval  $[i/f_s, j/f_s]$  we distinguish few scenarios depending on how many detected change-points are in  $[(i - L + 1)/f_s, (j + L - 1)/f_s]$ . If no change-point is detected we test a constant function versus the alternative of an additional event on  $[i/f_s, j/f_s]$  with an arbitrary function value. If one change-point is detected we test a signal with a change-point at the location that maximizes the likelihood, details are given in the next paragraph, versus the alternative of an additional event on  $[i/f_s, j/f_s]$  with arbitrary function value. If the test rejects, the single change-point is replaced by two and the locations and function value are obtained in the upcoming deconvolution step. A precise definition of the hypothesis is given below. In the rare situation that two or more change-points are present we perform no local test to save computation time, since the parameter of more than two change-points can anyway not be estimated in the deconvolution step. However, similar as for the deconvolution it is a straightforward extension to test at presence of more change-points or for more change-points in the alternative, i.e., to test signals with  $k_1$  versus  $k_2$  change-points, but at the price of a larger computation time. However, a different and much more challenging question is the model selection task for how many change-points  $k_2$  we should test. For instance, if J-SMURF found two change-points, whether we want to test for an additional third change-point or for an additional peak given by two more change-points. Moreover, we only test on intervals with start and end point at the observation grid. Also this can be refined, again at the price of a larger computation time. We found that our choices are sufficient for the data we analyze and allow fast computation.

Recall that change-points detected by J-SMURF are typically shifted slightly to the right, since the convolution is ignored. Hence, even if the true underlying signal has only one change-point in  $[(i - L + 1)/f_s, (j + L - 1)/f_s]$  the hypothesis is often rejected. To correct for this, we reestimate the locations of all isolated change-points by local deconvolution as described in Section 3.6 without any refinement, only at the observation grid. This includes a reestimation of the function values on long segments by medians as described in Section 3.6.1. In other words, as the hypothesis we assume the signal which will be obtained by deconvolution if no test rejects, up to refinements using finer grids. At the same time, estimation of the function values on long segments by the median guarantees that they are not too badly estimated even if few short peaks are missed.

Let  $[\tau_L, \tau_R] = [i/f_s, j/f_s]$  be now the interval on which we test. And let  $\tau$  be the only change-point in  $[(i - L + 1)/f_s, (j + L - 1)/f_s]$  with function value  $m_L$  before and  $m_R$  afterwards and assume that all three parameters are already reestimated. This includes the scenario of no change-point by setting  $m_L = m_R$ . Then, we decide whether a new

segment on  $[\tau_L, \tau_R]$  is significantly better by testing the hypothesis

$$\mu_0(t) = \begin{cases} m_L & \text{if } t < \tau, \\ m_R & \text{if } t \geq \tau \end{cases} \quad (3.23)$$

versus the alternative

$$\mu_1(m)(t) = \begin{cases} m_L & \text{if } t < \tau_L, \\ m & \text{if } \tau_L \leq t < \tau_R, \\ m_R & \text{if } t \geq \tau_R, \end{cases} \quad (3.24)$$

with  $m \in \mathbb{R}$  arbitrary. In case of heterogeneous noise the same structure is assumed for variance functions  $\sigma_0^2$  and  $\sigma_1^2$  with values  $s_L^2, s^2$  and  $s_R^2$ . On long segments, in addition to the expectation, the variance is estimated by (3.11) using the same observations as used for estimating the expectation. However, we require at least 25 instead of ten observations in the definition of a long segment to guarantee a reasonable estimation of the variance.

### 3.7.3. Local test statistic for homogeneous noise

If the noise is homogeneous we use the (regularized) likelihood ratio test statistic

$$T_i^j(\mathbf{Y}) := (\mathbf{Y}_{i+1, j+L-1} - (\mathbf{m}_0)_{i+1, j+L-1})^t \Sigma_{i+1, j+L-1}^{-1} (\mathbf{Y}_{i+1, j+L-1} - (\mathbf{m}_0)_{i+1, j+L-1}) \\ - (\mathbf{Y}_{i+1, j+L-1} - (\mathbf{m}_1(\hat{m}))_{i+1, j+L-1})^t \Sigma_{i+1, j+L-1}^{-1} (\mathbf{Y}_{i+1, j+L-1} - (\mathbf{m}_1(\hat{m}))_{i+1, j+L-1}),$$

with

$$\hat{m} := \operatorname{argmax}_{m \in \mathbb{R}} (\mathbf{Y}_{i+1, j+L-1} - (\mathbf{m}_1(m))_{i+1, j+L-1})^t \Sigma_{i+1, j+L-1}^{-1} (\mathbf{Y}_{i+1, j+L-1} - (\mathbf{m}_1(m))_{i+1, j+L-1}).$$

Here,  $(\mathbf{m}_0)_{i+1, j+L-1}$  and  $(\mathbf{m}_1(m))_{i+1, j+L-1}$  are the vectors

$$(\mathbf{m}_0)_{i+1, j+L-1} = ((F_L * \mu_0)((i+1)/f_s), \dots, (F_L * \mu_0)((j+L-1)/f_s))^t, \\ (\mathbf{m}_1(m))_{i+1, j+L-1} = ((F_L * \mu_1(m))((i+1)/f_s), \dots, (F_L * \mu_1(m))((j+L-1)/f_s))^t$$

and  $\Sigma_{i+1, j+L-1}$  the covariance matrix of the observations  $Y_{i+1}, \dots, Y_{j+L-1}$  given by (3.5), regularized by Tikhonov regularization with parameter  $\gamma^2 = \sigma_0^2$ .

### 3.7.4. Local test statistic for heterogeneous noise

For detecting events we aim to use the information provided by potential variance changes at the change-point locations of the expectation. As discussed in Section 2.1, this is different to the test problem to be robust against variance changes for which we proposed H-SMUCE in Section 2. Note that if expectation and variance are tested simultaneously, a multiscale regression estimator is difficult to compute. That's because the usual computation by a dynamic program relies on minimizing a cost functional on an intersection of acceptance regions of tests included in the multiscale test, see Section 4.1. And these

regions are for this setting two dimensional sets instead of one dimensional intervals. Although such an optimization problem can be solved by Dykstra's projection algorithm (Dykstra, 1983) (and similar algorithm) it lasts too long when in the dynamic program such problems have to be computed many times for a large number of sets. But since here only testing is required it is computational feasible. This is another gain of the approach to use the previous estimation by J-SMURF and to obtain the final parameters by deconvolution.

The test problem is also different to the one in (Enikeeva et al., 2016), since we still allow the variance to be constant ( $s^2 = s_L^2$  or  $s^2 = s_R^2$ ) when the expectation changes and also  $m$  and  $s^2$  are unknown and have to be estimated from the observations. However, an approach for known parameter by assuming an alphabet for the function values and variance is described in Section 6.2.2. We stress that in the setting of this section the parameter space for the alternative is large and even for unfiltered observations it is unknown whether an adaptive test exists that is at least rate optimal for *all* alternatives. A theoretical exploration of these questions is interesting but difficult and beyond the scope of this work, where we aim for an estimator that works well in the algorithmically difficult model of filtered observations and can be computed quickly. Hence, in the following we aim for a test that provides a good power for the events in the measurements in Section 5.2, see Figure 3.3, and is computational feasible.

It follows from (3.14) that under the alternative the covariance is given by

$$\text{Cov}[Y_l, Y_{l+r}] = \begin{cases} w_{l,r}s^2 + s_{LR,l,r}^2 & \text{for } |r| = 0, \dots, L, \\ 0 & \text{for } |r| > L, \end{cases} \quad (3.25)$$

with

$$w_{l,r} := \mathcal{A}(l/f_s - \tau_L, r/f_s) - \mathcal{A}(l/f_s - \tau_R, r/f_s) \quad (3.26)$$

and

$$s_{LR,l,r}^2 := s_L^2 [\mathcal{A}(\infty, r/f_s) - \mathcal{A}(l/f_s - \tau_L, r/f_s)] + s_R^2 \mathcal{A}(l/f_s - \tau_R, r/f_s). \quad (3.27)$$

We remark that the assumed structure in (3.23) and (3.24) for variance function leads to an approximation for the covariance in (3.25), since further change-points in the variance function are ignored. However, since the kernel decays exponentially and the constant segments are by construction rather long, the resulting errors should be negligible, but computation is simplified tremendously.

Since the likelihood function involves the inverse and the determinant of the covariance matrix given by (3.25), computing the likelihood ratio test statistic requires time consuming numerical optimization. Hence, computation of this test statistic on various intervals for in total hundred thousands observations is not feasible. This already applies for the computation of the maximum likelihood estimator for  $(m, s^2)$ . And even the computation of other sensible estimators for  $(m, s^2)$  jointly lasts probably very long.

Instead we use for  $m$  the least squares estimator

$$\begin{aligned}\hat{m} &:= \operatorname{argmin}_{m \in \mathbb{R}} \sum_{l=i+1}^{j+L-1} (Y_l - \mathbb{E}[Y_l])^2 = \operatorname{argmin}_{m \in \mathbb{R}} \sum_{l=i+1}^{j+L-1} (Y_l - v_l m - m_{LR,l})^2 \\ &= \frac{\sum_{l=i+1}^{j+L-1} v_l (Y_l - m_{LR,l})}{\sum_{l=i+1}^{j+L-1} v_l^2},\end{aligned}\tag{3.28}$$

where it follows from (3.10) that

$$v_l := \mathcal{F}_L(l/f_s - \tau_L) - \mathcal{F}_L(l/f_s - \tau_R)\tag{3.29}$$

and

$$m_{LR,l} := m_L[1 - \mathcal{F}_L(l/f_s - \tau_L)] + m_R \mathcal{F}_L(l/f_s - \tau_R).\tag{3.30}$$

This estimator is identical with the maximum likelihood estimator when the covariance structure is ignored and the observations are assumed to be independent and to have homogeneous noise, but the convolution of the signal with the filter is taken into account. Moreover, it illustrates that taking into account more observations does not help, since  $v_l = 0$  for  $l < i + 1$  or  $l > j + L - 1$ .

For estimating the variance  $s^2$  we use the weighted estimator

$$\hat{s}^2 := \max \left( \frac{\left( \sum_{l=i+1}^{j+L-1} w_{l,0} (Y_l - v_l \hat{m} - m_{LR,l})^2 - B(s_L^2, s_R^2) \right)}{A}, 0 \right),\tag{3.31}$$

with  $A$  and  $B(s_L^2, s_R^2)$  such that

$$\mathbb{E} \left[ \sum_{l=i+1}^{j+L-1} w_{l,0} (Y_l - v_l \hat{m} - m_{LR,l})^2 \right] =: A s^2 + B(s_L^2, s_R^2).\tag{3.32}$$

It follows from (3.28) that the random variable of which we take the expectation in (3.32) can be written as a quadratic form  $(\mathbf{Y}_{i+1,j+L-1} - \mathbb{E}[\mathbf{Y}_{i+1,j+L-1}])^t C (\mathbf{Y}_{i+1,j+L-1} - \mathbb{E}[\mathbf{Y}_{i+1,j+L-1}])$ , where all entries of the matrix  $C$  are non-negative and depend only on  $v_l$  and  $w_{l,0}$ ,  $l = i + 1, \dots, j + L - 1$ . Together with (3.25) the structure proposed in (3.32) follows and  $A$  and  $B(s_L^2, s_R^2)$  can be computed explicitly. The estimator  $\hat{m}$  is unbiased, while for  $\hat{s}^2$  this would be true without the projection of negative values to zero in (3.31), which however reduces the mean squared error.

Using these estimators, under the alternative the observation  $Y_l$  has estimated expectation  $\hat{m}_{1,l} := v_l \hat{m} + m_{LR,l}$  and estimated variance  $\hat{s}_{1,l}^2 := w_{l,0} \hat{s}^2 + s_{LR,l,0}^2$ . It follows from (3.8) and (3.14) that under the null hypothesis the observation  $Y_l$  has expectation  $m_{0,l} := m_L [1 - \mathcal{F}_L(t - \tau)] + m_R \mathcal{F}_L(t - \tau)$  and variance  $s_{0,l}^2 := s_L^2 [\mathcal{A}(\infty, 0) - \mathcal{A}(l/f_s - \tau)] + s_R^2 \mathcal{A}(l/f_s - \tau, 0)$ . To obtain a test statistic these estimators can be plugged into the likelihood ratio, but its computation is rather slow and it is unclear how to regularize the covariance matrix, see a similar discussion in Section 3.7.7 for the deconvolution. Thereto, we ignore the

correlation of the observations and use the likelihood ratio test for this model which leads to

$$T_i^j(\mathbf{Y}) := \sum_{l=i+1}^{j+L-1} \log \left( \frac{s_{0,l}^2}{\hat{s}_{1,l}^2} \right) + \frac{(Y_l - m_{0,l})^2}{s_{0,l}^2} - \frac{(Y_l - \hat{m}_{1,l})^2}{\hat{s}_{1,l}^2}. \quad (3.33)$$

We are aware that this test statistic and its underlying estimators might be improvable with respect to efficiency of the estimators and the power of the resulting test for its various alternatives, but, as stressed before, we aimed for a test that has at least a good power for the recordings in Section 5.2 and can be computed efficiently. This will be confirmed by the simulations in Section 3.9.4 and in Section 4.5, respectively.

### 3.7.5. Multiple dependent rejections

Usually one event in the data causes rejections of multiple tests. We only keep the event with the largest test statistic among all rejections on intervals that adjoin each other at least. More precisely, two rejections on intervals  $[i_1/f_s, j_1/f_s]$  and  $[i_2/f_s, j_2/f_s]$ , are only considered as two separated events if the intervals are disjoint, let then w.l.o.g.  $j_1 < i_2$ , and there exists an  $l \in \{j_1 + 1, \dots, i_2 - 1\}$  such that all tests on intervals containing  $l/f_s$  accept the hypothesis. The choice of the largest test statistic is a natural choice for all tests on intervals of the same length, since they share the same distribution (under their respective null hypotheses and alternatives). For tests on intervals of different lengths this is an approximation, which we found works very well. That's because usually the test statistics are much larger when their alternative is true than when their null hypothesis is true or the underlying model is different to their null hypothesis and alternative, which outweighs the (slightly) different distributions (under their respective null hypotheses and alternatives). Also note that a slight misestimation of the change-point locations does not have a noticeable effect, since the final estimation of them is obtained by the upcoming deconvolution step.

### 3.7.6. Critical values

We combine the multiscale tests of J-SMURF and the tests on small scales in the improving step by a weighted Bonferroni correction to satisfy an overall significance level  $\alpha$ . The multiscale test of J-SMURF will be performed at a reduced significance level  $\alpha_1 < \alpha$  and the local tests are combined to a multiple test at significance level  $\alpha_2 = \alpha - \alpha_1$ . To this end, we obtain the critical values as described in Section 2.2.1 by using equal weights  $\beta_1, \dots, \beta_{l_{\max}} = 1/l_{\max}$ . To speed up the required Monte-Carlo simulations we use a small simplification. We generate observations according to the assumed model, see Section 3.9.1. But for computing the statistics we assume a constant estimation by J-SMURF instead of performing the J-SMURF step. Since the estimation by J-SMURF leads with probability at least  $1 - \alpha_1$  to a constant estimation this error is negligible. Like as for JULES, this construction aims that the probability to overestimate the number of change-points is bounded from above by  $\alpha$ . Again this bound cannot be proven, but we found in

simulations that for typical time series the probability is even smaller.

### 3.7.7. Local deconvolution

The final estimation is obtained by deconvolution as described in Section 3.6 with an adjustment on how to choose the grids. For J-SMURF and for the detection step of JULES the estimated change-point locations are shifted to the right, since the convolution was ignored. Hence, we use for a change-point  $\hat{\tau}$  detected by J-SMURF (and not replaced by two detected change-points by the local tests) still the grid  $\{\hat{\tau} - L/f_s, \hat{\tau} - (L+1)/f_s, \dots, \hat{\tau}\}$ . For a change-point  $\hat{\tau}$  detected by the local tests we use instead  $\{\hat{\tau} - \lceil L/2 \rceil / f_s, \dots, \hat{\tau} + \lceil L/2 \rceil / f_s\}$ , since the locations are not estimated precisely, but also not systematically shifted to one side. The prior reestimation of the function values is adapted in the same way. Everything else is performed in the same way as explained in Section 3.6.

This in particular means that we still use the likelihood function of observations with homogeneous noise, although heterogeneous noise is assumed. Simulations show, see Sections 3.9.4 and 3.9.7, that this works reasonably well for the recordings we analyze in Section 5.2. If the noise is larger and the relative difference of the noise values is large, taking into account the heterogeneous noise might be beneficial. However, as discussed before, the computation of the maximum likelihood estimator is too slow. Alternatively, the estimators (3.28) and (3.31) could be used and plugged in the likelihood function. Although this approach is computational feasible, it is unclear how to regularize the covariance matrix in the likelihood function. On the one hand, homogeneous noise is still part of the model and hence a regularization is required to overcome the issues raised by the ill-conditioned covariance matrix. On the other hand, in case of heterogeneous noise Tikhonov regularization equalizes the variances which results in a biased estimation of the change-point locations. All in all, the task of finding an appropriate regularization appears to be difficult and unbiased estimation might be even impossible, since keeping the variance levels and avoiding an ill-conditioned matrix might not be achievable at the same time. Hence, we still use the homogeneous approach which appears to be robust enough for our recordings, see Figure 3.5 and Section 3.9.4. Interestingly, we found in simulations (not displayed) that this approach using the regularized correlation matrix of homogeneous observations, which is for heterogeneous noise not anymore proportional to the covariance matrix, works still better than a least squares approach.

## 3.8. Analysis of the gating dynamics

After we discussed in the previous sections extensively the estimation of the underlying signal, we discuss in this section how to analyze the dynamics of an ion channel based on these estimations. Such an analysis depends of course very much on the data and on which questions are relevant. Nonetheless, we will outline some concepts in the following. We analyze exemplary one PorB wild type measurement in the presence of ampicillin.

An example of such a data set was shown in Figure 1.1 and an estimation by JULES in Figure 1.3. The channel is most of the time open and the ion flow is only briefly blocked by an ampicillin molecule. The main questions are how large is the conductance loss by a blockage and how long and how frequently the channel is blocked. For more details, for results based on all measurements and their interpretation see Section 5.1. We will also analyze the Gramicidin A recordings in Section 3.10, the PorB recordings without ampicillin, but with heterogeneous noise, in Section 5.2, and simulated data in Sections 3.9.6 and 3.9.7 in a similar way. This is also the reason why we discuss the analysis of the PorB recordings in this section and not later. We stress that this is an empirical analysis including approximations and subjective choices. Several analysis steps can be formalized and improved, but the focus of this work is on estimating the underlying signals.

Naturally, the first step should always be to plot the data together with the estimation of the signal to examine whether the estimation looks well. This also helps to identify artifacts that might disturb the later analysis and have to be taken into account or be removed by postfilter steps. Frequently, base line fluctuations, i.e., small conductance changes, occur, see for instance the oscillations of the conductance around 11.5 s in Figure 3.21. They are caused by small defects in the membrane, which is unavoidable in the recordings. There might be also periodic oscillations, resulting from the electronic or from building vibrations (although damped). Moreover, short, but large conductance spikes, often with a fixed frequency, caused by the electronic might be in the data as well. We also suggest to zoom into single change-points or if occurring into single peaks together with a plot of the convolution of the estimated signal with the assumed lowpass filter, see the lower panels in Figures 1.3, 3.5, 3.22 and 3.23. If the convolution fits the observations well, as it is the case in these figures, this a good verification of the assumed model, in particular that the filter is specified correctly.

It is crucial for the analysis that not all events are observable, shorter events are *missed*, see the simulations in Section 3.9.3. More precisely, we assume that only events with a dwell time in an interval  $[a, b]$  are observed with probability one and no artifacts of that length are detected. Other estimated events will be ignored for the analysis, since taking into account events that are detected with a probability less than one either complicate or even falsify the analysis. The same is true for lengths on which artifacts are observed. We found in simulations, similar to them in Section 3.9.3, that for the PorB recordings in Section 5.1 JULES is able to detect events of length at least 0.08 ms with almost probability one. Hence, we take into account all events with length between  $a = 0.08$  ms and  $b = 0.2$  ms, since some shorter events are missed and some larger events might be rare gating events of the channel itself and not caused by an ampicillin blockage.

Usually, before other questions can be answered, the number and values of the underlying conductance levels have to be estimated. More precisely, we assume that the functions values of the underlying signal are all contained in a finite, usually small, alphabet  $\{l_1, \dots, l_c\}$ . The number of conductance levels, also called *states*, i.e., the cardinality  $c$  of the alphabet, and the values  $l_1 < \dots < l_c$  are the parameter of interest. Using such an alphabet also as

an assumption for estimating the underlying signals is discussed in Section 6.2.2. Note that in these experiments the zero level does not correspond necessarily to zero conductance, since the base line has to be set manually. Hence, the *amplitudes*, i.e., the conductance differences  $l_2 - l_1, \dots, l_c - l_1$ , are often more informative. The estimated function values are assumed to be given by a mixture distribution with  $c$  components and the single components have expectation  $l_1, \dots, l_c$ . The errors result from measurement errors and the estimation error and are often assumed to be Gaussian distributed.

To estimate the components of the mixture, including the model selection task to estimate the number of components, many statistical methods are available, for an overview see (McLachlan and Peel, 2004) and the references therein. Moreover, histograms of the raw data, also called point amplitude histograms, and of the estimated function values, also called event histograms, can be used for visualization, see for instance (Bartsch et al., 2017, Figure 5 C, D). Note that if events are short in time a raw data histogram might be misleading due to filtering and that the channel is much longer, but not more frequently, in some states than in other states. In case of only two or very few more conductance levels histograms of the amplitudes, the differences between consecutive function values, are often more informative. Such a histogram is shown in Figure 3.6 for the recordings in Figure 1.1. Amplitude histograms for other data sets can be found in (Bartsch et al., 2017, Figure 6).

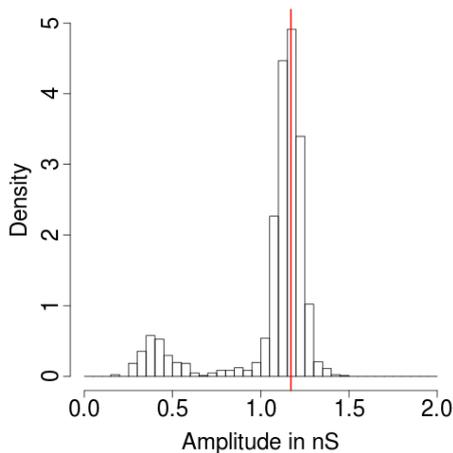


Figure 3.6.: Histograms of the amplitudes between 0 nS and 2 nS of all blockage events with residence time between 0.08 ms and 0.2 ms. Vertical red line indicates the estimated amplitude of 1.1708 nS by the half sample mode.

Comparing them with the measurement before ampicillin was added, see (Bartsch et al., 2017, Figure S9 in the Supporting Information), reveals one additional well separated peak. Boundaries of this peak  $[l, u] = [0.9, 1.5]$  nS are visually determined. All other events are neglected as they cannot be associated with an ampicillin blockage, most likely they are detected base line fluctuations. This is supported by the fact that the amount of events in the measurements without ampicillin and in the measurements with ampicillin, but outside of the interval  $[l, u]$ , are roughly the same. Here, the number of components

$c = 2$  is obtained visually as the number of distinguishable "significant" peaks in the histogram, of course a subjective choice. A conductance loss by an ampicillin blockage, i.e., the amplitude  $l_2 - l_1$ , of 1.1708 nS is estimated by the half sample mode (Robertson and Cryer, 1974), computed in  $R$  by using the *modeest* package. Alternatively, other mode estimators or estimators based on a Gaussian assumption can be used. But, we prefer the half sample mode estimator, since this estimator is rather robust against outliers.

In the following, we analyze the dwell times of the channel in the constant conductance segments. We restrict ourself to the simple model of only two conductance levels,  $c = 2$ , and assume that the channel is most of the time in one level, only briefly interrupted by short changes to the other level. For the recordings we analyze in this section this is fulfilled, since the ion flow is only briefly interrupted by ampicillin blockage events. Moreover, we will see that such a model is valid for PorB recordings without ampicillin in Section 5.2 and for the Gramicidin A recordings in Section 3.10. Extensions to more complicated models are possible, although not always straightforward, but beyond the scope of this work where we focus on estimating the underlying signal.

For many recordings, and as we will see also for the PorB recordings, a time continuous Markov model is a reasonable assumption. Note that often for each conductance level a mixture of exponential distributions is assumed, resulting in a Markov model with more than two states, but for the PorB recordings a simple two state model will be enough. More precisely, we assume a sequence of independent times  $\tilde{T}_1, \dots, \tilde{T}_{2\tilde{n}+1}$ . These times describes how long the channel dwells in a state before a change occur. Moreover, we assume with  $\tilde{C}_1, \dots, \tilde{C}_{\tilde{n}}$  exponentially distributed residence times of the ampicillin molecules, or more generally the dwell times in the closed state, and with  $\tilde{O}_1, \dots, \tilde{O}_{\tilde{n}+1}$  exponentially distributed times between two blockage events, in general the dwell times in the open state. Thus, we have  $\tilde{O}_i = \tilde{T}_{2i-1}$ ,  $i = 1, \dots, \tilde{n} + 1$ , and  $\tilde{C}_i = \tilde{T}_{2i}$ ,  $i = 1, \dots, \tilde{n}$ , whereby we assumed without loss of generality that channel is open at the beginning and end. And we denote the exponential rates with  $\lambda_O$  for the open state and with  $\lambda_C$  for the closed state. The rate  $\lambda_O$  is the frequency with which blockage events occur and  $1/\lambda_C$  is the average residence time.

Next, we have to discuss the influence of missing events. Missing events shorter than a resolution limit  $a$  is heavily discussed in the literature. The exact distribution is calculated by Hawkes et al. (1990), an estimator called MIL of the Q-matrix is suggested by Qin et al. (1996) and integrated in the QuB software package (Nicolai and Sachs, 2013), the exact maximum likelihood estimator for the Q-matrix for two conductance level is obtained by Colquhoun et al. (1996) and recently a Bayesian approach is proposed by Epstein et al. (2016). In this work for the small model of only two states we follow a simpler, more direct approach which allows at the same time to verify the assumption of an underlying two state Markov model.

By assumption we only observe closing events of length in the interval  $[a, b]$ . More precisely, we have closing times  $C_1, \dots, C_n$  such that for all  $C_i$ ,  $i \in \{1, \dots, n\}$ , there exists an  $j_i$  in  $\{1, \dots, \tilde{n}\}$  with  $C_i = \tilde{C}_{j_i}$  and for all  $j \in \{j_1, \dots, j_n\}$  we have  $\tilde{C}_j \in [a, b]$  and for all

other  $j \in \{1, \dots, \tilde{n}\} \setminus \{j_1, \dots, j_n\}$  we have  $\tilde{C}_j \notin [a, b]$ . For the opening times  $O_1, \dots, O_{n+1}$  the connection is given by  $O_i := \sum_{l=j_{i-1}+1}^{j_i-1} \tilde{O}_l + \tilde{C}_l + \tilde{O}_{j_i}$ ,  $i = 1, \dots, n+1$ , with  $j_0 = 0$  and  $j_{n+1} = \tilde{n}$ , i.e., by the sum of all times until the next observable closing event occurs. Both together, we observe the times  $T_1, \dots, T_{2n+1}$ , with  $O_i = T_{2i-1}$ ,  $i = 1, \dots, n+1$ , and  $C_i = T_{2i}$ ,  $i = 1, \dots, n$ . Also these times are independent by construction and since all times  $\tilde{T}_1, \dots, \tilde{T}_{2\tilde{n}+1}$  are independent. In the following we assume that the estimated times are realizations of  $T_1, \dots, T_{2n+1}$ , since the estimation is (almost) unbiased and the errors are small, see Sections 3.9.3 and 3.9.4. This means  $C_1, \dots, C_n$  are given by all  $\hat{\tau}_{i+1} - \hat{\tau}_i$  for which  $\hat{\tau}_{i+1} - \hat{\tau}_i \in [a, b]$  and  $\hat{m}_i - \hat{m}_{i-1} \in [l, u]$  and  $O_1, \dots, O_{n+1}$  analogue. We now want to verify the Markov assumption and estimate the rates  $\lambda_C$  and  $\lambda_O$ .

Instead of independence we only confirm that the times are uncorrelated by plotting the empirical autocorrelations of lag  $l$

$$\gamma_l := \begin{cases} \frac{1}{2n+1-l} \sum_{i=1}^{2n+1-l} \frac{T_i T_{i+l} - \bar{C} \bar{O}}{\hat{\sigma}_C \hat{\sigma}_O}, & \text{if } l \text{ odd,} \\ \frac{1}{2n+1-l} \sum_{i=1}^{n-l/2} \frac{C_i C_{i+l/2} - \bar{C}^2}{\hat{\sigma}_C^2} + \frac{1}{2n+1-l} \sum_{i=1}^{n+1-l/2} \frac{O_i O_{i+l/2} - \bar{O}^2}{\hat{\sigma}_O^2}, & \text{if } l \text{ even,} \end{cases}$$

with  $\bar{C}$  and  $\bar{O}$  the means and  $\hat{\sigma}_C$  and  $\hat{\sigma}_O$  the empirical standard deviations of the closing times  $C_1, \dots, C_n$  and the opening times  $O_1, \dots, O_{n+1}$ , respectively. Figure 3.7 confirms that the times are uncorrelated. An extensions to statistical tests for uncorrelation or independence might be possible as well. However, note that this confirmation and also the upcoming confirmation of the exponential distributions are only given for events with residence times between  $a = 0.08$  ms and  $b = 0.2$  ms, for all other lengths it is just an assumption.

By construction,  $C_1, \dots, C_n$  have density

$$f_{C_1}(x) = \frac{\lambda_C \exp(-\lambda_C x) \mathbb{1}_{x \in [a, b]}}{\exp(-\lambda_C a) - \exp(-\lambda_C b)}. \quad (3.34)$$

The maximum likelihood and the moment estimator  $\hat{\lambda}_C$  for  $\lambda_C$  coincide and can be computed by maximizing the likelihood function by the quasi-Newton method BFGS published simultaneously by Broyden (1970); Fletcher (1970); Goldfarb (1970); Shanno (1970). This is performed in R by the function `optim` with method 'BFGS' and start value  $\bar{C}$ . Its good performance is confirmed by simulations (not displayed). With this estimate at hand we can verify the distribution assumption for  $C_1, \dots, C_n$  derived from the Markov model by tests, for instance by the Kolmogorov-Smirnov test, but note that its critical value have to be corrected for the prior estimation of the rate (Lilliefors, 1969; Babu and Rao, 2004). Also here, we only give an empirical verification by comparing visually the estimated exponential distribution with the histogram of the times  $C_1, \dots, C_n$ , see Figure 3.8a. For this data set we estimate an average residence time of  $1/\hat{\lambda}_C = 0.0391$  ms.

Since the residence times are much shorter than the times between two events we make

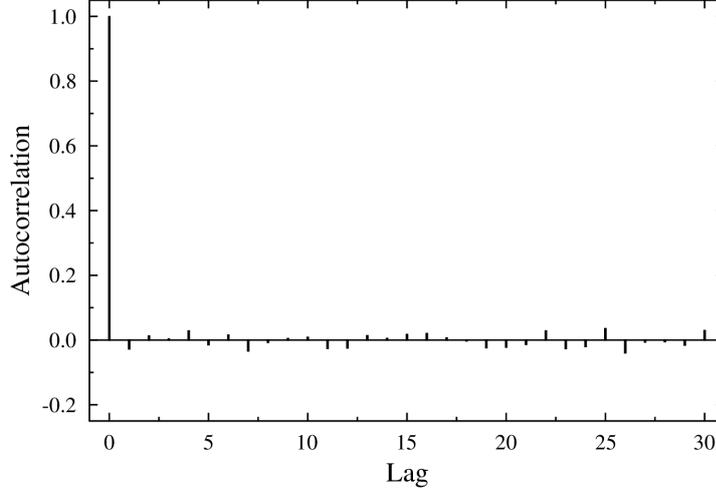


Figure 3.7.: Empirical autocorrelations  $\gamma_0, \dots, \gamma_{30}$ . The values close to zero confirm uncorrelated times.

the approximation  $O_i \approx \sum_{l=j_{i-1}+1}^{j_i} \tilde{O}_l$ . Let  $p := p(\lambda_C) := \mathbb{P}(\tilde{C}_1 \in [a, b])$  be the probability that a blockage event is observable. Then, each  $O_i$  is the sum of  $N$  independent exponentially distributed random variables, with rate  $\lambda_O$ , where  $N$  is a geometrically distributed random variable with probability  $p$  and independent of the exponentially distributed random variables. It follows that  $O_1, \dots, O_{n+1}$  are still exponentially distributed, but with rate  $p\lambda_O$ , see Proposition 23. This rate is estimated by the usual maximum likelihood estimator for exponentially distributed observations  $1/\bar{O}$  and the probability  $p$  is estimated by  $\hat{p} = p(\hat{\lambda}_C)$ . Thus, the rate  $\lambda_O$  is estimated by  $\hat{\lambda}_O = 1/(\hat{p}\bar{O}) = 17.0744$  Hz. And the distribution assumption can be checked in the same way as for the residence times, see Figure 3.8b.

In addition to estimation also confidence intervals can be obtained. For  $\lambda_O$  an exact confidence interval is given by  $[\chi_{2n}^2(\alpha/2)/\sum_{i=1}^n O_i, \chi_{2n}^2(1-\alpha/2)/\sum_{i=1}^n O_i]/p$ , with  $\chi_t^2(\alpha)$  the  $\alpha$ -quantile of a chi-squared distribution with  $t$  degrees of freedom. Replacing  $p$  by the estimate  $\hat{p}$  leads to an approximated confidence interval of  $[16.2134, 17.9572]$  Hz. For  $1/\lambda_C$  an asymptotic confidence interval of  $[0.0365, 0.0419]$  ms is obtained from the asymptotic normal distribution of  $\hat{\lambda}_C$  with an analytically computable variance, see Proposition 24.

In Section 5.1 we have measurements with more than one channel inserted in the membrane. In general, the analysis of such measurements is much more difficult due the interference of events and that it cannot be decided which channel opens or closes. In this special situation, where all events are so short that no inference is observed, the residence times are not altered and the blockage frequency for one channel is obtained by dividing the frequency of all channels by the number of channels. This number is determined by the unblocked conductance level divided by the estimated amplitude.

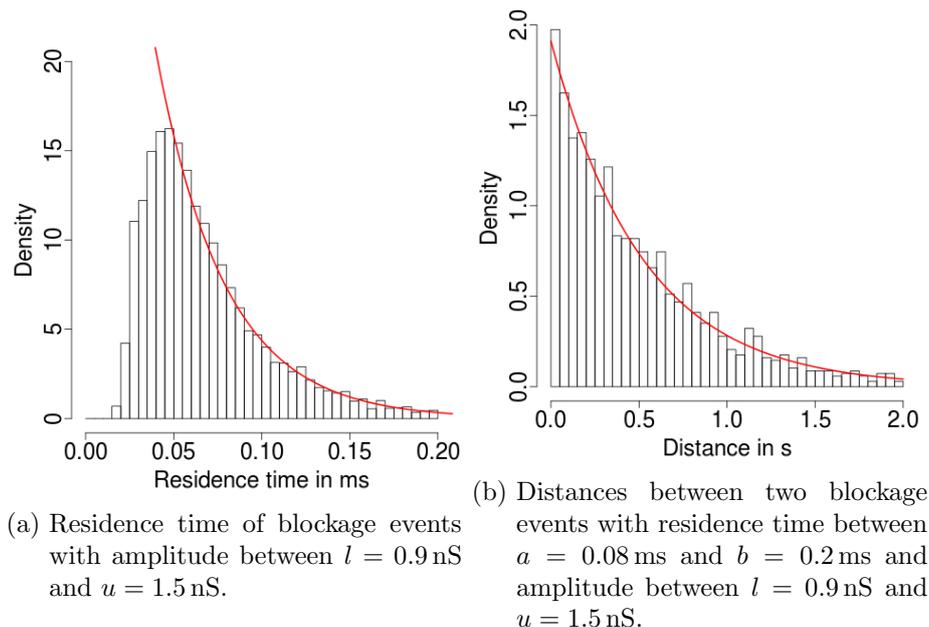


Figure 3.8.: Histograms of the residence times and distances between two blockage events with amplitude between  $l = 0.9$  nS and  $u = 1.5$  nS.

Finally, we have to discuss how to combine the results of several measurements, i.e., the same experiment is repeated  $m$  times under identical conditions. We model the amplitudes and the rates  $\lambda_C$  and  $\lambda_O$  as random variables drawn from an unknown distribution  $F$ . Uncertainty result from effects like how the channel is exactly positioned in the membrane or whether the membrane is bent a little bit. Since  $m$  is typically very small, in Section 5.1 only four, its distribution is hard to explore. In some experiments this error might be negligible, but for the PorB recordings it dominates, since  $n$  is rather large such that the estimation error for the rates is small. To visualize results the empirical mean and standard deviation of the estimated amplitudes and rates are used, although it cannot be ensured that the first and second moment of  $F$  exists. Nonetheless, two different measurements, e.g., measurements from two different channels, like in Section 5.1 of the PorB wild type and a mutant, can be compared, with statistically strict conclusions, by non-parametric tests like the Wilcoxon signed-rank test.

### 3.9. Simulations

In this section we present the simulation study from (Pein et al., 2017d) in which we compared JULES with the slope thresholding method TRANSIT (VanDongen, 1996) and the multiresolution method J-SMURF (Hotz et al., 2013). In addition, we report in this work results for JILTAD. Besides the version of JILTAD for homogeneous noise, we also include its version for heterogeneous noise to examine how much detection power is lost when the assumption of homogeneous noise is missing. Moreover, we perform similar simulations with heterogeneous noise. For this model we focus on JILTAD assuming

heterogeneous noise, since to our best knowledge no other non-parametric method is known that takes into account the filter explicitly. For illustration of missing heterogeneous noise we also include JULES in one simulation.

The purpose of the simulation study is threefold. First of all, we assess the ability of the methods to detect and estimate an isolated peak. Then, we identify the minimal distance at which the methods are able to separate two consecutive peaks. Finally, although none of the methods rely on a hidden Markov model assumption, but since a hidden Markov model is a very common assumption to analyze ion channel recordings, we examine their ability to recover the parameters of a Hidden Markov model with one open state and two closed flickering states, in particular how well the two flickering states can be separated. All of these simulations focus on estimating events that are short in time, since this is the most challenging task due to the filter and a satisfying estimation of events on larger scales but with a small amplitude was already shown in (Hotz et al., 2013) by J-SMURF which behaves on larger scales similar to JULES and JILTAD.

### 3.9.1. Data generation

We generate all signals and observations accordingly to the homogeneous and heterogeneous ion channel models we described in Section 3.1 such that they are in line with the measured data we analyze in Sections 3.10 and 5. This means in particular that amplitudes, dwell times and noise levels of the generated observations are chosen such that they are similar to those of the analyzed datasets. We also simulate a 4-pole Bessel filter with 1 kHz cut-off frequency and sample the observation at 10 kHz.

The expectation of the observations, given by the convolution of the signal with the truncated kernel  $F_L$  of the lowpass Bessel filter, can be computed explicitly by (3.6). For the errors we use for the homogeneous model the Durbin-Levinson algorithm (Brockwell and Davis, 2006, proposition 5.2.1) to compute the coefficients of the moving average process corresponding to the desired covariance structure. For the heterogeneous model we oversample the observations by a factor of 100, i.e., we generate 100 times as many independent Gaussian observations, discretize the filter accordingly, compute a discrete convolution and rescale the observations such that they have the desired variance.

### 3.9.2. Parameter choices

We use the following specifications in these simulations and in the real data applications. For the methods J-SMURF, JULES and JILTAD we apply their implementations in the R package *clampSeg* (Pein et al., 2017b). To control the probability to overestimate the number of change-points we choose a conservative significance level of  $\alpha = 0.05$ . The correlation matrix is regularized with parameter  $\gamma^2 = 1$ . And, as mentioned before, we truncate the kernel and autocorrelation function of the filter at  $L = 11$  which corresponds to the fact that the autocorrelation function is below  $10^{-1}$  afterwards. All of these choices are the default parameters. For TRANSIT we use the R function `transit` in the package

*stepR* (Pein et al., 2017a) with default parameters.

### 3.9.3. Isolated peak with homogeneous noise

In this section we examine in simulations with 4 000 observations the detection and estimation of an isolated peak. We consider a situation that is comparable to the analyzed data in Section 3.10. More precisely, we simulate a signal with function values  $m_0 = m_2 = 40$  and  $m_1 = 20$  and change-points at  $\tau_1 = 2000/f_s$  and  $\tau_2 = (2000 + \ell)/f_s$ , c.f. (3.1) and Figure 3.9. The standard deviation of the errors  $\epsilon_i$  in (3.4) is  $\sigma_0 = 1.4$ . We are interested in the performance of the methods in detecting the peak and estimating the change-point locations  $\tau_1$  and  $\tau_2$  and the function value  $m_1$  as a function of  $\ell$ , the length of the peak relative to the sampling rate  $f_s$ .

For  $\ell = 3$ , Figure 3.9 shows an example of simulated observations, estimations and their convolutions with the Bessel filter in a neighborhood of the peak. Figure 3.10 shows for  $\ell \in \{0.1, 0.2, \dots, 5\}$  in how many simulations the signal is correctly identified, i.e., only the two change-points of the peak and no other change-point is detected. Finally, Tables 3.1-3.3 summarize our results based on 10 000 repetitions for  $\ell = 2, 3, 5$ . More precisely, we define a peak as detected if there exists a  $j$  such that  $|\hat{\tau}_j - \tau_1| < L/f_s$  and  $|\hat{\tau}_{j+1} - \tau_2| < L/f_s$  as a detected change-point is shifted at most  $L/f_s$  by the filter. If only one change-point but not a peak is within these boundaries we do not count it as a true detection, but also not as a false positive, whereas all other change-points are counted as false positives. For the estimated change-point locations and function values we only take into account simulations in which the peak is detected and report the mean squared error, the bias and the standard deviation. We also report trimmed versions for the estimated function values, where we compute these error quantities only based on simulations with estimated function value between 0 and 40.

Table 3.1.: Performance of JULES, JILTAD, TRANSIT and J-SMURF in estimating a signal with an isolated peak having function values  $m_0 = m_2 = 40$ ,  $m_1 = 20$  and change-points at  $\tau_1 = 0.2$  and  $\tau_2 = \tau_1 + \ell/f_s$ ,  $\ell = 2, 3, 5$ . The standard deviation of the error is  $\sigma_0 = 1.4$ . Results are based on 10 000 pseudo samples.

Method	Length ( $\ell$ )	Correctly identified (%)	Detected (%)	False positive (Mean)
JULES	2	65.17	65.17	0.0290
JILTAD	2	95.99	96.16	0.0029
JILTAD (het.)	2	87.09	87.18	0.0098
TRANSIT	2	0.01	96.02	19.9692
J-SMURF	2	0.00	0.00	0.1998
JULES	3	99.82	99.82	0.0004
JILTAD	3	99.95	100.00	0.0010
JILTAD (het.)	3	99.70	99.78	0.0037
TRANSIT	3	0.02	74.65	19.8013
J-SMURF	3	0.00	0.00	0.1846
JULES	5	100.00	100.00	0.0000
JILTAD	5	99.98	100.00	0.0004
JILTAD (het.)	5	99.94	99.97	0.0009
TRANSIT	5	0.00	98.20	19.8484
J-SMURF	5	0.00	0.00	0.1892

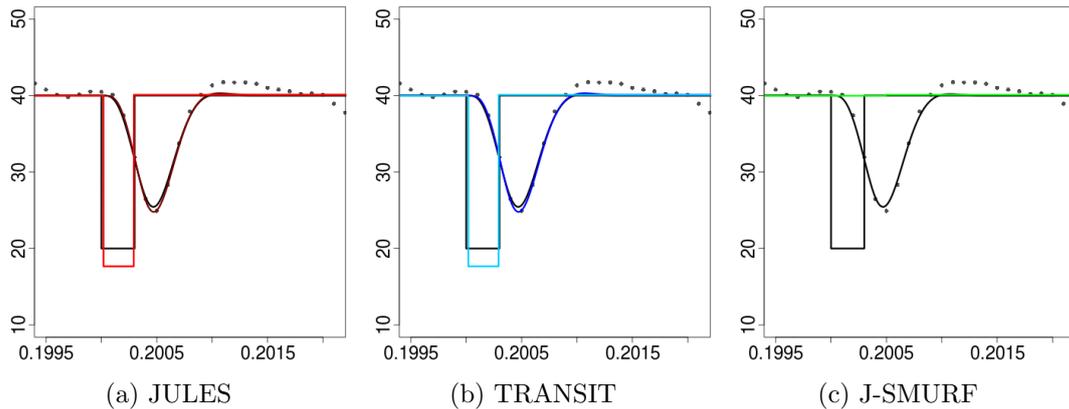


Figure 3.9.: Simulated observations (grey points), true signal  $\mu$  (black) and estimations by JULES (red), TRANSIT (blue) and J-SMURF (green). The estimations by JILTAD when assuming homogeneous noise and when assuming heterogeneous noise are identical with the estimation by JULES. We also show the convolutions of the true signal (black) and of the estimations by JULES (darkred) and TRANSIT (lightblue) with the lowpass 4-pole Bessel filter. JULES and JILTAD provide a very accurate estimation, whereas TRANSIT shifts the two change-points to the right and estimates a too small amplitude and J-SMURF misses the peak.

We found that the detection power of JULES increases with the length  $\ell$  in range between 1.1 to 3.1 times the sampling rate and is (almost) one for larger lengths while at the same time almost no false positives are detected, resulting in a correct identification with a probability of almost one for lengths a little bit larger than one quarter of the filter length. Taking into account the convolution for testing, as JILTAD does, improves the detection power noticeable, its detection power increases between 0.8 and 2.3. Allowing heterogeneous noise decreases the detection power mildly with an increase between 1 and 2.8, but the detection power of JILTAD assuming heterogeneous noise is nonetheless still larger

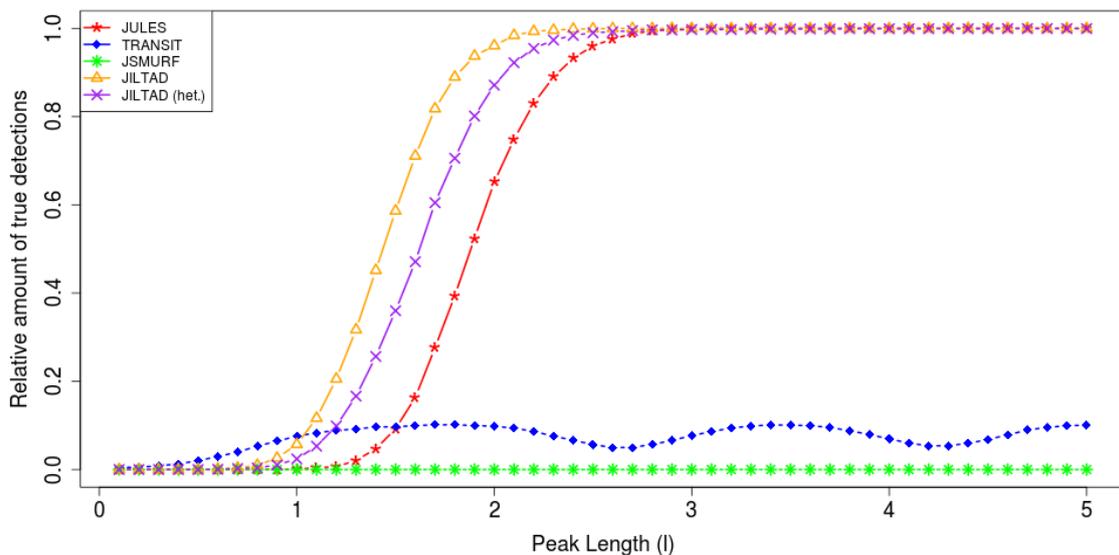


Figure 3.10.: Correct identification rates of the methods for an isolated peak with length  $\ell$ . Curves are based on 10000 repetitions, each.

Table 3.2.: Performance of JULES, JILTAD and TRANSIT in estimating a signal with an isolated peak having function values  $m_0 = m_2 = 40$ ,  $m_1 = 20$  and change-points at  $\tau_1 = 0.2$  and  $\tau_2 = \tau_1 + \ell/f_s$ ,  $\ell = 2, 3, 5$ . The standard deviation of the error is  $\sigma_0 = 1.4$ . Results are based on 10000 pseudo samples and are given as multiples of the sampling rate  $f_s = 10^4$ . We excluded J-SMURF, since this method did not detect peaks of these lengths.

Method	Length ( $\ell$ )	$f_s^2 \text{MSE}(\hat{\tau}_1)$	$f_s \text{BIAS}(\hat{\tau}_1)$	$f_s \text{SD}(\hat{\tau}_1)$	$f_s^2 \text{MSE}(\hat{\tau}_2)$	$f_s \text{BIAS}(\hat{\tau}_2)$	$f_s \text{SD}(\hat{\tau}_2)$
JULES	2	0.4022	-0.1047	0.6255	0.2677	0.0587	0.5141
JILTAD	2	0.2980	0.0141	0.5458	0.2979	-0.0131	0.5457
JILTAD (het.)	2	0.4178	-0.0638	0.6433	0.3872	0.0625	0.6191
TRANSIT	2	9.3427	2.6218	1.5713	24.1103	4.7920	1.0709
JULES	3	0.1170	0.0044	0.3420	0.1087	-0.0012	0.3297
JILTAD	3	0.1088	0.0070	0.3299	0.1094	-0.0024	0.3308
JILTAD (het.)	3	0.1334	0.0001	0.3653	0.1668	0.0080	0.4083
TRANSIT	3	13.0761	2.9386	2.1074	24.2032	4.7115	1.4162
JULES	5	0.0670	-0.0026	0.2588	0.0669	0.0025	0.2587
JILTAD	5	0.0670	-0.0026	0.2588	0.0669	0.0026	0.2587
JILTAD (het.)	5	0.0694	-0.0038	0.2635	0.0823	0.0051	0.2869
TRANSIT	5	15.4710	3.5622	1.6679	18.6559	4.0973	1.3670

than the one of JULES assuming homogeneous noise which illustrates how much detection power is lost by not taking into account the convolution. In comparison, TRANSIT has a slightly larger detection power for very small lengths, but detects much more false positives, resulting in almost no correctly identified signals. Moreover, its detection power fluctuates for larger lengths in an uncontrollable way. J-SMURF is not able to detect such small filtered peaks as it does not take into account the corresponding scales. However, J-SMURF detect in most simulations a single change-point instead of the peak. But, in roughly one fifth of the simulations the location is misestimated such that it is counted as a false positive. We also briefly investigated MDL from Gnanasambandam et al. (2017) (not displayed). To decrease the amount of false positives the authors suggested to assume a minimal length for the events. Although this might be problematic for real data applications with short events, we used for the simulations with  $l = 5$  the assumption of that length as prior information. But even under these conditions the number of events is heavily overestimated. More precisely, on average 98.87 change-points are estimated in the

Table 3.3.: Performance of JULES, JILTAD and TRANSIT in estimating a signal with an isolated peak having function values  $m_0 = m_2 = 40$ ,  $m_1 = 20$  and change-points at  $\tau_1 = 0.2$  and  $\tau_2 = \tau_1 + \ell/f_s$ ,  $\ell = 2, 3, 5$ . The standard deviation of the error is  $\sigma_0 = 1.4$ . Results are based on 10000 pseudo samples. We excluded J-SMURF, since this method did not detect peaks of these lengths.

Method	Length ( $\ell$ )	$\text{MSE}(\hat{m}_1)$	$\text{BIAS}(\hat{m}_1)$	$\text{SD}(\hat{m}_1)$	$\text{MSE}_{\text{Trim}}(\hat{m}_1)$	$\text{BIAS}_{\text{Trim}}(\hat{m}_1)$	$\text{SD}_{\text{Trim}}(\hat{m}_1)$
JULES	2	222978.6771	-83.6452	464.7745	32.5343	-0.6010	5.6726
JILTAD	2	364264.8273	-146.4294	585.5415	35.2761	-0.2986	5.9322
JILTAD (het.)	2	232908.4428	-89.4288	474.2749	34.9639	-0.1520	5.9115
TRANSIT	2	112.7163	10.1733	3.0367	99.3660	9.7529	2.0609
JULES	3	552.8490	-1.0170	23.4919	12.3488	-0.7219	3.4393
JILTAD	3	1038.9324	-1.3235	32.2069	12.3993	-0.7256	3.4459
JILTAD (het.)	3	1041.2124	-1.2786	32.2441	12.5987	-0.6855	3.4828
TRANSIT	3	69.6075	7.0327	4.4891	47.2984	6.2367	2.8988
JULES	5	2.7763	-0.1081	1.6628	2.7763	-0.1081	1.6628
JILTAD	5	2.7762	-0.1079	1.6628	2.7762	-0.1079	1.6628
JILTAD (het.)	5	2.8281	-0.1017	1.6787	2.8281	-0.1017	1.6787
TRANSIT	5	26.8554	2.0811	4.7462	11.5351	1.4210	3.0849

4 000 observations instead of the two change-points of the peak. Note that the simulated observations are rather heavily filtered and as noted by Gnanasambandam et al. (2017) their approach works better for unfiltered or hardly filtered recordings.

JULES and JILTAD estimate almost all detected peaks with high accuracy which increases with the length of the peak. For short peaks in a few cases the amplitude is heavily over-estimated, resulting in large mean squared error, bias and standard deviation, but much smaller trimmed values. Notably, the results are very similar for JULES and both versions of JILTAD. In many simulations their estimations are even identical as for instance in Figure 3.9a, since they rely on almost the same deconvolution approach. Moreover, we found that the convolved estimations fit the observations always very well, see Figure 3.9a for an illustration. In comparison, Figure 3.9b and the values for the bias in Tables 3.2 and 3.3 show that TRANSIT miss-estimates the peak systematically, in particular the amplitude is underestimated if the peak is smoothed by the filter. MDL shows a very similar behavior as also this approach does not take into account the filter explicitly, see Figure 3.26 for an illustration. Results for J-SMURF are not reported, since this method did not detect the peaks.

The major difficulty of the deconvolution problem is that convolutions of signals with larger amplitude but smaller length can look very similar to the convolution of a signal with smaller amplitude but larger length. In fact, it is possible to show that the change-points locations  $\tau_i$  can by no method be estimated better than at  $1/\sqrt{n}$  rate, instead of the sampling rate  $1/n$  without convolution, see (Boysen et al., 2009a) for a similar setting and (Goldenshluger et al., 2006; Frick et al., 2014b) for further theoretical results for estimating a peak from filtered data. Note that the filter is rather short ranged with a filter length of eleven, but the peaks are very short in time with five or less observations and the signal to noise ratio is with  $20/1.4 \approx 14.29$  of medium size, since Gramicidin A has a small conductance in comparison to other proteins, but is measured at the Port-a-Patch with solvent-free lipid bilayers, which typically leads to a better signal to noise ratio as for instance measurements using black lipid membranes.

Similar results were obtained in additional simulation studies (not included) with different amplitudes and noise levels. For instance for the PorB recordings with ampicillin from Figure 1.1 we found that a peak has to be at least of length  $l = 4$  to be detected by JULES with probability almost one. Also simulations (not displayed) where we shift the two change-point locations by  $0.1/f_s, 0.25/f_s, 0.5/f_s, 0.75/f_s$  and  $0.9/f_s$  lead to almost identical results for JULES and JILTAD. This confirms their ability to estimate correctly the location and amplitude of changes between sampling points.

#### 3.9.4. Isolated peak with heterogeneous noise

In this section we carry out simulations similar to them in the previous section, but with heterogeneous noise. More precisely, in accordance with the model in Section 3.1.2 and with the estimated values in Section 5.2 for the observations in Figure 3.3, we choose

function values  $m_0 = m_2 = 0$ ,  $m_1 = 0.32$ , variances  $s_0^2 = s_2^2 = 6.1 \cdot 10^{-5}$  and varying variance  $s_1^2 \in \{2 \cdot 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 2 \cdot 10^{-3}, 5 \cdot 10^{-3}\}$  to examine the influence of different noise levels. All other parameters, in particular the number of observations and the change-point locations are as before. We only report results for JILTAD assuming heterogeneous noise, since no other method is known that deals explicitly with this model. Results are shown in Tables 3.4-3.6.

Table 3.4.: Performance of JILTAD in estimating a signal with an isolated peak having change-points at  $\tau_1 = 0.2$  and  $\tau_2 = \tau_1 + \ell/f_s$ ,  $\ell = 2, 3, 5$ , function values  $m_0 = m_2 = 0$ ,  $m_1 = 0.32$ , variances  $s_0^2 = s_2^2 = 6.1 \cdot 10^{-5}$  and varying variance  $s_1^2 \in \{2 \cdot 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 2 \cdot 10^{-3}, 5 \cdot 10^{-3}\}$ . Results are based on 10 000 pseudo samples.

Method	Length ( $\ell$ )	Correctly identified (%)	Detected (%)	False positive (Mean)
$s_1^2 = 2 \cdot 10^{-4}$	2	99.96	100.00	0.0008
$s_1^2 = 5 \cdot 10^{-4}$	2	99.96	100.00	0.0008
$s_1^2 = 10^{-3}$	2	99.94	99.98	0.0010
$s_1^2 = 2 \cdot 10^{-3}$	2	99.07	99.11	0.0014
$s_1^2 = 5 \cdot 10^{-3}$	2	90.04	90.08	0.0042
$s_1^2 = 2 \cdot 10^{-4}$	3	99.97	100.00	0.0006
$s_1^2 = 5 \cdot 10^{-4}$	3	99.97	100.00	0.0006
$s_1^2 = 10^{-3}$	3	99.97	100.00	0.0006
$s_1^2 = 2 \cdot 10^{-3}$	3	99.93	99.96	0.0006
$s_1^2 = 5 \cdot 10^{-3}$	3	96.08	96.11	0.0024
$s_1^2 = 2 \cdot 10^{-4}$	5	99.95	100.00	0.0010
$s_1^2 = 5 \cdot 10^{-4}$	5	99.95	100.00	0.0010
$s_1^2 = 10^{-3}$	5	99.95	100.00	0.0010
$s_1^2 = 2 \cdot 10^{-3}$	5	99.94	100.00	0.0012
$s_1^2 = 5 \cdot 10^{-3}$	5	99.42	99.48	0.0018

Table 3.5.: Performance of JILTAD in estimating a signal with an isolated peak having change-points at  $\tau_1 = 0.2$  and  $\tau_2 = \tau_1 + \ell/f_s$ ,  $\ell = 2, 3, 5$ , function values  $m_0 = m_2 = 0$ ,  $m_1 = 0.32$ , variances  $s_0^2 = s_2^2 = 6.1 \cdot 10^{-5}$  and varying variance  $s_1^2 \in \{2 \cdot 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 2 \cdot 10^{-3}, 5 \cdot 10^{-3}\}$ . Results are based on 10 000 pseudo samples and are given as multiples of the sampling rate  $f_s = 10^4$ .

Method	Length ( $\ell$ )	$f_s^2 \text{MSE}(\hat{\tau}_1)$	$f_s \text{BIAS}(\hat{\tau}_1)$	$f_s \text{SD}(\hat{\tau}_1)$	$f_s^2 \text{MSE}(\hat{\tau}_2)$	$f_s \text{BIAS}(\hat{\tau}_2)$	$f_s \text{SD}(\hat{\tau}_2)$
$s_1^2 = 2 \cdot 10^{-4}$	2	0.0331	0.0092	0.1818	0.0381	-0.0076	0.1951
$s_1^2 = 5 \cdot 10^{-4}$	2	0.0515	0.0113	0.2267	0.0427	-0.0115	0.2062
$s_1^2 = 10^{-3}$	2	0.0677	0.0255	0.2590	0.0595	-0.0266	0.2424
$s_1^2 = 2 \cdot 10^{-3}$	2	0.1532	0.0935	0.3801	0.1570	-0.0942	0.3848
$s_1^2 = 5 \cdot 10^{-3}$	2	0.6628	0.3275	0.7454	0.6223	-0.3252	0.7188
$s_1^2 = 2 \cdot 10^{-4}$	3	0.0120	0.0001	0.1097	0.0117	0.0010	0.1083
$s_1^2 = 5 \cdot 10^{-4}$	3	0.0177	0.0040	0.1329	0.0177	-0.0022	0.1332
$s_1^2 = 10^{-3}$	3	0.0391	0.0181	0.1970	0.0388	-0.0152	0.1965
$s_1^2 = 2 \cdot 10^{-3}$	3	0.1569	0.0846	0.3870	0.1533	-0.0802	0.3833
$s_1^2 = 5 \cdot 10^{-3}$	3	1.1681	0.4668	0.9748	1.1058	-0.4515	0.9498
$s_1^2 = 2 \cdot 10^{-4}$	5	0.0070	-0.0014	0.0835	0.0084	0.0013	0.0914
$s_1^2 = 5 \cdot 10^{-4}$	5	0.0176	0.0036	0.1326	0.0189	-0.0037	0.1375
$s_1^2 = 10^{-3}$	5	0.0572	0.0217	0.2381	0.0591	-0.0215	0.2421
$s_1^2 = 2 \cdot 10^{-3}$	5	0.2473	0.0985	0.4874	0.2464	-0.0976	0.4867
$s_1^2 = 5 \cdot 10^{-3}$	5	2.0808	0.6363	1.2946	2.1605	-0.6520	1.3174

In most scenarios JILTAD has a good detection power and detects almost no false positives, see Table 3.4, only for a five times larger variance than in the real data and when the peaks are short significantly many events are missed. Notably, contrary to (Enikeeva et al., 2016), where the variance values are assumed to be known and at the peak larger

Table 3.6.: Performance of JILTAD in estimating a signal with an isolated peak having change-points at  $\tau_1 = 0.2$  and  $\tau_2 = \tau_1 + \ell/f_s$ ,  $\ell = 2, 3, 5$ , function values  $m_0 = m_2 = 0$ ,  $m_1 = 0.32$ , variances  $s_0^2 = s_2^2 = 6.1 \cdot 10^{-5}$  and varying variance  $s_1^2 \in \{2 \cdot 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 2 \cdot 10^{-3}, 5 \cdot 10^{-3}\}$ . Results are based on 10 000 pseudo samples.

Method	Length ( $\ell$ )	MSE( $\hat{m}_1$ )	BIAS( $\hat{m}_1$ )	SD( $\hat{m}_1$ )	MSE <sub>Trim</sub> ( $\hat{m}_1$ )	BIAS <sub>Trim</sub> ( $\hat{m}_1$ )	SD <sub>Trim</sub> ( $\hat{m}_1$ )
$s_1^2 = 2 \cdot 10^{-4}$	2	0.1320	0.0194	0.3628	0.1320	0.0194	0.3628
$s_1^2 = 5 \cdot 10^{-4}$	2	0.6953	0.0322	0.8333	0.1225	0.0215	0.3494
$s_1^2 = 10^{-3}$	2	2.3290	0.0888	1.5236	0.9686	0.0604	0.9824
$s_1^2 = 2 \cdot 10^{-3}$	2	21.7896	0.7801	4.6025	11.7570	0.5806	3.3795
$s_1^2 = 5 \cdot 10^{-3}$	2	294.6030	5.5640	16.2380	85.3154	3.6929	8.4668
$s_1^2 = 2 \cdot 10^{-4}$	3	0.0002	0.0009	0.0152	0.0002	0.0009	0.0152
$s_1^2 = 5 \cdot 10^{-4}$	3	0.0007	0.0018	0.0259	0.0007	0.0018	0.0259
$s_1^2 = 10^{-3}$	3	0.0023	0.0055	0.0473	0.0023	0.0055	0.0473
$s_1^2 = 2 \cdot 10^{-3}$	3	1.7338	0.0806	1.3143	0.6608	0.0603	0.8107
$s_1^2 = 5 \cdot 10^{-3}$	3	334.7102	5.0891	17.5739	62.1012	2.6210	7.4322
$s_1^2 = 2 \cdot 10^{-4}$	5	0.0001	0.0001	0.0077	0.0001	0.0001	0.0077
$s_1^2 = 5 \cdot 10^{-4}$	5	0.0003	0.0007	0.0179	0.0003	0.0007	0.0179
$s_1^2 = 10^{-3}$	5	0.0013	0.0033	0.0354	0.0013	0.0033	0.0354
$s_1^2 = 2 \cdot 10^{-3}$	5	0.0055	0.0158	0.0725	0.0055	0.0158	0.0725
$s_1^2 = 5 \cdot 10^{-3}$	5	203.0129	2.7499	13.9811	27.0030	1.1807	5.0608

than the background noise, the detection power is decreasing with increasing variance. Hence, for recordings with a small change in the expectation in relation to the variances but a large relative variance difference it would be valuable to use a different test statistic, for instance one that focuses on comparing the variances.

We remark that the scenario of a mostly open channel with short closings is slightly more difficult for JILTAD. More precisely, for function values with the same difference but variances  $s_0^2 = s_2^2 = 10^{-3}$  and  $s_1^2 = 6.1 \cdot 10^{-5}$  JILTAD misses most events of length  $l = 2$ , but detects 97% events of length  $l = 3$  and almost all of length  $l = 5$ . In Section 5.2 we also have two measurements with variances roughly  $2.5 \cdot 10^{-5}$  and 0.05 and an amplitude of 0.95. Notably, for these data sets event detection works very well, almost with probability one even for events of length  $l = 0.1$ . This can be explained by the larger amplitude of 0.95 and the smaller noise level for the closed state in which the channel is most of the time.

In Tables 3.5 and 3.6 we found that estimation of the change-point locations and the function value works well for variances similar to the real data, but has some issues when the variance of the peak is larger, in particular in the scenario of a five times larger variance. For such observations it might be desirable to take into account the heterogeneous variance. For smaller variances the results for estimating the change-point locations are better when the peak is longer, but for larger variances results are even worse when the peak is longer. An explanation might be two effects with opposite influences. The different function value, and also the relative variance difference which is not used by JILTAD, provide more information when the peak is longer, but then also the variance of the observations is larger which hinders estimation. Estimation of the function value  $m_1$  is always more accurate when the peak is longer. It seems that here the first influence dominates. In the reversed scenario of short closing events, estimation of the change-point locations has a large bias and variance for  $\ell = 2$ , but improves dramatically with the length of the peak,

since in this setting also the variance of the observations gets smaller when the length of the peak increases. In the scenario of a larger variance heterogeneity and a larger amplitude a small bias the estimation of all parameters is slightly biased which does not improve in the length of the peak for the change-point locations, but for the function value  $m_1$ . All in all, these simulations confirm that JILTAD performs well for observations comparable to them in Section 5.2, although the local tests ignore the correlation structure and in the deconvolution step we still use the deconvolution approach designed for homogeneous noise.

### 3.9.5. Separation of two consecutive peaks

In this simulation study we are interested in the performance of JULES and JILTAD in separating two consecutive peaks as a function of  $d$ , the distance between them. We only include JULES and JILTAD as these methods rely on deconvolution and, hence, separation signifies an issues for them. Separation is necessary in the detection step, i.e., two peaks have to be detected, but in particular in the deconvolution step as we are only able to deconvolve both peaks individually if they are isolated. To this end, we consider a signal with change-points at  $\tau_1 = 2000/f_s$ ,  $\tau_2 = \tau_1 + 5/f_s$ ,  $\tau_3 = \tau_2 + d$  and  $\tau_4 = \tau_3 + 5/f_s$ , with  $\tau_0 = 0$  and  $\tau_{\text{end}} = 4000/f_s$ , and function values  $m_0 = m_2 = m_4 = 40$  and  $m_1 = m_3 = 20$ . The standard deviation is assumed to be constant with  $\sigma_0 = 1.4$ . Here we fixed the length of the consecutive peaks at  $5/f_s$ , since want to focus on separation and we found in Section 3.9.3 that peaks of this size are detected with (almost) probability one.

We identified three outcomes, all illustrated in Figure 3.11. First of all, perfect separation, i.e., four change-points are detected and the local deconvolution yields estimates for all parameters. Secondly, separation fails in the detection step, i.e., only 2 change-points are detected and identified as one peak whose parameters are obtained by deconvolution. Finally, separation fails in the deconvolution step, i.e., two peaks are detected but the distance is so small that the deconvolution method cannot separate them, in other words, no long segment is in between.

Figures 3.12-3.14 show the frequency at which each scenario occurred as a function of  $d$ , the distance between the two peaks. For each value of  $d = \{1, 2, \dots, 70\}$  we performed 10 000 simulations. JULES potentially fails to detect both peaks when  $d \leq 12$ , but the frequency decreases rapidly with  $d$ . For  $d \in [4, 31]$  detection is often possible, but not the separation in the deconvolution, whereas for  $d > 33$  this scenario is no longer observed. Finally, separation of the two peaks is possible with high probability as soon as the distance between them is at least  $32/f_s$ , roughly three times the filter length. This is equal to the minimal distance of a long segment of ten observations plus the two shifts on the left and right side of the segment by  $L = 11$  to take the filter into account. This corresponds to 3.2 ms for the Gramicidin A recordings, whereas the estimated average distance is with  $1/3.28 \text{ s} \approx 0.3 \text{ s}$ , see Section 3.10, much larger.

For JILTAD assuming homogeneous noise separation in the deconvolution step fails poten-

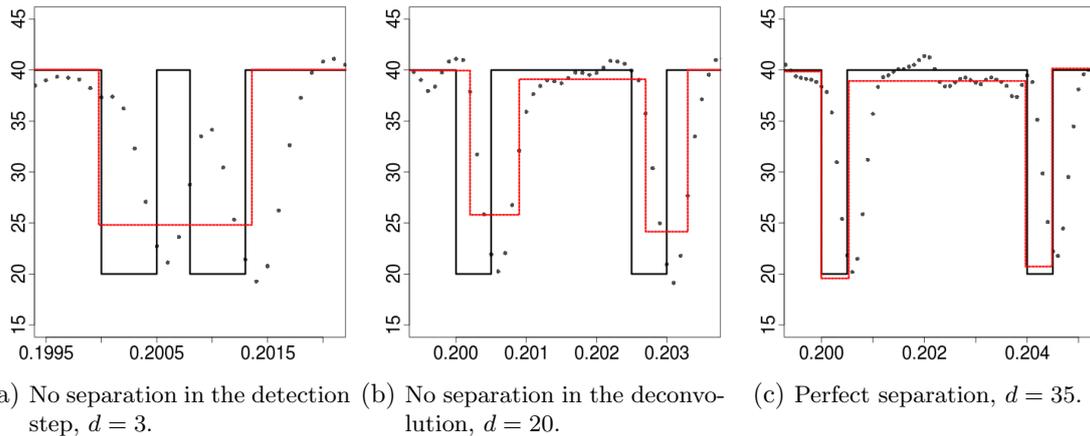


Figure 3.11.: Examples for the different outcomes. Simulated observations (grey points), true signal  $\mu$  (black) and estimation (red), here exemplary by JULES.

tially for distances  $d$  between 2 and 5 and between 19 and 32, whereas for distances  $d \geq 33$  separation was almost always perfect. For all other distances the method failed to detect two peaks. For very short distances only one change-point is detected by J-SMURF and we can apply local tests which quickly detect two peaks, in particular faster than JULES, but separation fails in the deconvolution step. But for  $d \geq 6$  J-SMURF detects two changes resulting in a segment on which local tests can not be applied. For  $d > 19$  J-SMURF still detects two change-points, but the distance between them is large enough such that local tests can be applied again. These tests detect the two peaks resulting in another segment with no separation in the deconvolution step and for  $d \geq 32$  in two separated peaks. We stress that the same length is required for JULES and JILTAD to separate two peaks,

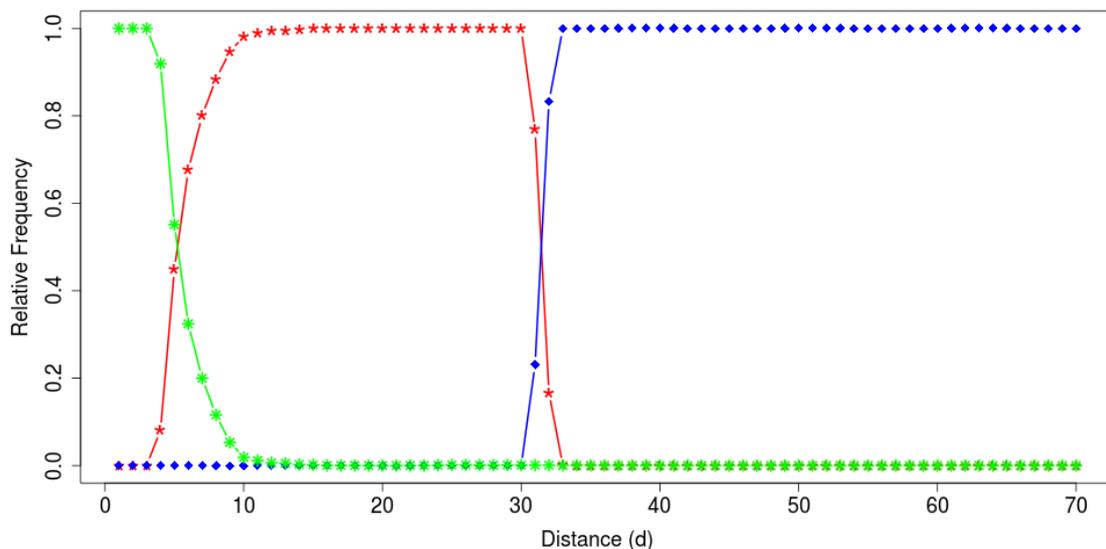


Figure 3.12.: Results for JULES in estimating two consecutive peaks separated by distance  $d$ . Its frequencies for no separation in the detection step (green), for successful detection, but no separation in the deconvolution step (red) and for successful detection and deconvolution (blue). Results are based on 10 000 simulations for each value of  $d$ .

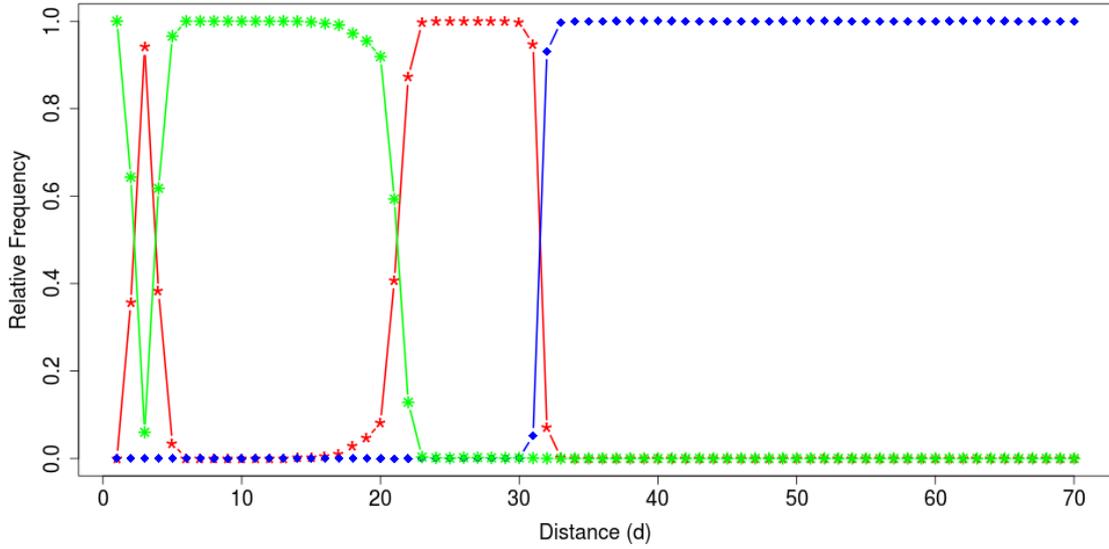


Figure 3.13.: Results for JILTAD assuming homogeneous noise in estimating two consecutive peaks separated by distance  $d$ . Its frequencies for no separation in the detection step (green), for successful detection, but no separation in the deconvolution step (red) and for successful detection and deconvolution (blue). Results are based on 10 000 simulations for each value of  $d$ .

but JULES indicates more reliably that more than one peak occurred, but could not be deconvolved.

When heterogeneous noise is assumed, J-SMURF does not detect two change-points for short distances and also the detection of two peaks by the local tests requires a longer distance between them. Thus, the two peaks are not detected until  $d = 6$  and separation

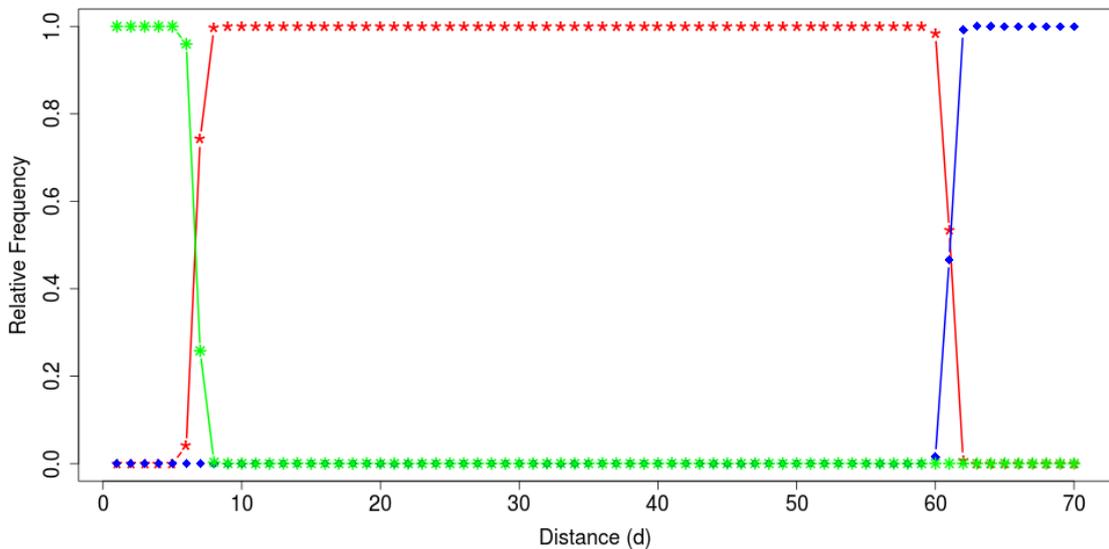


Figure 3.14.: Results for JILTAD assuming heterogeneous noise in estimating two consecutive peaks separated by distance  $d$ . Its frequencies for no separation in the detection step (green), for successful detection, but no separation in the deconvolution step (red) and for successful detection and deconvolution (blue). Results are based on 10 000 simulations for each value of  $d$ .

in the deconvolution step requires a much longer distance of  $d \geq 62$ . We do not show results for simulations with heterogeneous noise, since the results are identical as separation depends on the method and distance between the peaks but not on the noise.

### 3.9.6. Hidden Markov model

In this section we simulate data from a three state hidden Markov model. None of the considered methods rely on a hidden Markov assumption, but as such a model is often assumed for ion channel recordings, it is instructive to investigate these methods in such a scenario. More precisely, we simulate one open state with 40 pS and two closed states with 20 pS and  $(20 + \Delta)$  pS, i.e., the amplitudes of the two flickering states, 20 pS and  $(20 - \Delta)$  pS, differ by  $\Delta$  pS. The standard deviation of the errors  $\epsilon_i$  in (3.4) is  $\sigma_0 = 1.4$  pS. The dwell time in the open state is exponentially distributed with rate 2.5 Hz and the channel switches to both closed states equally likely. For both closed states the channel reopens quickly with rate 0.8 kHz. We generate five time series with 600 000 observations, each.

In the following we analyze these data as outlined in Section 3.8. In accordance with the definition of a short segment in Section 3.6 we define a closing event as a flickering event if its dwell time is smaller than or equal to 2.6 ms. But, results are qualitative the same, amplitude histograms are even quantitative very similar, if we increase or decrease this threshold within a reasonable range. We do not show results for J-SMURF, since this method is not able to detect such short flickering events, c.f. Figure 3.10. Figure 3.15 shows histograms of the estimated amplitudes of flickering events with a small bin width of 0.5, visually chosen, to already see small indications of two different levels.

In Table 3.3 we found that the estimation of the function values by JULES and JILTAD have a large mean squared error, mainly due to heavy overestimations in rare cases. Nevertheless, we see in this simulation that both methods estimate and separate the two amplitudes very well. For  $\Delta = 2$  a tiny indication of two different amplitudes is seen, for  $\Delta = 3$  the two peaks are distinct and for larger differences the two different states are clearly detected. In comparison, TRANSIT finds some smaller amplitudes and hence the separation is slightly less clear, but still possible for  $\Delta \geq 3$ . We already saw in Figure 3.9 and Table 3.3 that TRANSIT estimates a too small amplitude, since short peaks are smoothed by the filter.

Figure 3.16 shows the dwell times in the closed state and an exponential fit. For the fit we take into account events with amplitude between  $l = 10$  pS and  $u = 30$  pS and dwell times between  $a = 0.24$  ms and  $b = 2.6$  ms, since we cannot detect events with a dwell time smaller than 0.24 ms reliably and they would disturb the analysis. For this case we use  $\Delta = 0$ , i.e., consider only one closed state.

Apart from the fact that all methods miss extremely short events,  $< 0.3$  ms, which coincides with Figure 3.10, JULES and JILTAD confirm an exponential distribution and estimate with 0.80 kHz and 0.82 kHz, respectively, the rate very accurately. TRANSIT detects

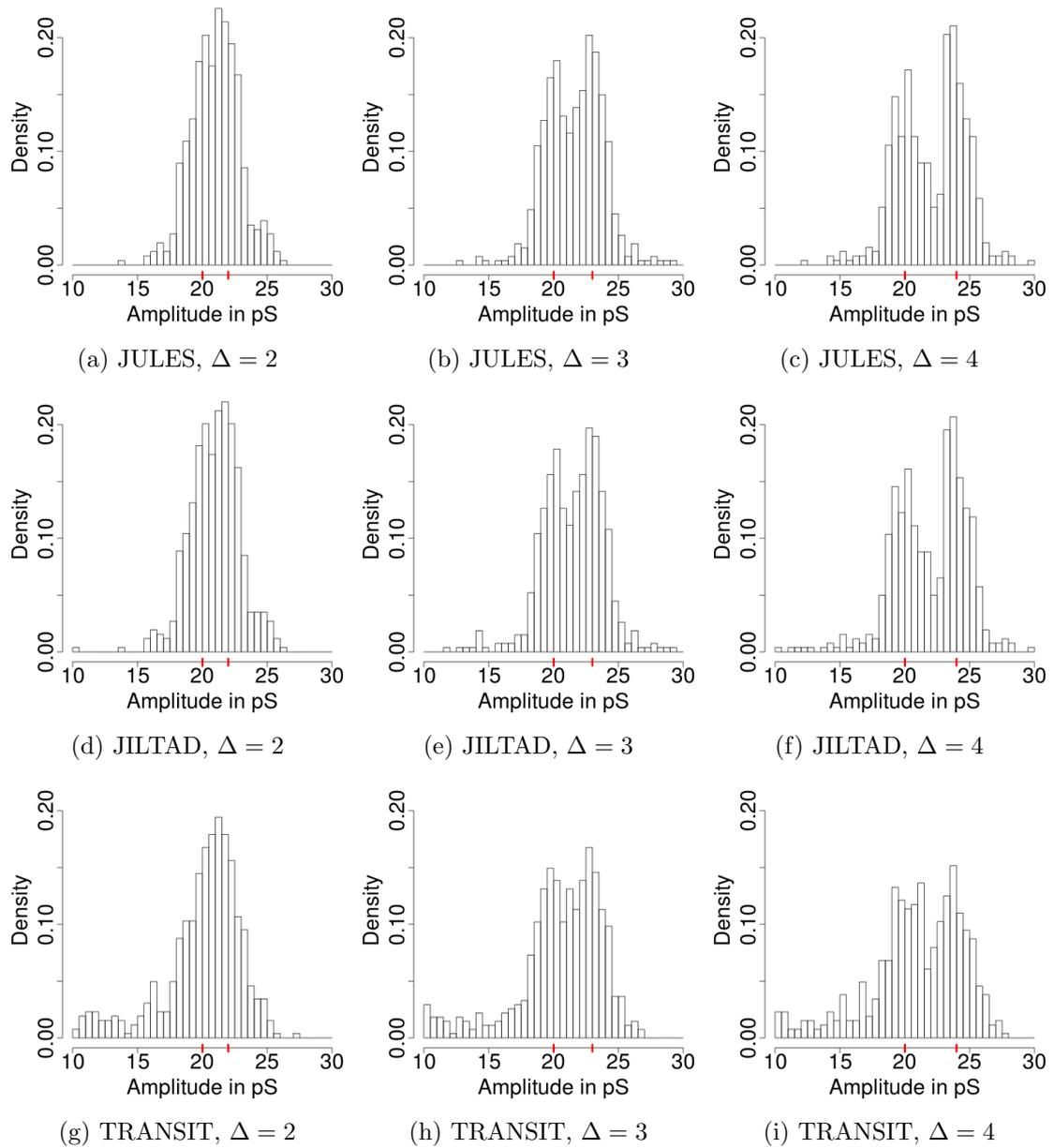


Figure 3.15.: Histograms of the amplitudes of detected events for various  $\Delta$  and dwell time below 2.6 ms. Red tick marks are the true amplitudes, both with equal probability.

additional spurious events with lengths between 0.4 ms and 0.5 ms and hence overestimates the rate with 0.99 kHz slightly.

Figure 3.17 shows the distance between two flickering events which coincides with the dwell time in the open state if other closing events are considered as spurious events. We include events with a distance between 0.032s and 1s, as Figures 3.12-3.13 show that JULES and JILTAD are not able to separate peaks with a smaller distance.

JULES, JILTAD and TRANSIT suggest an exponential distribution and estimate with 2.67 Hz, 2.70 Hz and 2.49 Hz, respectively, roughly the correct rate. All in all, we found

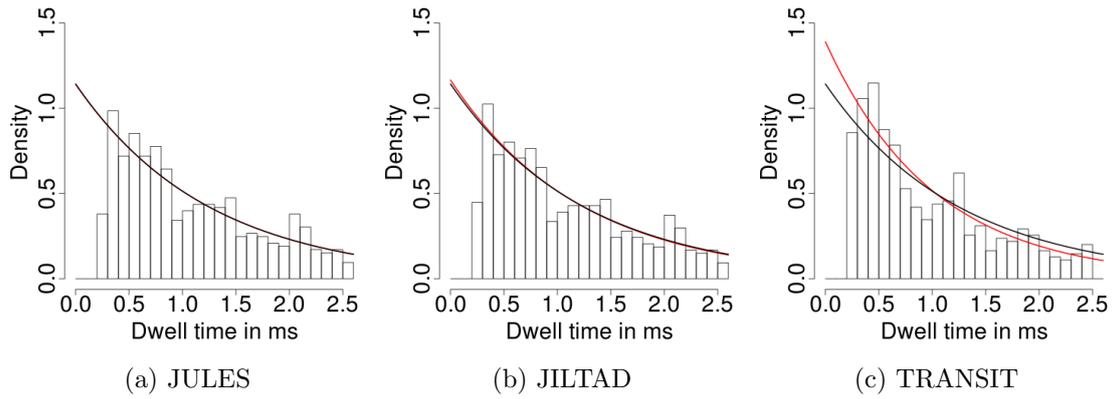


Figure 3.16.: Histograms of the dwell time in the closed state for  $\Delta = 0$  for closing events with amplitude between 10 and 30 pS together with the true exponential distribution with rate 0.8 kHz (black) and exponential fits (red) taking into account missed events. We rescaled all lines such that the area under them are standardized to one to make them comparable to the histograms.

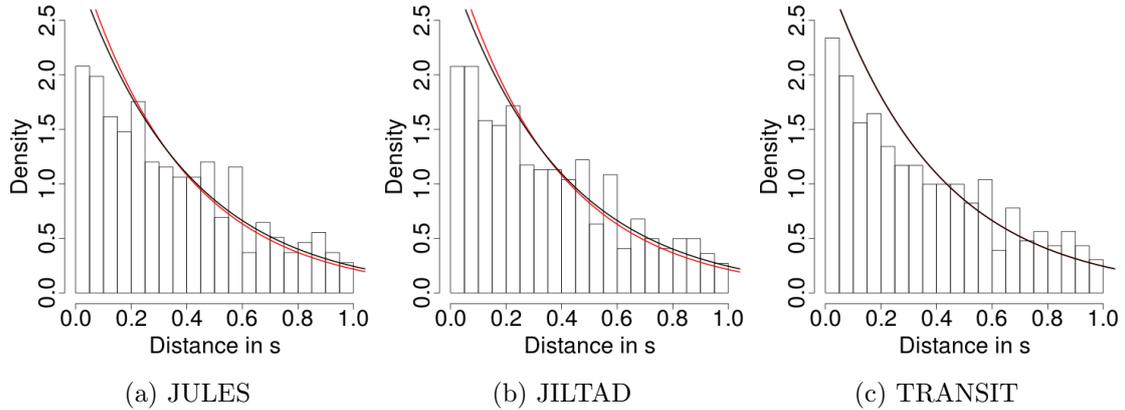


Figure 3.17.: Histograms of the distance between two flickering events, i.e., events with amplitude between  $l = 10$  pS and  $u = 30$  pS and dwell time between  $a = 0.24$  ms and  $b = 2.6$  ms, for  $\Delta = 0$  together with the true exponential distribution with rate 2.5 Hz (black) and exponential fits corrected for missed events (red). We rescaled all lines such that the area under them are standardized to one to make them comparable to the histograms.

that JULES and JILTAD recover the hidden Markov model well, while TRANSIT tends to underestimate the amplitude.

### 3.9.7. Hidden Markov model with heterogeneous noise

We now extend these simulations to heterogeneous noise. We choose a closed state with 0.04 nS and two open states with 0.36 nS and  $(0.36 + \Delta)$  nS, i.e., the amplitudes that differ by  $\Delta$  nS. The standard deviation is  $\sqrt{6.1 \cdot 10^{-5}}$  nS in the closed state and  $\sqrt{10^{-3}}$  nS in both open states. The dwell time in the closed state is exponentially distributed with rate 5 hertz and the channel switches to both closed states equally likely. For both open states

the channel closes quickly with rate 0.4 kHz. Note that up to a second open state these values are similar to the recordings in Figure 3.3. We generate again five time series with 600 000 observations, each.

We analyze these data sets with JILTAD assuming heterogeneous noise. Additionally, to illustrate ignoring heterogeneous noise we include JULES which assumes homogeneous noise. We do not show results for H-SMUCE, which could serve as an illustration for ignoring the filtering, since its lack of detection power on small scales was already clearly visible in Figure 3.4. Figure 3.18 shows histograms of the estimated amplitudes of all events with an amplitude between  $l = 0.2$  nS and  $u = 0.5$  nS with a small bin width of 0.005 nS.

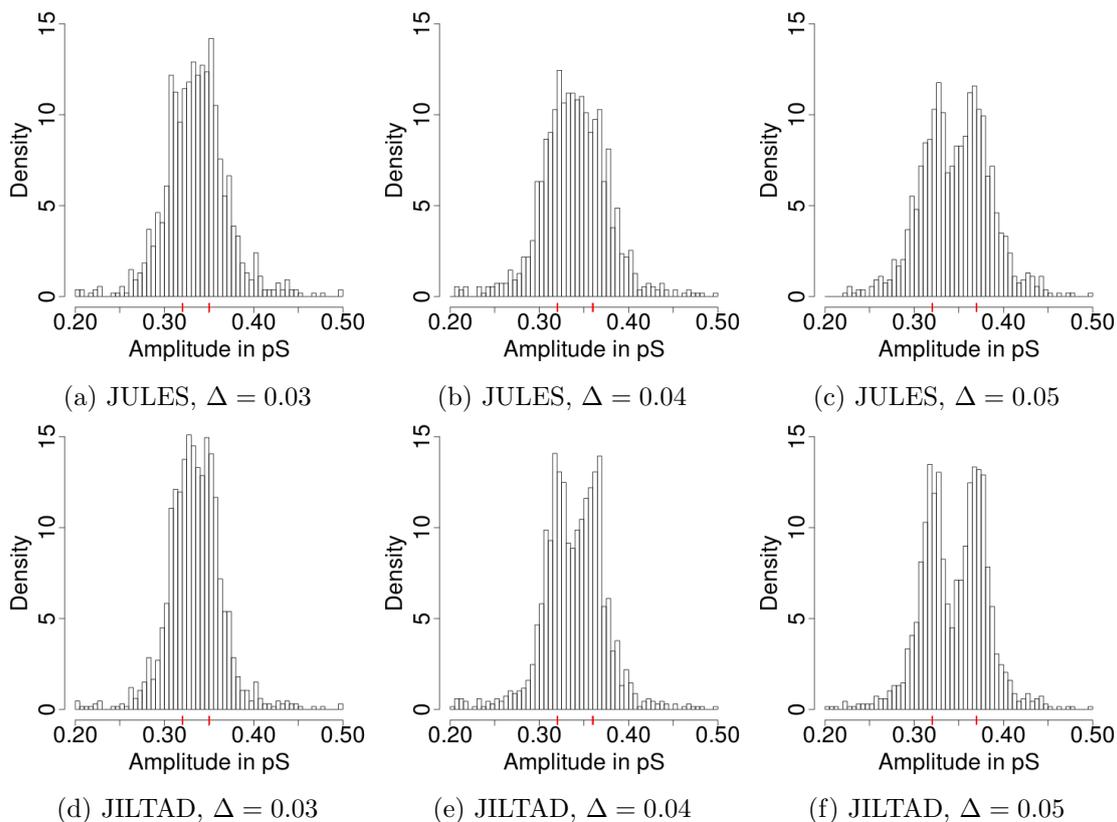


Figure 3.18.: Histograms of the estimated amplitudes for various  $\Delta$ . Red tick marks are the true amplitudes, both with equal probability.

Using JULES we only find two open states when  $\Delta = 0.05$ . Contrary, with JILTAD we already see for  $\Delta = 0.03$  an unreliable tendency and a clear detection of two states for  $\Delta \geq 0.04$ . This confirms that the detection of additional events as in Figure 3.4 hinders indeed the analysis and unsurprisingly taking into account the heterogeneity explicitly improves results.

We continue with an analysis of the dwell time in the open state and the distance between two opening events. For this analysis we use  $\Delta = 0$  and take into account all events with dwell time between  $a = 0.28$  ms and  $b = 10$  ms. Figure 3.19 shows the dwell times in the

open state and Figure 3.20 a histogram of the distances between two opening events.

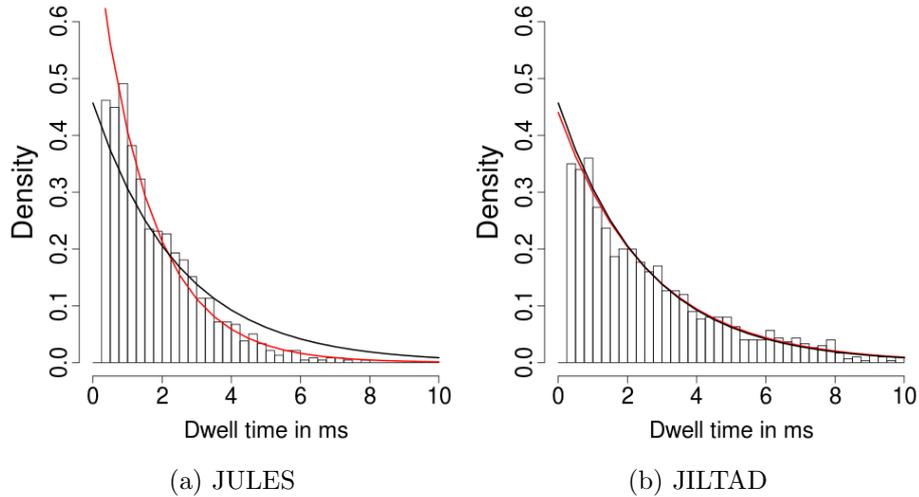


Figure 3.19.: Histograms of the dwell time in the open state for opening events with amplitude between  $l = 0.2$  nS and  $u = 0.5$  nS together with the true exponential distribution with rate 0.4 kHz (black) and an exponential fit (red) taking missed events into account. We rescaled all lines such that the area under them are standardized to one to make them comparable to the histograms.

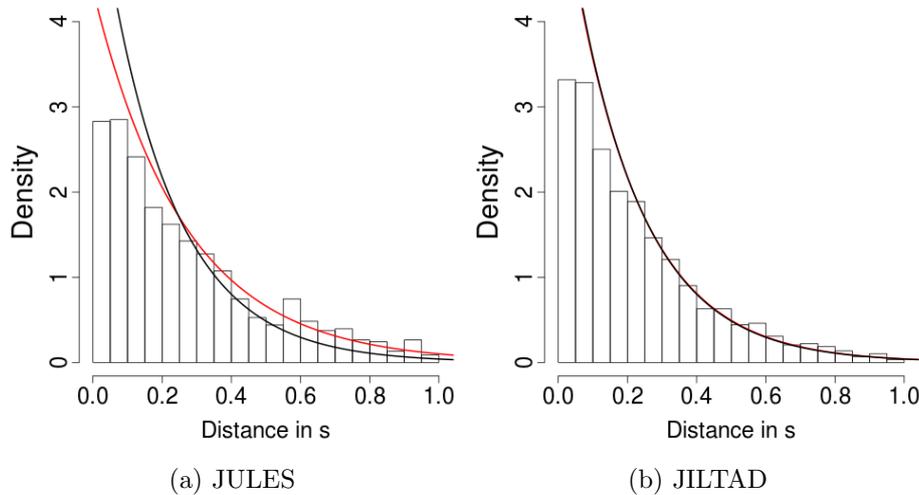


Figure 3.20.: Histograms of the distance between two opening events, i.e., events with amplitude between  $l = 0.2$  nS and  $u = 0.5$  nS and dwell time between  $a = 0.28$  ms and  $b = 10$  ms, together with the true exponential distribution with rate 5 Hz (black) and exponential fits corrected for missed events (red). We rescaled all lines such that the area under them are standardized to one to make them comparable to the histograms.

We found that JILTAD recovers in both cases the exponential distribution very well and estimates both rates with 0.3856 kHz and 5.1931 Hz very accurately. In comparison, JULES is not able to deconvolve all events due to the detection of additional spurious events, compare Figure 3.4. This results in a huge misestimation of 0.6463 kHz and 3.9729 Hz.

Notably, the times seem still to be exponentially distributed, probably since longer events have a larger probability for additional false positives.

### 3.10. Analysis of Gramicidin A recordings

After simulated data generated by a hidden Markov model, we analyze in this section real measurements of the small antibiotic peptide Gramicidin A produced by the soil bacterium *Bacillus brevis* (O’Connell et al., 1990; Andersen et al., 2007). It forms an ion channel by interleaflet dimerization (O’Connell et al., 1990; Andersen et al., 2005, 2007). Apart from the classical slow gating by dimerization, also short closing events of the channel have been observed which could be linked to the physical and chemical properties of the lipid membrane; especially its hydrophobic thickness (Armstrong and Cukierman, 2002; Ring, 1986).

Measurements are performed in the Steinem lab (Institute of Organic and Biomolecular Chemistry, University of Göttingen) with solvent-free lipid bilayers using the Port-a-Patch and are recorded at a sampling rate of 10 kHz using a 1 kHz 4-pole Bessel filter. For more details see (Pein et al., 2017d, Section V A). Note that we use these recordings more as an illustration of the methods and less to obtain relevant biochemical results, for such a study see Section 5.

#### 3.10.1. Estimation of the underlying signals

Estimations are obtained by JULES, JILTAD, TRANSIT and J-SMURF with parameter choices as in Section 3.9.2 and the MDL method (Gnanasambandam et al., 2017), here applied without a minimal event length.

The recorded observations, Figure 3.21, show gating events between two states on various time scales, but also several noise effects like outliers or varying conductance levels, compare for instance the conductance from 10 s–25 s and 55 s–60 s. Note that such effects raise substantial difficulties for methods that assume a Hidden Markov model or a similar model. Contrary, our non-parametric methods JULES and JILTAD, Figures 3.22 and 3.23, provide reasonable estimations covering all major features of the data and some smaller fluctuations. In comparison, JILTAD finds additional events that are too short for JULES to detect which is in accordance with the findings in Section 3.9.3. J-SMURF, Figure 3.24, works well on larger time scales but misses flickering events. TRANSIT, Figure 3.25, is able to detect those events, but detects at the same time far too many false positives. The zooms into single peaks, Figures 3.22–3.25 lower panels, demonstrate that JULES and JILTAD fit the observation well, whereas J-SMURF detects small changes instead of peaks and TRANSIT estimates too small amplitudes and also misestimates their location due to the smoothing by the filter. MDL, Figure 3.26, has the same flaws as TRANSIT, but detects even more events that are very likely false positives. Hence, in the following we do not report results for J-SMURF and MDL.

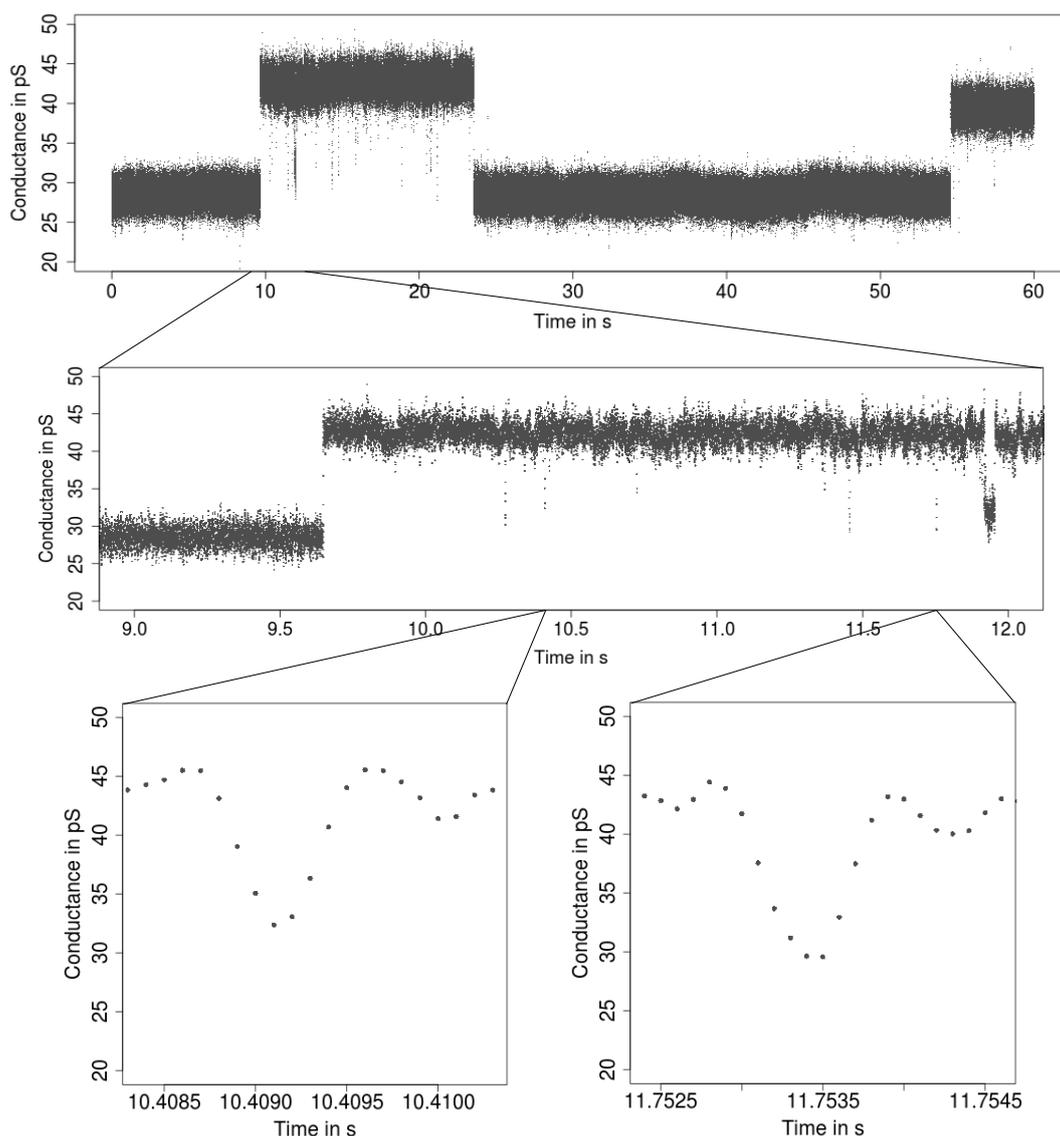


Figure 3.21.: Observations (grey points) of a representative conductance time series obtained by the patch clamp technique for active Gramicidin A inserted into a solvent-free lipid bilayer at 100 mV.

### 3.10.2. Analysis of flickering dynamics

The following analysis of the flickering dynamics of Gramicidin A is similar to the analysis in Section 3.8. Note that flickering does not occur in all recordings. For this analysis we focus on five time series, each with 600 000 data points, which show significant flickering. We define a closing event as a flickering event if its dwell time is smaller than or equal to  $b = 2.6$  ms. For comparison we also study briefly the amplitude of the slow gating, i.e., closing events with dwell time longer than 10 ms. Figure 3.27 compare the estimated amplitudes of flickering events in  $[l, u] = [10, 30]$  pS (307 events) with the amplitudes of slow gating events (44 events) of the same magnitude by kernel density estimates,

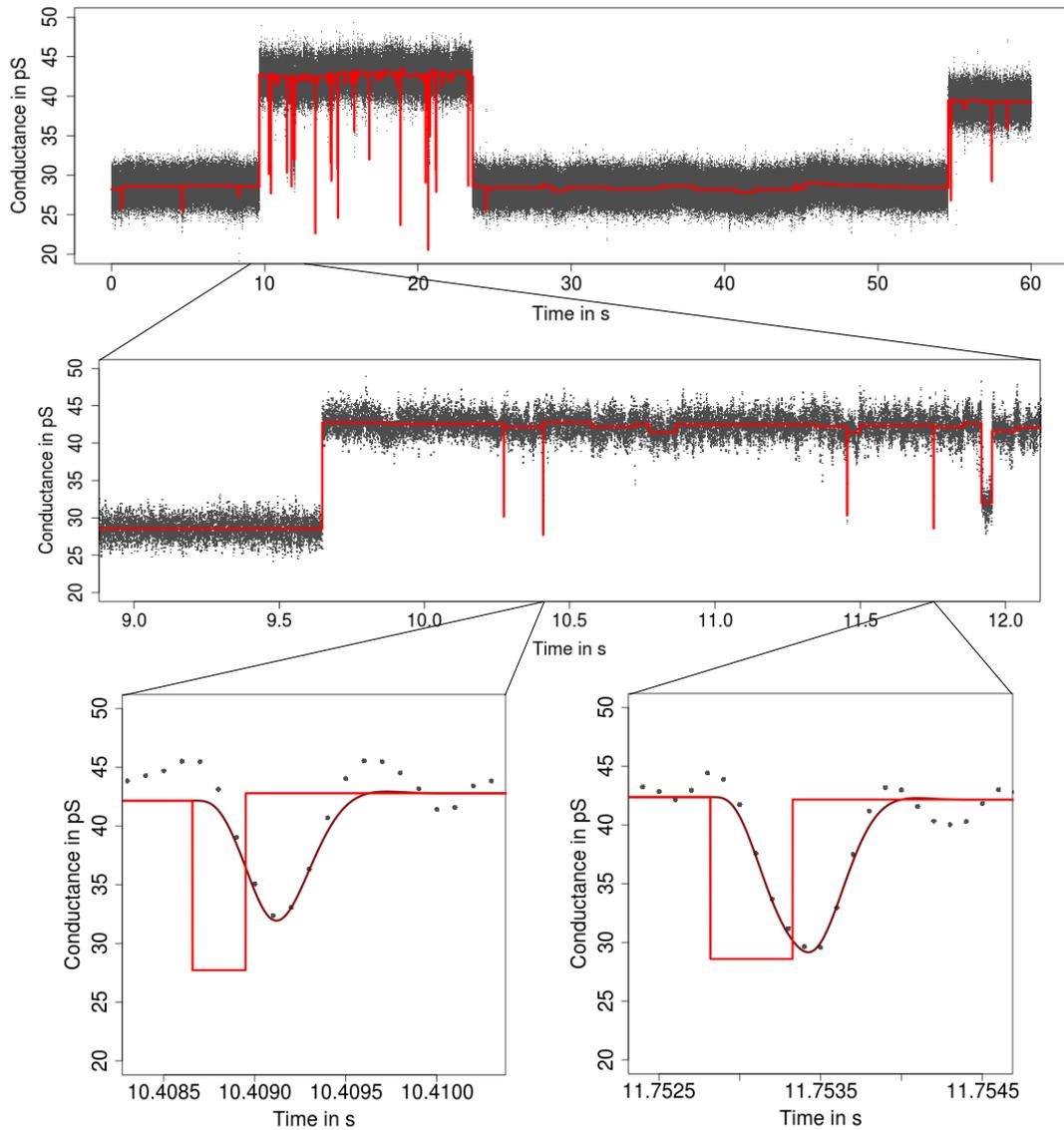


Figure 3.22.: Observations (grey points) as in Figure 3.21 together with an estimation by JULES (red) and its convolution with the lowpass filter (darkred).

performed by the R function `bkde` in the package `KernSmooth` (Wand and Jones, 1994) with bandwidth 2, visually chosen. Other events are most likely caused by artifacts, for instance base line fluctuations.

We see no distinct peak as in the simulations in Figure 3.15. Note that this also true when we show histograms instead of kernel density fits, but kernel density fits allow a better comparison of flickering with the gating on larger time scales. Hence, either multiple levels are underlying or more likely additional errors occur. Recall for instance the conductivity fluctuations in Figure 3.21. Moreover, we found that the flickering events are full-sized, i.e., have the same amplitude as the slow gating events. In comparison, TRANSIT suggests a smaller amplitude which can be explained by the fact that TRANSIT ignores the

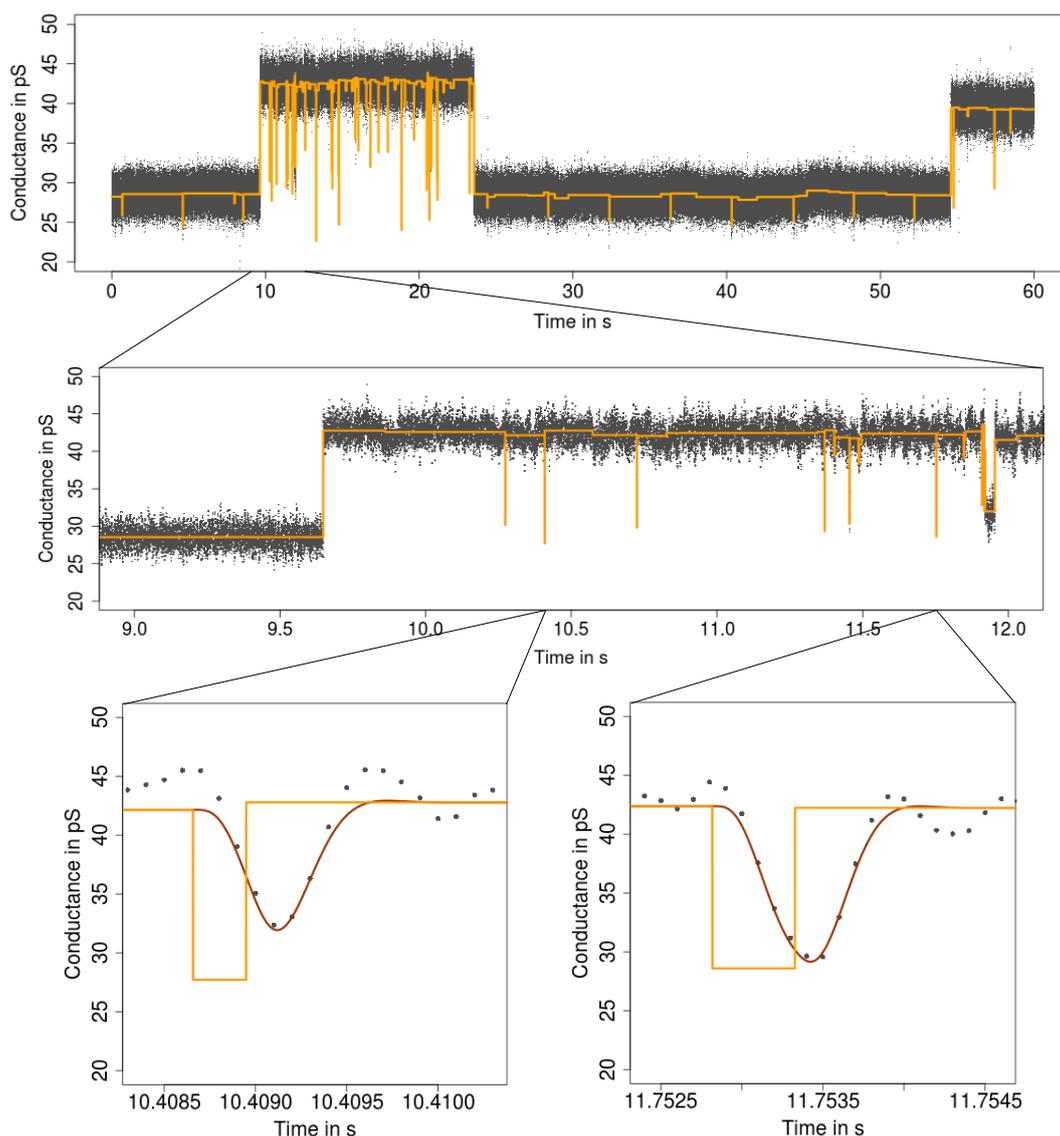


Figure 3.23.: Observations (grey points) as in Figure 3.21 together with an estimation by JILTAD (orange) assuming homogeneous noise and its convolution with the lowpass filter (darkorange).

smoothing effect by the filter. We already saw this effect in the simulations in Section 2.4, see Figures 3.9 and 3.15 as well as Table 3.3.

Figure 3.28 shows the dwell times in the closed state and Figure 3.29 shows the distance between two flickering events. The latter coincides with the dwell time in the open state if other events, in particular events with smaller amplitude, are considered as artifacts and not as gating. We take into account events with a dwell time between  $a = 0.31$  ms for JULES,  $a = 0.24$  ms for JILTAD and  $a = 0.24$  ms for TRANSIT and  $b = 2.6$  ms.

Although the lipid system is totally different, our results are of a similar order than in (Ring, 1986; Armstrong and Cukierman, 2002). Flickering events occur on average roughly every second and are only around a millisecond long. We miss extremely short events in

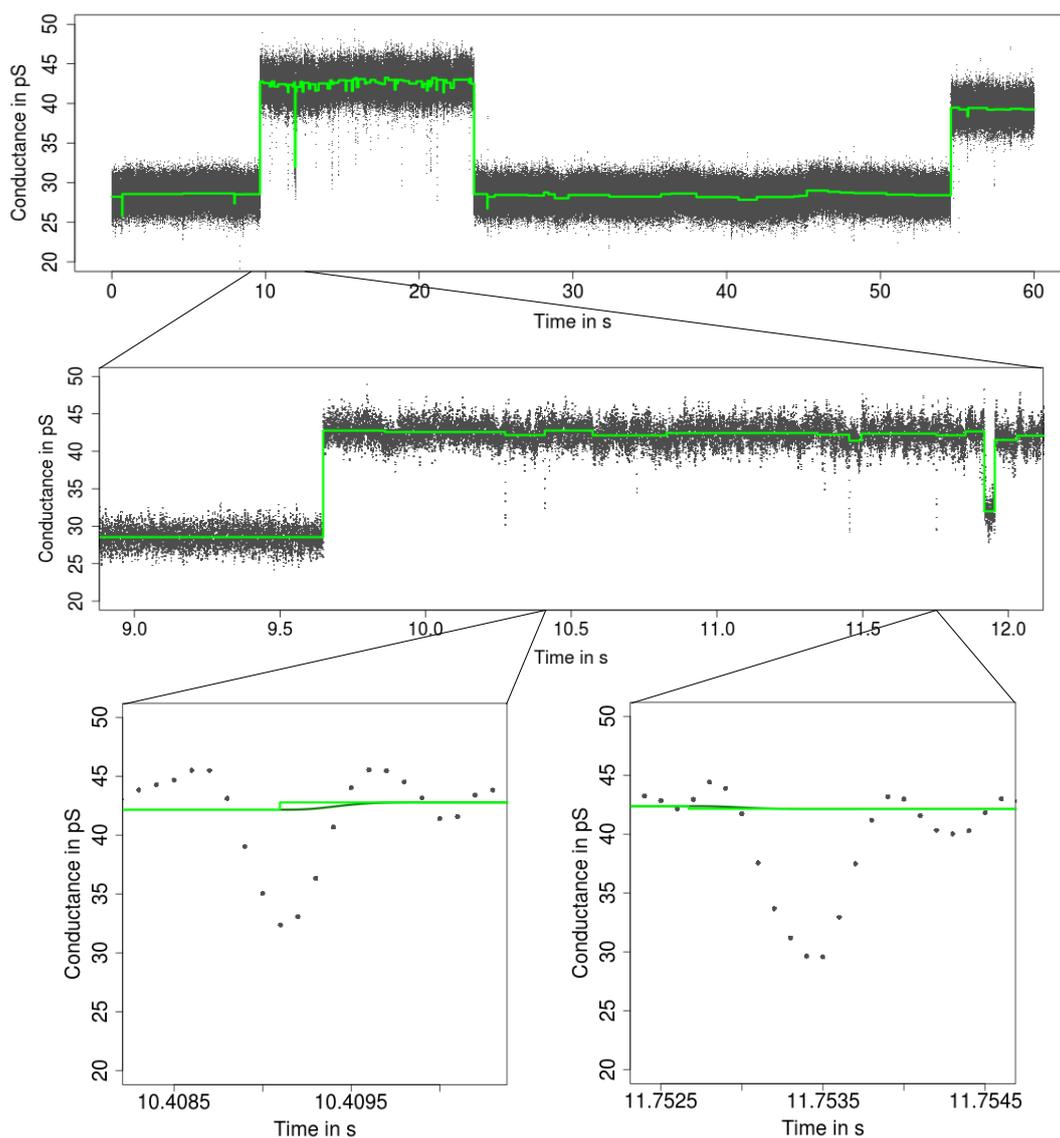


Figure 3.24.: Observations (grey points) as in Figure 3.21 together with an estimation by J-SMURF (green) and its convolution with the lowpass filter (darkgreen).

accordance with the simulations, see Figures 3.10 and 3.16. By taking missed events into account as described in Section 3.8 we confirm an exponential distribution, with rate 0.77 kHz based on JULES and with rate 0.85 kHz based on JILTAD, for the dwell times in the closed state. In comparison, TRANSIT detects more short events resulting in a rate of 1.46 kHz. However, many of them are most likely false positives, see Figure 3.25 and Table 3.1 in the simulations.

In Figure 3.29, the histogram of the distances between two events, too many short distances are observed for a good exponential fit. Since such an effect was not observed in the simulations, compare Figure 3.17, methodology reasons are unlikely. We speculate that this is caused by artifacts, although most of them are of smaller amplitude and not taken

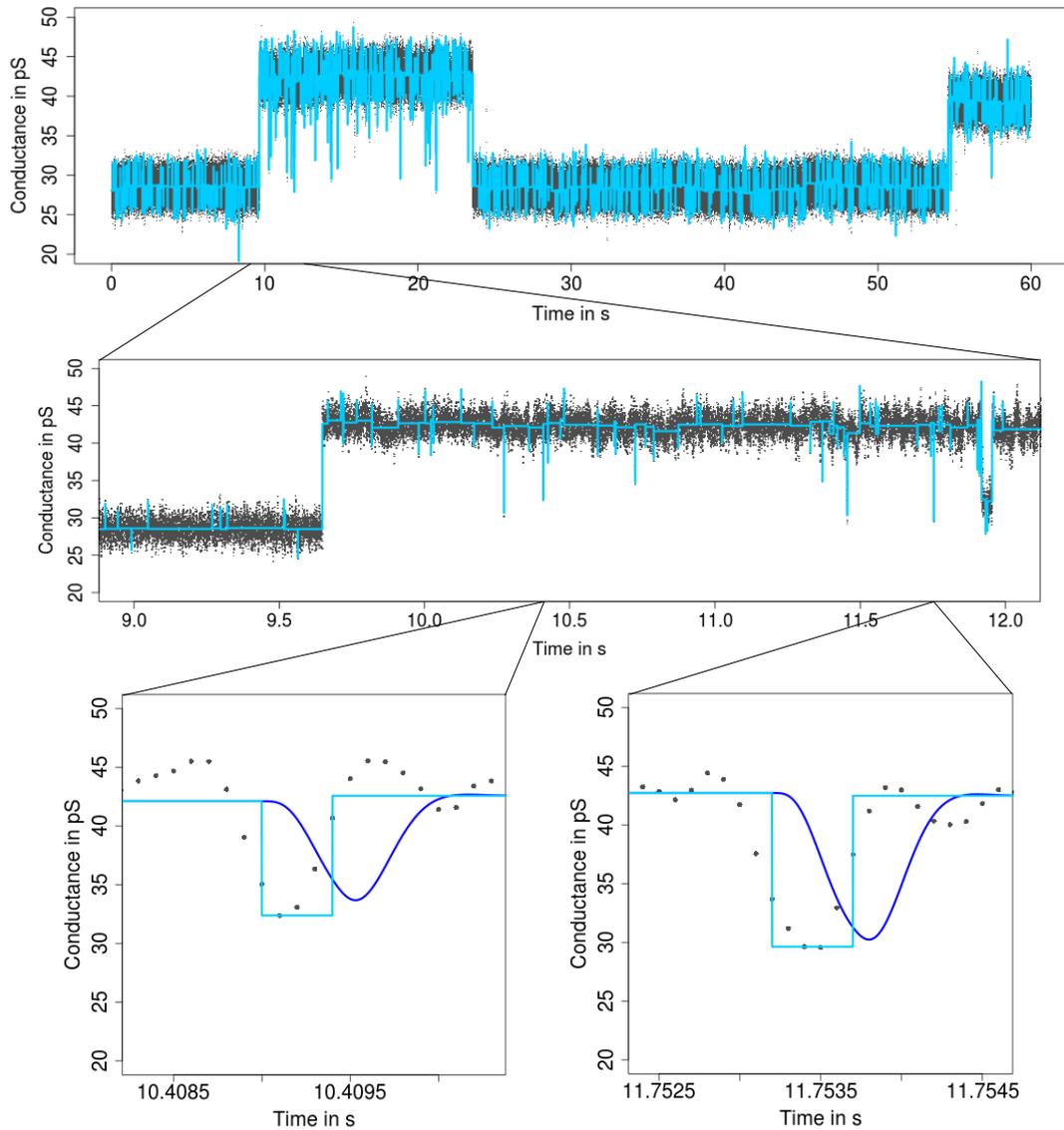


Figure 3.25.: Observations (grey points) as in Figure 3.21 together with an estimation by TRANSIT (lightblue) and its convolution with the lowpass filter (blue).

into account. Hence, to estimate the frequency of the events we only take into account distances of length between  $a = 0.1$  s and  $b = 1$  s. Using also here a missed event correction leads to estimated rates of 2.59 Hz and 3.33 Hz based on the estimations by JULES and JILTAD, respectively. In contrast, when we use TRANSIT we estimate a higher rate of 5.62 Hz due to its many additional findings which are likely to be wrong, confer the number of false positives in Table 3.1. We stress that we cannot exclude that a mixture of two (or more) exponential distributions is underlying, but a more detailed analysis would require more events or less artifacts.

All in all, the comprehensive simulation study from Section 3.9 and the real data application from this section confirm indeed that JULES and JILTAD are able to detect and

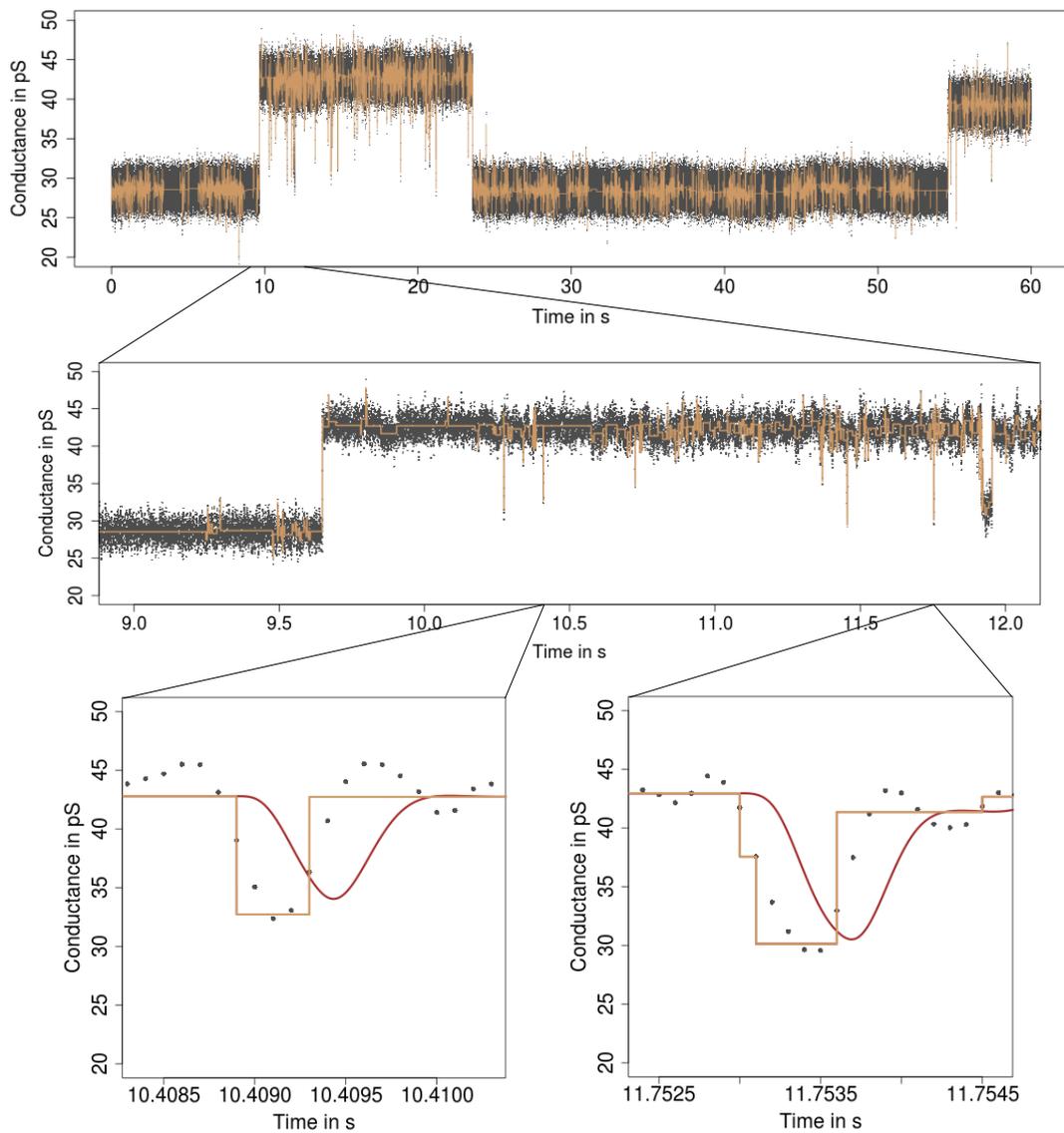


Figure 3.26.: Observations (grey points) as in Figure 3.21 together with an estimation by MDL (lightbrown) and its convolution with the lowpass filter (brown).

estimate events that are short in time with very high precision, JULES only when the noise is homogeneous, but JILTAD also when the noise is heterogeneous.

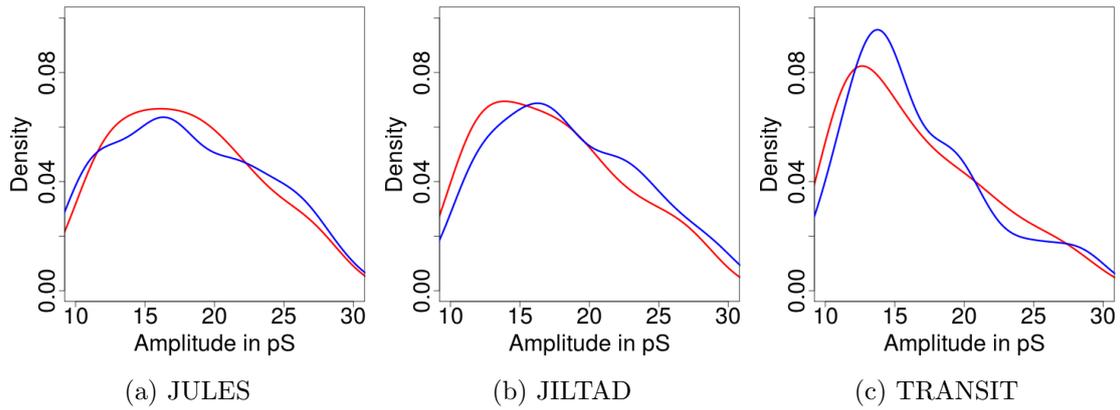


Figure 3.27.: Kernel density estimates with bandwidth 2 of the amplitudes of events with dwell time below 2.6 ms (—) and above 10 ms (—).

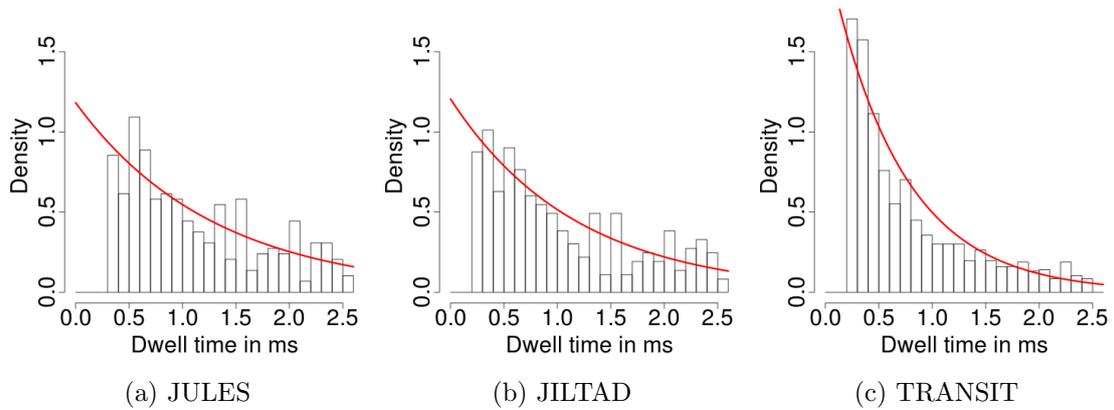


Figure 3.28.: Histograms of the dwell time in the closed state for closing events with amplitude between  $l = 10$  pS and  $u = 30$  pS together with exponential fits (—) taking missed events into account. All lines are rescaled such that the area under them are standardized to one to make them comparable to the histograms.

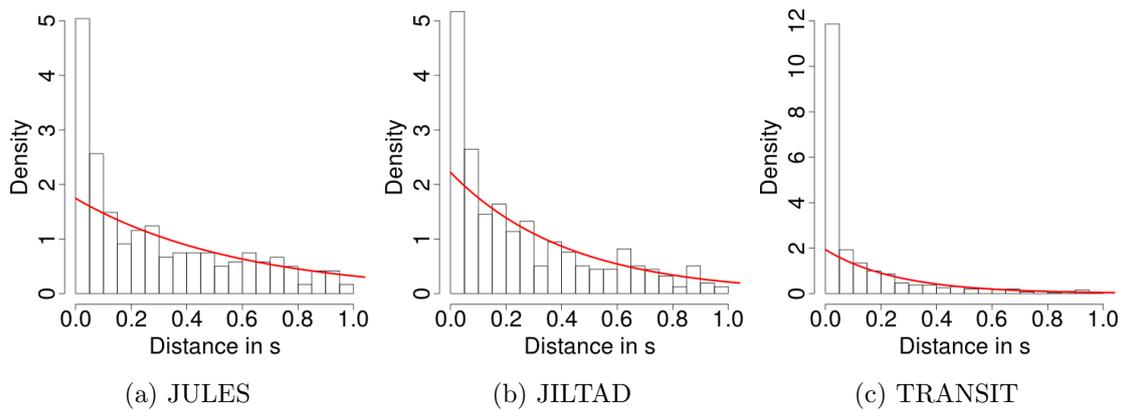


Figure 3.29.: Histograms of the distance between two flickering events, i.e., events with amplitude between  $l = 10$  pS and  $u = 30$  pS and dwell time below  $b = 2.6$  ms, together with exponential fits (—) and the fits corrected for missed events (—). All lines are rescaled such that the area under them are standardized to one to make them comparable to the histograms.

## 4. Computation and software

In Section 4.1 we describe how the multiscale regression estimator (1.2) and its variants presented in this work can be computed efficiently by a pruned dynamic program. We also detail the computation of the vector of critical values  $\mathbf{q}$  in Section 4.2. Our estimators are implemented in two *R* packages *stepR* (Pein et al., 2017a) and *clampSeg* (Pein et al., 2017b), both are available on CRAN. An overview about the functions provided by these two packages is given in Sections 4.3 and 4.4, respectively. Finally, in Section 4.5 we examine the computation time of our estimators theoretically and empirically including comparisons with state of the art methods. A discussion how these algorithms can be parallelized see Section 6.2.3.

### 4.1. Computation of the multiscale regression estimator

The following dynamic program to compute the multiscale regression estimator (1.2) was proposed in (Frick et al., 2014a) and presented in more detail in (Sieling, 2013), we highlight in particular the pseudocode (Sieling, 2013, Algorithm 1). In the following we give only a compact summary and present in Section 4.1.1 a faster algorithm for small interval sets.

Let  $\underline{b}_{i,j}, \bar{b}_{i,j} \in \mathbb{R}$  be the so called *bounds*

$$[\underline{b}_{i,j}, \bar{b}_{i,j}] := \begin{cases} \{m_{ij} \in \mathbb{R} : T_i^j(\mathbf{Y}, m_{ij}) \leq q_{j-i+1}\} & \text{if } [i/n, j/n] \in \mathcal{I}, \\ [-\infty, \infty] & \text{else.} \end{cases} \quad (4.1)$$

Note that the set in the first line of (4.1) is indeed an interval for all models considered in this work, since the test statistics are unimodal, and can be computed by solving  $T_i^j(\mathbf{Y}, m_{ij}) = q_{j-i+1}$ . Moreover, let  $\underline{B}_{i,j}, \bar{B}_{i,j} \in \mathbb{R}$  be the *intersected bounds*

$$\underline{B}_{i,j} := \max_{\substack{i \leq s \leq t \leq j \\ [s/n, t/n] \in \mathcal{I}}} \underline{b}_{s,t} \text{ and } \bar{B}_{i,j} := \min_{\substack{i \leq s \leq t \leq j \\ [s/n, t/n] \in \mathcal{I}}} \bar{b}_{s,t}. \quad (4.2)$$

Consequently, to be accepted by the multiscale test a candidate function  $\mu$  has to satisfy  $m_{ij} \in [\underline{B}_{i,j}, \bar{B}_{i,j}]$  on all intervals  $[i/n, j/n]$  on which it is constant with function value  $m_{ij}$ . The other way around, if  $\underline{B}_{i,j} > \bar{B}_{i,j}$  no function in  $C(\mathbf{Y}, \mathbf{q})$  can be constant on  $[i/n, j/n]$ . We define  $R_k$  as the smallest index  $1 < r \leq n$  such that for  $Y_1, \dots, Y_r$  no solution with  $k - 1$  change-points exists that is accepted by the multiscale test. These right limits are

computed iteratively by  $R_0 := 1$  and

$$R_k := \min \left\{ R_{k-1} < r \leq n : \underline{B}_{R_{k-1}, r} > \overline{B}_{R_{k-1}, r} \right\}, \quad (4.3)$$

for  $k = 1, \dots, \hat{K}$ , which allows at the same time to compute  $\hat{K}$ . Note that this definition is slightly different to its counterpart in (Sieling, 2013), but allows an easier definition of the confidence intervals, see below.

We now assume that for a piecewise constant function  $\mu \in \mathcal{M}$  as in (2.4) the functional  $L(\mathbf{Y}, \mu)$  can be written as

$$L(\mathbf{Y}, \mu) := \sum_{k=0}^K L((Y_{l_k}, \dots, Y_{l_{k+1}-1}), m_k), \quad (4.4)$$

with  $l_k = \lceil n\tau_k \rceil$ . With this assumption we can define the *optimal local costs*

$$c_{i,j} := \begin{cases} L((Y_i, \dots, Y_j), \hat{m}_{i,j}) & \text{if } \underline{B}_{i,j} \leq \overline{B}_{i,j}, \\ \infty & \text{else,} \end{cases} \quad (4.5)$$

with *local estimate*

$$\hat{m}_{i,j} := \operatorname{argmin}_{m \in [\underline{B}_{i,j}, \overline{B}_{i,j}]} L((Y_i, \dots, Y_j), m). \quad (4.6)$$

Let  $k$  be the smallest number of change-points allowed by the multiscale test to reach design point  $p$ , i.e.,  $R_{k-1} \leq p < R_k$ . Then, the *cumulated costs*  $c_p$  to reach design point  $p$  are defined and computed iteratively by  $c_0 := 0$  and  $c_p := c_{l(p)} + c_{l(p)+1,p}$ ,  $p = 1, \dots, n$ , with *optimal last change-point location* to reach design point  $p$

$$l(p) := \operatorname{argmin}_{R_{k-2} \leq l < R_{k-1}} c_l + c_{l+1,p}. \quad (4.7)$$

Finally, the multiscale regression estimator can be obtained by computing and storing the cumulated costs, the local estimates  $\hat{m}_{l(p),p}$ ,  $p = 1, \dots, n$ , and optimal last change-point locations iteratively until the last design point  $n$  is reached and by *backtracking*. Later means to obtain the estimated change-point locations by setting  $\tau_{\hat{K}} := l(n)/n$  and recursively  $\tau_k := l(n\tau_{k+1})$ ,  $k = \hat{K} - 1, \dots, 1$ , and the estimated functions value by  $\hat{m}_k := \hat{m}_{n\tau_k, n\tau_{k+1}}$ ,  $k = 0, \dots, \hat{K}$ . Hence, this approach requires only linear storage space. A discussion of its computation time is given in Section 4.5.

#### 4.1.1. Speed up for small interval sets

In what follows we give a modification of this algorithm which reduces the computation time remarkably when the interval set  $\mathcal{I}$  is small, for example H-SMUCE uses the dyadic partition  $\mathcal{D}$ , which contains only  $\mathcal{O}(n)$  intervals. We observe that we only have to compute the optimal costs for those  $p$  in  $\{R_{k-1}, \dots, R_k - 1\}$  for which the last design point  $n$  is

reachable with  $\hat{K} - k$  change-points, since otherwise the solution has more than  $\hat{K}$  change-points and is not contained in  $C(\mathbf{Y}, \mathbf{q})$ . By the same arguments, we have to consider in (4.7) only those  $l(p) \in \{R_{k-2}, \dots, R_{k-1} - 1\}$  for which the last design point  $n$  is reachable with  $\hat{K} - k + 1$  change-points. Consequently, we define  $L_k$  as the smallest number  $1 < r \leq n$  such that for  $Y_r, \dots, Y_n$  a piecewise constant solution with  $\hat{K} - k$  change-points exists that is accepted by the multiscale test. These left limits can be computed iteratively by  $L_{\hat{K}+1} := n + 1$  and

$$L_k := \min \left\{ 1 \leq r \leq L_{k+1} - 1 : \underline{B}_{r, L_{k+1}-1} \leq \overline{B}_{r, L_{k+1}-1} \right\},$$

for  $k = \hat{K}, \dots, 1$ . It follows that the change-point  $\hat{\tau}_k$  has to be in the interval  $[L_k/n, R_k/n]$ , since otherwise an additional change-point is necessary to fulfill the constraints proposed by the multiscale test. Hence, we only consider  $p \in \{L_k, \dots, R_k - 1\}$  and replace in (4.7)  $R_{k-2}$  by  $L_{k-1}$ . The intervals  $[L_k/n, R_k/n]$ ,  $k = 1, \dots, \hat{K}$ , are confidence intervals for the change-point locations. For the computation of confidence bands we refer to (Frick et al., 2014a; Sieling, 2013). Note that  $R_{k-1} \leq L_k$  always such that the dynamic program requires at the maximum the same amount of steps and often less, but at the price of computing  $L_1, \dots, L_{\hat{K}}$  before. This can be done by looking at the observations in reverse order using dynamic programming, too. Since only the intersected bounds are required its computation time scales with the size of the interval set  $\mathcal{I}$ . We use this algorithm when confidence intervals are requested by the user and the left limits have to be computed anyway and when the interval system  $\mathcal{I}$  is the dyadic partition  $\mathcal{D}$  or all intervals that contain a dyadic number of observations  $\mathcal{L}$ , but not when it is the system of all intervals and confidence intervals are not desired. See also Section 4.5 for a discussion in which situations, in particular for which interval systems, this algorithm is faster than the previous one.

## 4.2. Computation of the vector of critical values

In this section we explain how the vector of critical values  $\mathbf{q} = (q_1, \dots, q_n)$  can be computed based on Monte-Carlo simulations.

### 4.2.1. Scale balancing by penalization

We now describe how to obtain the global quantile  $q_\alpha$  at significance level  $\alpha$  for a finite  $n$ . Let  $M$  be the number of Monte-Carlo simulations and  $T_{n,1}^{pl,n}, \dots, T_{n,M}^{pl,n}$  be i.i.d. copies of the penalized multiscale statistic  $T_n^{pl,n}(\boldsymbol{\epsilon}, 0)$  as in (1.5). To this end, the choice  $M = 10\,000$  appears to be a good trade-off between computation time and approximation accuracy. Then, the global quantile  $q_{\alpha,n}$  at significance level  $\alpha$  is obtained as the empirical  $(1 - \alpha)$  quantile of these copies. Note that the  $(1 - \alpha)$  quantile of the limit distribution of  $T_n^{pl,n}(\boldsymbol{\epsilon}, 0)$ , see (Frick et al., 2014a, Theorem 1), can be used as well. However, an analytic computation is very difficult. Hence, this quantile is typically approximated by the computation above

for a very large  $n$ . From the global quantile  $q_{\alpha,n}$  the vector of critical values can be obtained easily by  $(p_{1,n}^{-1}(q_{\alpha,n}), \dots, p_{n,n}^{-1}(q_{\alpha,n}))$ .

#### 4.2.2. Scale balancing by weights

We now detail the computation of the vector of critical values obtained by weights. To be more general, we describe the computation of the vector of critical values defined in Remark 2. The algorithm only uses the continuity and monotonicity of the cumulative distribution functions of the statistics  $T_1, \dots, T_n$ , which ensures its flexibility.

Let  $M$  be the number of simulations and  $(T_{1,1}, \dots, T_{n,1}), \dots, (T_{1,M}, \dots, T_{n,M})$  be i.i.d. copies of the vector  $(T_1, \dots, T_n)$ . We again suggest  $M = 10\,000$  as a default choice. Moreover, let  $F_M(\cdot)$  denote the empirical distribution function of  $(T_1, \dots, T_n)$  and  $F_{M,k}(\cdot)$  the empirical distribution function of the random variable  $T_k$ . Then, we aim to find a vector of critical values  $\hat{\mathbf{q}}_M = (\hat{q}_{M,1}, \dots, \hat{q}_{M,n})$  which satisfies with

$$\alpha - \frac{1}{M} < 1 - F_M(\hat{\mathbf{q}}_M) \leq \alpha, \quad (4.8)$$

an empirical version of condition (2.12), and with

$$\frac{1 - F_{M,j_1}(\hat{q}_{M,j_1})}{\beta_{j_1}} \leq \frac{1 - F_{M,j_2}(\hat{q}_{M,j_2}) + \frac{1}{M}}{\beta_{j_2}} \quad \text{for all } j_1, j_2 \in \{1, \dots, n\}, \quad (4.9)$$

an empirical version of condition (2.14). In the following we propose an iterative algorithm to determine such a vector and show afterwards that this vector converges almost surely to the vector of critical values  $\mathbf{q} = (q_1, \dots, q_n)$  defined by (2.12) and (2.14). As the  $k$ -th entry of the starting vector we choose the empirical  $(1 - \alpha\beta_k)$ -quantile of the statistic  $T_k$ , since the vector with these values satisfies condition (4.9) and the inequality

$$1 - F_M(\cdot) \leq \alpha.$$

Afterwards, we reduce the entries until the lower bound from condition (4.8) is satisfied, too. To ensure condition (4.9) in every iteration, we always reduce the entry which has the smallest ratio

$$\frac{1 - F_{M,k}(\hat{q}_{M,k})}{\beta_k}.$$

In Algorithm 3 the determination of the critical values is summarized in pseudocode.

The method has the advantage that we do not need specific assumptions on the distribution of the vector  $(T_1, \dots, T_n)$  and still get critical values which are adapted to the exact finite sample distribution of  $(T_1, \dots, T_n)$  and ensure therefore even for a finite number of observations the significance level  $\alpha$ .

The following theorem shows the convergence of this algorithm to  $\mathbf{q} = (q_1, \dots, q_n)$ .

**Theorem 14** (Consistency of Monte-Carlo critical values). *The empirical vector of critical values  $\hat{\mathbf{q}}_M = (\hat{q}_{M,1}, \dots, \hat{q}_{M,n})$  converges almost surely in the number of simulations  $M$  to*

---

**Algorithm 3** Determination of the critical values.

---

**Input:** The statistics  $T_1, \dots, T_n$  as well as the significance level  $\alpha \in (0, 1)$ , the weights  $\beta_1, \dots, \beta_n > 0$ , with  $\sum_{k=1}^n \beta_k = 1$ , and the number of simulations  $M \in \mathbb{N}$ .

**Output:** The vector of critical values  $\hat{\mathbf{q}}_M = (\hat{q}_{M,1}, \dots, \hat{q}_{M,n})$  fulfilling (4.8) and (4.9).

```
1: for  $i = 1, \dots, M$  do
2:    $(T_{1,i}, \dots, T_{n,i}) \leftarrow$  realisation of  $(T_1, \dots, T_n)$ 
3: end for
4: for  $k = 1, \dots, n$  do
5:    $(S_{k,1}, \dots, S_{k,M}) \leftarrow$  sort  $((T_{k,1}, \dots, T_{k,M}))$ 
6:    $w_k \leftarrow M - \lfloor \alpha \beta_k M \rfloor$ 
7: end for
8: repeat
9:    $\hat{k} \leftarrow \operatorname{argmin}_{k=1, \dots, n} \beta_k^{-1} (1 - F_{M,k}(S_{k,w_k}))$ 
10:   $w_{\hat{k}} \leftarrow w_{\hat{k}} - 1$ 
11: until  $1 - F_M(S_{1,w_1}, \dots, S_{m,w_n}) > \alpha$ 
12:  $w_{\hat{k}} \leftarrow w_{\hat{k}} + 1$ 
13: return  $S_{1,w_1}, \dots, S_{m,w_n}$ 
```

---

the vector of critical values  $\mathbf{q} = (q_1, \dots, q_n)$  defined by (2.12) and (2.14).

### 4.2.3. Storing the results of the Monte-Carlo simulations

A Monte-Carlo simulation to generate  $M$  i.i.d. copies of the penalized multiscale statistic  $T_n^{\text{pl}|\epsilon, 0}$  or of the vector  $(T_1, \dots, T_n)$  lasts potentially much longer, up to several hours or days if the number of observations is in the millions, than the main calculations. For a more detailed comparison of the computation times see Section 4.5. Thus, we suggest to store the results and load them if they are required again. The *stepR* package, see Section 4.3, offers multiple possibilities to do so and by default some are executed.

The simulations can either be stored persistently on the file system for which the *R* package *R.cache* is used or in the workspace. Loading from the workspace is faster, but either the user has to store the workspace manually or in a new session simulations have to be performed again. Moreover, storing in and loading from variables and files with a user given path are supported.

Both, copies of the penalized multiscale statistic and of the vector  $(T_1, \dots, T_n)$ , can be stored by all options. Note that the penalized multiscale statistic can be computed quickly from the vector  $(T_1, \dots, T_n)$ . Hence, later is more flexible, but requires much more storage space.

To avoid memory problems and resimulations if the number of observations is only slightly different, Monte-Carlo simulations can also be performed for a (slightly) larger number of observations. The overestimation control, (Frick et al., 2014a, (7)) and Theorem 5, is still satisfied but the detection power is (slightly) smaller. By default, we perform simulations for  $2^{\lceil \log_2(n) \rceil} - 1$  observations and store the copies of the vector  $(T_1, \dots, T_n)$  on the file system and copies of the penalized multiscale statistic in the workspace. More details are given in the documentation of the function *critVal* in the *stepR* package.

### 4.3. The R package *stepR*

The R package *stepR* (Pein et al., 2017a) offers an efficient computation of the multiscale regression estimator (1.2), its confidence statements and connected terms, like the multiscale test statistic or bounds, by dynamic programs as described in Section 4.1. To speed up computations the time consuming dynamic programs are written in C++ and interfaced by the R code. All computations can be performed for various parametric families, which is in a strict sense the parametric distribution of the errors  $\epsilon_1, \dots, \epsilon_n$ , but we use this terminology also for the local test statistics and the functional  $L(\mathbf{Y}, \mathbf{q})$  that define the estimator, and several interval systems  $\mathcal{I}$ . An overview is given at the end of this section. To this end, the dynamic programs are written in an object oriented way with abstract classes for the parametric families and the interval systems. This allows the addition of a new family or interval system by only writing a new derived class with few functions instead of writing a whole complicated dynamic program. Moreover, if for instance a class for a new parametric family is implemented, its functionality is provided for all functions and for all interval systems. Note that this also includes testing. The dynamic program is tested well by several hundred checks such that only the new class has to be tested which is less time consuming and much easier.

A first version was written by Hotz and Sieling (2015). We rewrote their package completely and also incorporated ideas from the packages *HSMUCE* (Pein, 2016), our previous implementation of H-SMUCE, and *FDRSeg* (Li and Sieling, 2015). We added several new functionalities like obtaining critical values by weights as described in Section 4.2.2 or the computation of the multiscale test statistic. Moreover, in the new implementation also the interval system is object oriented and not only the parametric family as before. We also added the parametric family 'hsmuce' and the dyadic partition as an interval system. For smaller interval sets or when confidence statements are requested we compute the estimator by a faster dynamic program as detailed in Section 4.1.1. Finally, for a faster computation we do not pre-compute all bounds, but rather compute required bounds on-line in the dynamic program.

The main function *stepFit* allows to compute the multiscale regression estimator  $\hat{\mu}$  (1.2), confidence intervals for the change-point locations and a confidence band for  $\mu$ . The vector of critical values  $\mathbf{q}$  and the global quantile  $q_\alpha$  at significance level  $\alpha$  can be computed by the function *critVal*. Monte-Carlo simulations are automatically performed if required. Moreover, as explained in Section 4.2.3 several possibilities to store the results are offered. For given  $\mu \in \mathcal{M}$  the penalized multiscale statistic (1.5) and the vector of all unpenalized statistics can be computed by the function *computeStat*. And the bounds (4.1) can be computed by the function *computeBounds*. Monte-Carlo simulations are performed by the function *monteCarloSimulation*. For completeness, we remark that also the functions *compareBlocks*, *neighbours*, *sdrobnorm*, *stepcand*, *steppath*, *stepsel* from the previous version are still available. Moreover, the package contains functions that are marked as deprecated and will be removed in further versions, but are currently still available as an

immediate deletion might lead to misunderstandings for some users.

Currently the following parametric families are provided: SMUCE for independent Gaussian, Poisson, Binomial and Gaussian variance regression. H-SMUCE for independent heterogeneous Gaussian observations and for  $L$ -dependent Gaussian observations with known, but arbitrarily covariance structure, as it is requested by the detection step of JULES. And it supports the following interval systems: All intervals with start and end points at the design grid, all intervals that contain a dyadic number of observations  $\mathcal{L}$  and the dyadic partition  $\mathcal{D}$ .

For more details see the vignette (Pein and Munk, 2017) and the documentation of each function itself.

#### 4.4. The R package `clampSeg`

The `clampSeg` package is specialized on estimating the signal underlying ion channel recordings by non-parametric approaches. The current version (Pein et al., 2017b) only offers an implementation of JULES, a version containing also J-SMURF and JILTAD is not published so far. Note that J-SMURF as described in (Hotz et al., 2013) can be accessed by the R function `jsmurf` in the `stepR` package (Pein et al., 2017a), but shall be moved to the `clampSeg` package, recall the previous note on deprecated functions. Also for this package time consuming computations are written in C++ and interfaced by the R code.

The main function `jules` allows estimation by JULES. An object providing the lowpass filter functions, mainly the kernel function, its antiderivative and the autocorrelation function, can be created by the function `lowpassFilter`. Currently, only Bessel filters are supported. The global quantile  $q_\alpha$  can be derived in an universal manner by Monte-Carlo simulations and for every significance level  $\alpha$  by the function `getCritVal`, automatically called if required. By default,  $\alpha = 0.05$  is chosen. This function performs required Monte-Carlo simulations automatically and by default stores their results by calling the function `critVal` from the `stepR` package. The detection and estimation step of JULES, as described in Sections 3.5 and 3.6, can be obtained separately by the functions `stepDetection` and `deconvolveLocally`, respectively. As an example the Gramicidin A recording shown in the middle panel of Figure 3.21 is provided by the data object `gramA`.

#### 4.5. Computation time

In this section we discuss the computation time of the proposed methods. We will see that the computation time of the multiscale regression estimator is dominated by the Monte-Carlo simulation to compute the vector of critical values. However, such a simulation is required only once for each setting, since its result can be restored, see Section 4.2.3. Thus, we firstly discuss how long a Monte-Carlo simulation lasts. For the discussion of the computation time of the estimator itself we ignore this time. In the run time simulations we

ensure that all required results are stored and only the loading times from the workspace are included in the results. We do not discuss the complexity to compute the vector of critical values, since in practice the run time for it is vanishingly small if the result of the corresponding Monte-Carlo simulation is available. A discussion how the practical run times can be reduced by parallelization is given in Section 6.2.3.

We assume that the local test statistic  $T_i^j(\mathbf{Y}, m)$ , the local cost  $L((Y_i, \dots, Y_j), \hat{m}_{i,j})$  and the local estimate  $\hat{m}_{i,j}$  can be computed in constant time  $\mathcal{O}(1)$  as soon as cumulated sums  $\sum_{l=i}^j \gamma_l$  of certain quantities  $\gamma_l$  are derived. For instance for H-SMUCE the length  $\sum_{l=i}^j 1$ , the cumulated sum of observations  $\sum_{l=i}^j Y_l$  and cumulated sum of the squared observations  $\sum_{l=i}^j Y_l^2$  are required. We stress that this assumption is not true if the likelihood ratio test statistic is used for correlated observations, e.g., (3.18).

A simple way to compute these sums is to compute  $\sum_{l=1}^j \gamma_l$ ,  $j = 1, \dots, n$ , and then  $\sum_{l=i}^j \gamma_l = \sum_{l=1}^j \gamma_l - \sum_{l=1}^{i-1} \gamma_l$ . However, this computation can lead to large numerical errors, since for short intervals with large indices two large numbers of almost equal size are subtracted to obtain a much smaller number. This approach computes for instance sometimes negative empirical variances due to numerical errors. Hence, we obtain these sums iteratively. More precisely, for the interval system of all intervals we use the recursion  $\sum_{l=i}^j \gamma_l = \gamma_i + \sum_{l=i+1}^j \gamma_l = \sum_{l=i}^{j-1} \gamma_l + \gamma_j$  and for the set of all intervals containing a dyadic number of observations and the dyadic partition we use  $\sum_{l=i}^j \gamma_l = \sum_{l=i}^{(i+j-1)/2} \gamma_l + \sum_{l=(i+j+1)/2}^j \gamma_l$ .

Under these assumptions all test statistics, all bounds and all intersected bounds can be computed in  $\mathcal{O}(|\mathcal{I}|)$  steps. For the intersected bounds this follows from the recursion formulas  $\underline{B}_{i,j} := \max\{\underline{B}_{i+1,j}, \underline{B}_{i,j-1}, \underline{b}_{i,j}\}$  and  $\bar{B}_{i,j} := \min\{\bar{B}_{i+1,j}, \bar{B}_{i,j-1}, \bar{b}_{i,j}\}$ .

It follows that the computation of the penalized multiscale statistic  $T_n^{p_{j-i+1,n}}(\mathbf{Y}, \mu)$  as in (1.5) and of the vector  $(T_1, \dots, T_n)$  as in (2.15) requires  $\mathcal{O}(|\mathcal{I}|)$  steps and a whole Monte-Carlo simulation  $\mathcal{O}(M|\mathcal{I}|)$ . This means that the run time increases for the dyadic partition linearly in the number of observations  $n$ , for the system of all intervals containing a dyadic number of observations with order  $n \log(n)$  and quadratically for the system of all intervals. Table 4.1 shows exemplary run times. The simulations were performed on a single core system with 1.8 GHz and 8 GB RAM in a 64-bit OS.

Table 4.1.: Run times of Monte-Carlo simulations using different interval sets.

Number of observations	All intervals	dyadic	dyadic partition
$2^7 = 128$	5.79 s	2.96 s	2.63 s
$2^{10} = 1\,024$	183 s	11.76 s	7.41 s
$2^{16} = 65\,536$	11 997 min	16 min	6 min
$2^{20} = 1\,048\,576$	-	6 h	2 h
$2^{24} = 16\,777\,216$	-	103 h	32 h

Comparing them with the results in Figures 4.1 confirms that the Monte-Carlo simulation dominates the computation of the estimator and storing of results is beneficial. Moreover, it shows that Monte-Carlo simulations last much longer for the interval system of all intervals. Hence, we could not perform a Monte-Carlo simulation for hundred thousands

and more observations. And note that storing of Monte-Carlo simulations for all intervals require storage space of order  $nM$  and hence is expensive as well. For smaller interval sets we performed Monte-Carlo simulations up to 16 million observations and stored their results for SMUCE and H-SMUCE.

We now discuss the computation time of the estimator itself. For the algorithm in Section 4.1 without the speed up from Section 4.1.1 the cumulated costs have to be computed for each  $p$  in  $\{R_{k-1}, \dots, R_k - 1\}$ . And for each of these design points the last change-point location can be in  $\{R_{k-2}, \dots, R_{k-1} - 1\}$ , see (4.7), and the corresponding local costs have to be computed. Hence, the complexity is bounded by the data dependent term

$$\mathcal{O} \left( \sum_{k=2}^{\hat{K}} (R_k - R_{k-1})(R_{k-1} - R_{k-2}) \right), \quad (4.10)$$

since at most the same amount of bounds has to be computed and the right limits are obtained by the same iterations. This term can be bounded by  $\mathcal{O}(n^2)$ , but for many observations less computations are required, since from  $c_{l,p} = \infty$  it follows that  $c_{l',p} = \infty$  for all  $l' \leq l$ . Ignoring some start locations is called *pruning*.

The computation of the left limits in advance requires to compute all bounds in the intervals  $[L_k/n, (L_{k+1} - 1)/n]$ . Hence, the complexity for it is bounded by

$$\mathcal{O} \left( \left| \left\{ \left[ \frac{i}{n}, \frac{j}{n} \right] \in \mathcal{I} : \exists k \in \{0, \dots, \hat{K}\} \text{ s.t. } \left[ \frac{i}{n}, \frac{j}{n} \right] \subset \left[ \frac{L_k}{n}, \frac{L_{k+1} - 1}{n} \right] \right\} \right| \right). \quad (4.11)$$

These computations allow to store the intersected bounds  $\underline{B}_{r,L_{k-1}}$  and  $\overline{B}_{r,L_{k-1}}$  for all  $r = L_{k-1}, \dots, L_k - 1$  and for all  $k = 1, \dots, \hat{K}$ . Hence, by the same arguments as for (4.10) it follows that the complexity of the main dynamic program with the speed up in Section 4.1.1 is bounded by

$$\mathcal{O} \left( \sum_{k=2}^{\hat{K}} (R_k - L_k)(R_{k-1} - L_{k-1}) \right). \quad (4.12)$$

Combining (4.12) and (4.11) gives the complexity of the whole algorithm.

A general comparison of these computation times is difficult. Hence, we only highlight certain special cases. For the system of all intervals (4.11) is of the same order as (4.10), which shows that for this system no computation time is gained by the approach in Section 4.1.1. For smaller interval systems (4.11) is of a smaller complexity. An easy bound is the number of intervals in  $\mathcal{I}$ , i.e.,  $\mathcal{O}(|\mathcal{I}|)$ . Hence, for the dyadic partition this term is of order  $\mathcal{O}(n)$  and is always of the same or of smaller order than (4.12).

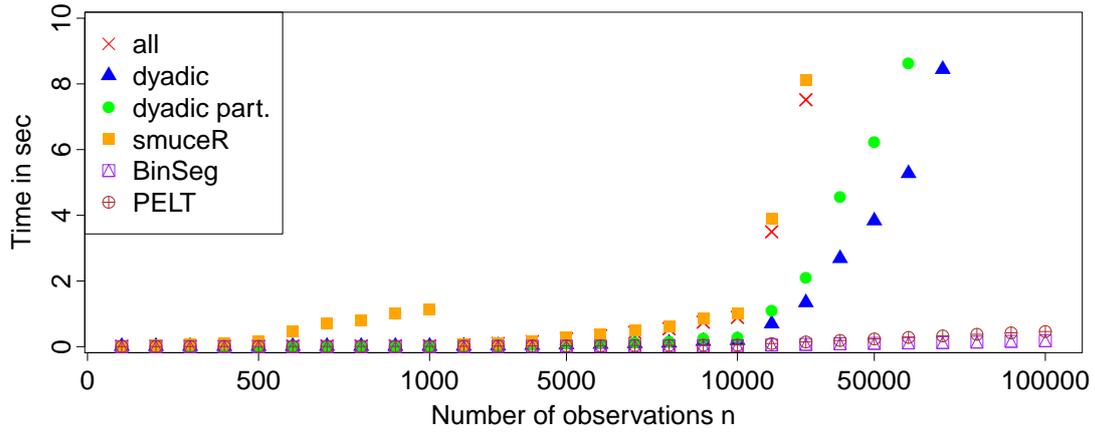
The term in (4.12) can again be bounded by  $\mathcal{O}(n^2)$  in the worst case, but the computation time is in many cases much smaller. In particular, if the signal to noise ratios are large such that the change-points are easy to detect and hence  $R_k - L_k$  is small. This is asymptotically for instance the case for a fixed signal, where  $R_k - L_k$  stays more or less constant in  $n$ . More precisely, by combining (4.12) with equation (2.24) we see that with probability

tending to one (4.12) is of linear order, if  $\alpha_n \rightarrow 0$ , but  $n^{-\frac{1}{2}} \log((\alpha_n \beta_{k_n, n})^{-1}) \rightarrow 0$ . (4.10), (4.11) and (4.12) are also  $\mathcal{O}(n)$  if the number of change-points increases linearly in the number of observations and the change-points are evenly enough distributed.

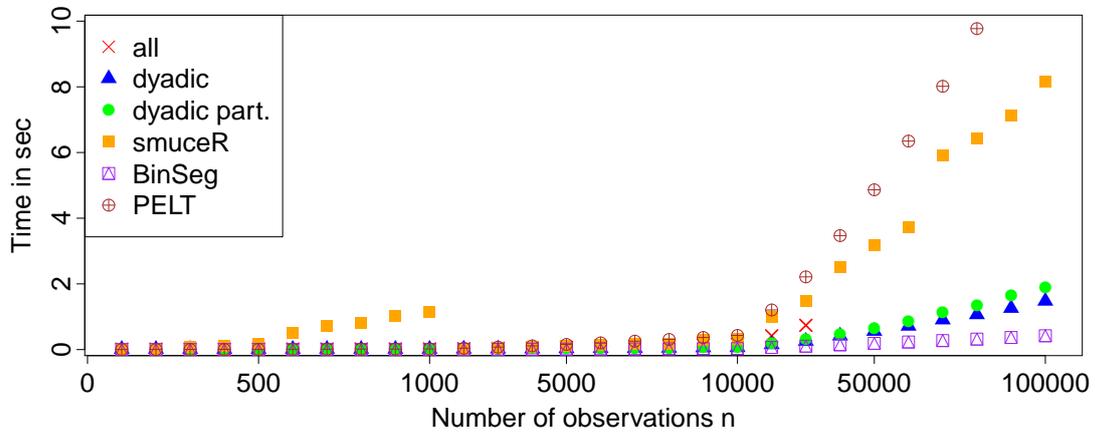
In the following we examine the computation time empirically in a similar simulation study as in (Maidstone and Pickering, 2014). More precisely, we generate data with varying number of observations  $n$  and equidistant change-points. Thereby, we consider  $K = 10$ ,  $K = \sqrt{n}$  and  $K = n/100$ . The standard deviation is chosen to be one and the jump size  $|m_k - m_{k+1}|$  equal to  $\sqrt{200/n * (K + 1)}$ . Upward and downward changes alternate. This construction guarantees that with high probability all methods are able to detect all change-points. All simulations are repeated 100 times. The simulations were performed on a single core system with 1.8 GHz and 8 GB RAM in a 64-bit OS and terminated after ten seconds.

We compare our implementation of the multiscale regression estimator SMUCE in (Pein et al., 2017a), using all intervals, all intervals containing a dyadic amount of observations and the dyadic partition, with the old implementation in (Hotz and Sieling, 2015) by applying the function *smuceR*. Note that later uses all intervals until 1 000 observations and all intervals containing a dyadic number of observations afterwards. We also compare it with PELT (Killick et al., 2012) and binary segmentation (Vostrikova, 1981), BinSeg, using the *R* function *cpt.mean* in the package *changeoint*. Both methods are known to be fast and serve as a benchmark. All methods are called for independent homogeneous Gaussian observations. The maximal number of change-points for binary segmentation is set to  $Q = 2 * K$ . Note, that this choice incorporates prior knowledge about the true signal. For all other parameters the default values are used. We stress that the computation times of H-SMUCE is only slightly larger than the one of SMUCE assuming homogeneous Gaussian observations, since computation of the test statistics and bounds are slightly more expensive. In particular, they require additionally the computation of cumulated sums of the squared observations. In (Pein et al., 2017c, Figure 8 in the Supplement) we compared H-SMUCE, using the old implementation in (Pein, 2016), with all other methods that are known to us to be robust against heterogeneous noise. We found that H-SMUCE was significantly faster than these methods.

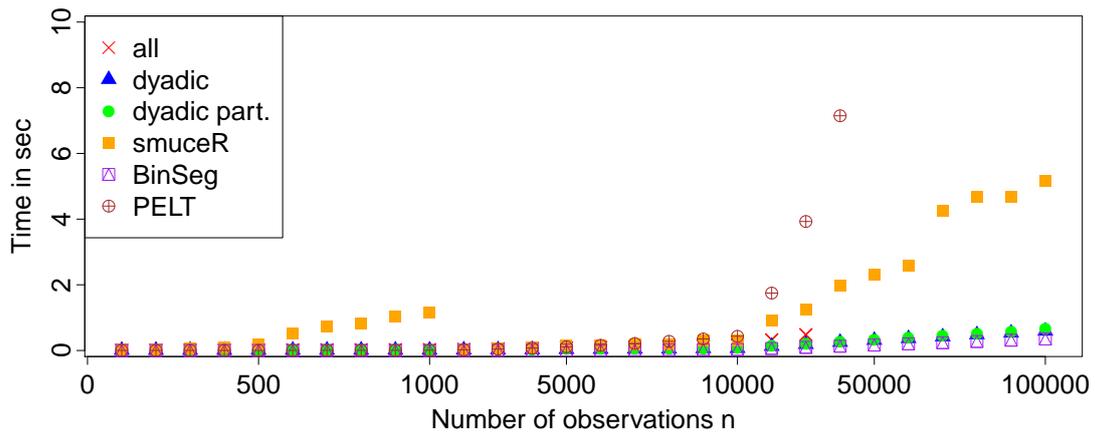
Figure 4.1 shows that in all settings the new implementation is much faster than the old implementation. It is even slightly faster when it uses with the system of all intervals much more intervals than *smuceR* (for  $n > 1\,000$  observations). Using all intervals that contains a dyadic amount of observations leads to a slightly smaller computation time than using the dyadic partition. This is surprising, since the dyadic partition has a smaller cardinality and hence leads to a smaller complexity. Most likely this is caused by the fact that for the dyadic partition it is more difficult to decide whether an interval is part of the system or not. We think that this can be improved by precomputing all bounds, but this is not implemented so far. This is possible for the dyadic partition, since only a linear storage space is required. The computation time is much smaller if more change-points occur. For a constant number of change-points the computation is much slower than PELT and



(a)  $K = 10$ .



(b)  $K = \sqrt{n}$ .



(c)  $K = n/100$ .

Figure 4.1.: Mean computation time of various methods in the different settings. Note that for purposes of visualization the x-axes are displayed non-equidistantly. Also note that for the system of all intervals Monte-Carlo simulations are only performed until  $n = 30\,000$ . Hence, no times are reported for a larger number of observations.

binary segmentation. In comparison, if the number of change-points increases linearly in the number of observations, the computation time seems to increase only linearly in the number of observations and the method is almost as fast as PELT, while binary segmentation has a much larger computation time. Hence, in this setting the computation of the multiscale regression estimator for one million observations lasts still less than a minute, which allows its usage in large data applications.

Finally, we discuss the computation times of the methods for the ion channel recordings. J-SMURF and the detection step of JULES are multiscale regression estimators and the previous discussion applies. Hence, the computation time is roughly  $\mathcal{O}(n)$ , since for most ion channel recordings change-points occur frequently. The additional tests in JILTAD are of complexity  $\mathcal{O}(l_{\max}^2 n)$ , since for each of the  $l_{\max}$  scales  $1, \dots, l_{\max}$  roughly  $n$  tests have to be performed and the complexity to compute a single test is at most of order  $\mathcal{O}(l_{\max})$ , since the expectation (and variance) of at most  $l_{\max} + L - 1$  observations have to be computed. The computation time of the local deconvolution is dominated by the iterative grid search to deconvolve a single event, all other computations can be ignored. The deconvolution of a single event is constant in the number of observations, since the number of involved observations and the grid sizes do not increase. Moreover, the number of involved observations is small and the covariance matrix is a band matrix, with band size equal to  $L$ , which allows fast computation. Hence, the complexity of the deconvolution increases linearly in the number of events which increases for ion channel recordings typically linearly in the number of observations. In summary, for a typical time series the complexity to compute J-SMURF, JULES or JILTAD increases only linearly in the number of observations.

This is confirmed by measuring the run times of JULES and JILTAD for the times series generated in the Sections 3.9.6 and 3.9.7 which contain 600 000 observations each. For the time series with homogeneous noise in Section 3.9.6 the computation of JULES lasts on average 16 seconds and of JILTAD 30 seconds. For generated time series with heterogeneous noise in Section 3.9.7 the computation time of JILTAD is with roughly five minutes a little bit larger. Similar results are obtained for the real data applications. Thus, the theoretical considerations as well as the empirical run times confirm that these methods can be computed efficiently, as requested in Section 3.2. Moreover, we found that for JULES the run time is neither by the detection step nor by the deconvolution step dominated. Depending on the number of events, in some measurements detection lasts longer and in other measurements deconvolution lasts longer.

## 5. Application

In this section we present the application from (Bartsch et al., 2017). We also refer to this paper for more details about the biological and medical background, as well as for the interpretations of the results. Against the background of multidrug-resistant bacteria we explored together with the Steinem lab and other external collaborators the interaction of the antibiotic ampicillin with channels of the bacterial porin PorB. Ampicillin is a broad-spectrum antibiotic that inhibits the enzyme transpeptidase which is required for the cell division of bacteria (Acred et al., 1962). One potential source of antibiotic resistance is a prevented access of antibiotics through the outer bacterial membrane (Delcour, 2009). For its passage porins in the membrane play a key role (Delcour, 2003; Nikaido, 2003; Tanabe et al., 2010). In (Bartsch et al., 2017) the outer membrane porin PorB from *Neisseria meningitidis*, a pathogenic bacterium in the human nose and throat region (Virji, 2009), is studied. It is frequently present in *Neisseria meningitidis* (Tanabe et al., 2010) and has three equivalent pores formed by beta barrels. The wild type is compared with the mutant G103K, since cells with this mutation are suspected to be more likely resistant to antibiotics (Oppenheim, 1997; Olesky et al., 2002). In comparison, the mutant has a very similar protein structure but a reduced pore size and differently charged surfaces at the so called constriction zone, see (Bartsch et al., 2017, Figure 2). The constriction zone is formed by an internal loop that folds back into each pore and is the smallest point to pass the pore (Schulz, 2002). Hence, the electric fields in this region influence the transport heavily (Delcour, 2003; Nikaido, 2003).

Besides other methods, in particular molecular dynamic simulations, patch-clamp experiments were performed for a quantitative characterization of the interaction of the two types with ampicillin. When an ampicillin molecule binds to the pore it blocks the ion flow temporarily and, hence, this event can be detected by a conductance loss. Note that it cannot be decided whether an ampicillin molecule really passes through the channel or only enters the channels but leaves to the same side. Single channel recordings using solvent-free bilayers at the Port-a-Patch were used to explore the conductivity of the wild type and the mutant without presence of ampicillin. These measurements contain heterogeneous noise and hence no automatic analysis routine was available when writing the paper. We now analyze these measurements using JILTAD, see Section 5.2. Secondly, measurements using black lipid membranes with solvents were used to measure how frequently and how long ampicillin blocks the pores of the wild type and the mutant as a function of the ampicillin concentration and the applied voltage. The analysis of these quantities using JULES was our main contribution to the paper and the results are presented in Section 5.1. Notably,

the presence of solvent changed the protein properties severely. No heterogeneous noise is present, the channel is mainly open (97% instead of 20%) which is desirable for analyzing a blockage process, conductance losses of mono- and dimer size are not observed anymore and also the conductance levels are slightly different. We start with the second type of measurements to stay in the logic of this work to discuss first homogeneous noise and then heterogeneous noise.

## 5.1. Results for the recordings with ampicillin

In this section we present the results of our analysis of the measurements using planar black lipid membranes (BLMs). All of these measurements are recorded at sampling rate 50 kHz and were filtered with a four-pole Bessel lowpass filter with cutoff frequency 5 kHz, resulting in a normalized cutoff frequency of 0.1. A more detailed description of the measurements is given in (Bartsch et al., 2017, Materials and Methods). For the wild type as well as for the mutant G103K four measurements with 1 mM ampicillin concentration and at different applied voltage levels of 40, 60, 80, 100 and 120 mV were recorded. For both proteins in one of the recordings the ampicillin was added in steps to obtain measurements with different ampicillin concentrations of 0.1, 0.2, 0.4, 0.6, 0.8 and 1 millimol (mM) at 80 mV. In each measurement the recordings last at least five minutes and, hence, always at least five time series, with 3 million observations each, were available. An example was shown in Figure 1.1.

The underlying signals of each time series are estimated by JULES at significance level  $\alpha = 0.05$  without any prior data cleaning. An exemplary estimation was shown in the introduction in Figure 1.3. We stress that the residence times are much shorter than in similar studies before (James et al., 2009; Danelon et al., 2006; Mahendran et al., 2010) which made a deconvolution approach like in Section 3.6 mandatory. Estimation of the amplitudes, residence times and the frequencies were described in Section 3.8. A two state hidden Markov model was supported for all measurements. We stress again that this confirmation is only given for events with residence times between  $a = 0.08$  ms and  $b = 0.2$  ms, for all other lengths it is only an assumption. The estimated amplitudes are in average 1.19 nS for the wild type and 0.81 nS for the mutant. A comparison with the estimated amplitudes in Section 5.2, although influenced by the absence of solvent, suggests that the observed events are a blockage of all tree pores at the same time. Such a cooperative blockage by ampicillin was not found for porins before (James et al., 2009; Nestorovich et al., 2002), but for instance for trimeric chitoporin trapping sugars (Suginta et al., 2016). Blockage events of only one or two pores could not be observed which either means that they are not present or that these events have much shorter residence time such that they are not detectable.

Results of the estimated residence times and frequencies as functions of the voltage and the concentration are shown in Figures 5.1 and 5.2. We stress that the averaged residence times and frequencies based on estimations by JILTAD and a hidden Markov approach

were for all measurements close to the values based on an estimation by JULES and lead in all cases to the same conclusions. For the hidden Markov approach this is illustrated in (Bartsch et al., 2017, Figure S9, Supporting information). In this section we focus on the results obtained by JULES.

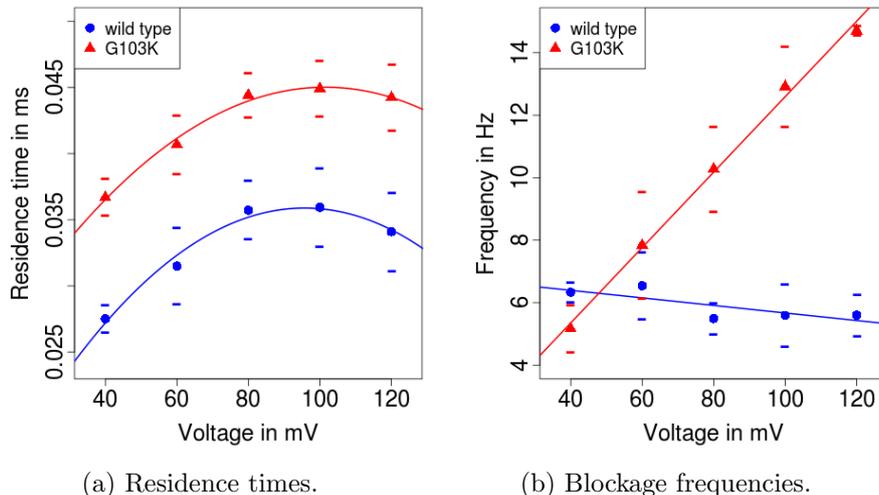


Figure 5.1.: Voltage-dependent residence times and blockage frequencies of ampicillin for PorB wild type and PorB G103K in the presence of 1 mM ampicillin. Four measurements were averaged for each protein. The bars indicate confidence intervals at 95% based on a normal approximation, but we stress that they are not reliable and only given for a rough tendency, see the discussion at the end of Section 3.8.

The residence times show a parabolic and the frequency a linear dependency on the voltage. For all voltage levels the residence times are statistically significant larger for the mutant than for the wild type, confirmed by the two-sample Wilcoxon signed-rank test at error level 0.05. The linear dependency of the frequencies on the voltage is different, for the mutant the frequency is increasing, but for the wild type decreasing. Interpretations of these major findings are given below. We found as expected no significant dependency of the residence times on the concentration. This means a larger amount of ampicillin molecules in the solution does not effect the single molecule in the pore. The frequencies increase linearly with the concentration which was also expected, since roughly described the ampicillin molecules diffuse trough the solution and enter the pore if they are close to it and in the right orientation. Consequently, in a fixed period of time each molecule has the same small probability to enter the pore which also explains the exponential distribution. If more molecules are available the waiting time is shorter, with a linear dependency. The larger slope for the mutant does not provide any new information when linearity and a larger value for the mutant than for the wild type at 1 mM is already known.

Molecular dynamics (MD) simulations revealed that during the passage through the pore an ampicillin molecule binds in the constriction zone to the channel protein. The binding is similar for the wild-type and the mutant, but the mutant G103K has one additional contact for ampicillin on the extracellular side of the constriction zone (Bartsch et al., 2017,

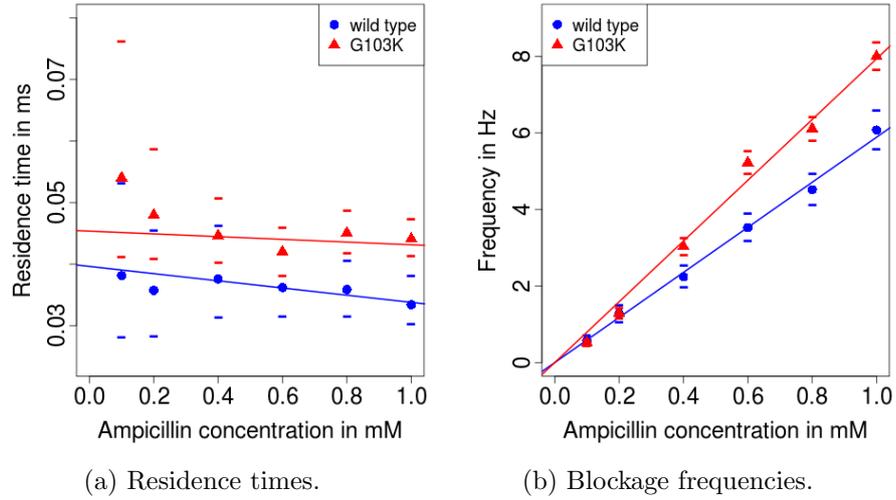


Figure 5.2.: Residence times and blockage frequencies at increasing ampicillin concentrations for PorB wild type and PorB G103K including confidence intervals (bars) at 95% significance. The recordings were performed at 80 mV. The linear fit weighted according to the confidence intervals shows no significant dependence of the residence times on the ampicillin concentration. Averaged residence times of 35  $\mu$ s for the wild type and of 44  $\mu$ s for the mutant were determined using weighted means. For both proteins, the frequencies increase linearly in the ampicillin concentration.

Figure 9) which results in a more stable bound to the pore and explains the longer residence times. The translocation of ampicillin require several reorientations of the ampicillin molecule. Thereby, different orientations of the ampicillin are allowed by the wild type and the mutant. An applied voltage changes the electric fields within the pore which makes reorientations easier or more difficult. Since different orientations are required for the wild type and the mutant protein, a different dependency on the applied voltage is reasonable. For more detailed explanations see (Bartsch et al., 2017, Figure 9C and the explanations around). In general a passage trough the mutant seems to be more difficult, i.e., higher energies are required, than trough the wild type which can explain an antibiotic resistance for cells with the G103K mutation. Such explanations can be helpful for the development of new drugs to mitigate the severe consequences of multidrug-resistant bacteria.

## 5.2. Analysis of PorB recordings with heterogeneous noise

In this section we analyze single channel recordings of the PorB wild type and of the mutant G103K using solvent-free bilayers at the Port-a-Patch. They are sampled with 10 kHz sampling rate and filtered with a four-pole lowpass Bessel filter with normalized cutoff frequency of 0.1. For each of the two proteins roughly ten measurements with one up to ten time series with 600 000 observations each are available. An example was shown in Figure 3.3. In this measurement the channel switches frequently between two conductance levels, roughly between 0.04 nS and 0.36 nS. The standard deviation in the closed

state is roughly  $\sqrt{6.1 \cdot 10^{-5}}$  nS and in the open state  $\sqrt{10^{-3}}$  nS. The noise is in many measurements of this order, but we also have measurements with (almost) homogeneous noise and two measurements with larger heterogeneity, there the variances are roughly  $\sqrt{2.5 \cdot 10^{-5}}$  nS and  $\sqrt{0.05}$  nS. Moreover, in Figure 3.3 several artifacts are present, see for instance the fluctuating base line in the first ten seconds. We stress that varying base lines, large peaks and other artifacts are present in many measurements, in some such that the channel dynamics could not be identified and we had to exclude these measurements from the analysis. Since PorB forms three pores, four different conductance levels are possible. However, in all measurements we analyzed the channel switches mostly between two conductance levels, but different amplitudes are found that correspond to the mono-, di- and trimer, see below. Such a cooperative opening and closing was observed before, see (Song et al., 1998).

Since most measurements contain heterogeneous noise we use JILTAD for estimating the signals, all parameters are chosen as described in Section 3.7. An exemplary estimation of the signal underlying the observations in Figure 3.3 was shown in Figure 3.5.

Amplitudes and dwell times are analyzed as outlined in Section 3.8 and for each measurement individually. Figure 5.3 shows a histogram of the estimated amplitudes in the measurement from Figure 3.3.

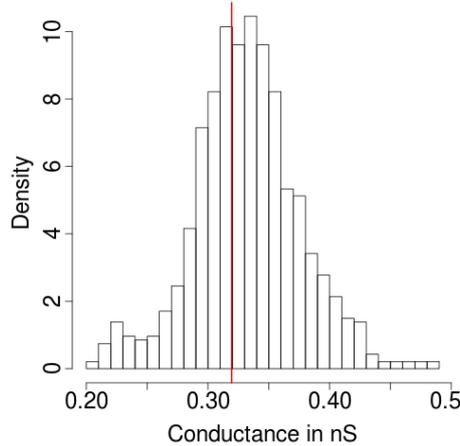


Figure 5.3.: Histograms of the amplitudes between  $l = 0.2$  nS and  $u = 0.5$  nS. Vertical red line indicates the estimated amplitude of 0.3194 nS by the half sample mode.

An amplitude of 0.3194 nS is estimated by the half sample mode. In total, the amplitudes are estimated for eight measurements of the mutant G103K and for seven measurements of the wild type. The estimated amplitudes cluster around 0.32 nS, 0.46 nS and 0.74 nS for the mutant and 0.42 nS, 0.91 nS and 1.65 nS for the wild type. These values are similar to the ones found in (Bartsch et al., 2017), with manually chosen events, of 0.13 nS, 0.36 nS and 0.76 nS for the mutant and 0.44 nS, 0.89 nS and 1.52 nS for the wild type. The only significant difference is an amplitude of 0.13 nS instead of 0.46 nS for the mutant. Note that in the histogram in (Bartsch et al., 2017, Figure 3 D) a peak at 0.46 nS is visible, but not found to be significant in the following estimation of the amplitudes by a Gaus-

sian mixture method. Instead a peak at 0.13 nS is found. Such events are visible in two measurements, but could not be distinguished from base line fluctuations in the analysis using JILTAD. In our analysis, however, three measurements have events with amplitudes around 0.46 nS which is to us a strong confirmation of such a level.

The amplitudes of the mutant are much smaller than those of the wild type. An explanation is given by atomistic simulations. The wild type has two distinct pathways for anions and cations, but in the mutant the cations pathway is disrupted, resulting in smaller conductance. For more details see (Bartsch et al., 2017, Figure 4) and the explanations around. In comparison to the amplitudes in Section 5.1 which correspond most likely to the amplitude of the whole channel, i.e., of all three pores simultaneously, the amplitudes are in these solvent-free measurements smaller for the wild type and slightly higher for the mutant. Such slight differences of the amplitudes for different membrane systems are common and reported for many channels (Neher and Eibl, 1977; Johannsson et al., 1981). We now analyze the dwell time in the open state and how frequently the channel opens. We take into account events with an amplitude between  $l = 0.2$  nS and  $u = 0.5$  nS and with a dwell time in the open state between  $a = 0.1$  ms and  $b = 200$  ms. Histograms of the dwell time in the open state are shown in Figure 5.4 for the measurement in Figure 3.3.

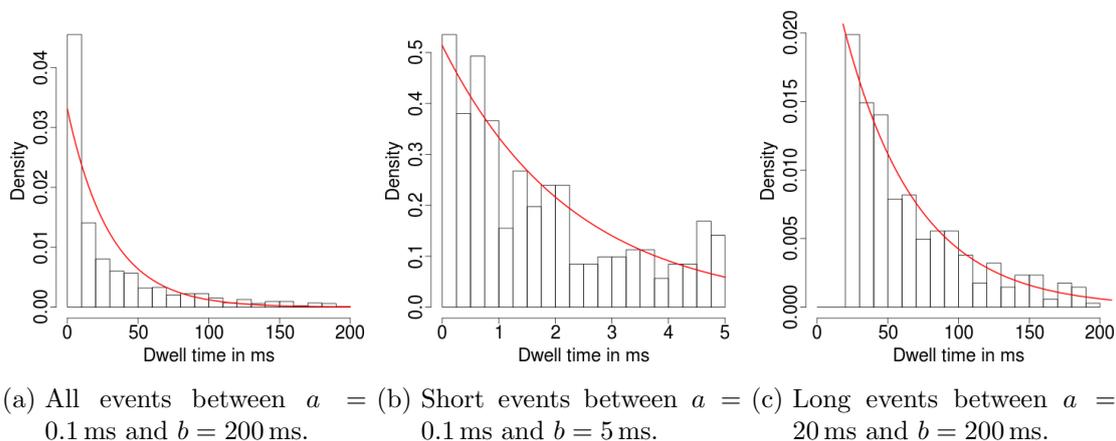
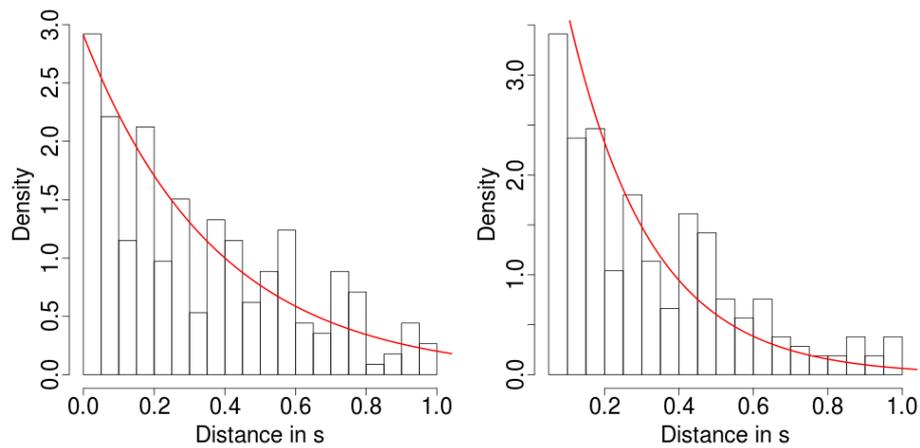


Figure 5.4.: Histograms of the dwell times of opening events with amplitude between  $l = 0.2$  nS and  $u = 0.5$  nS.

Interestingly, the dwell times do not fit a single exponential distribution, but when we split the events in short and long ones, both fit exponential distributions very well, with an estimated average duration of 51.62 ms and 2.31 ms, respectively. Note that these estimations are approximations, since an exponential distribution with a large rate generates with a small probability a long event and vice versa, but since the average dwell times are very different this error is negligible. Unfortunately, the result of a mixture of two exponential distributions could not be confirmed by further measurements using solvent-free membranes, since either the amount of events is very small or the gating events cannot be separated from detected artifacts, but by measurements using black lipid membranes without ampicillin. This hints that the PorB channels show flickering and normal gating

at the same time. Next, we analyze how frequently both types of events occur, see Figure 5.5. For events with a long dwell time a missed event correction was necessary and we only took into account distances between 50 ms and 1 s, since artifacts seems to disturb shorter distances.



(a) Distance between two events with dwell time between  $a = 0.1$  ms and  $b = 5$  ms. (b) Distance between two events with dwell time between  $a = 20$  ms and  $b = 200$  ms.

Figure 5.5.: Histograms of the distances between two short and between two long opening events with amplitude between  $l = 0.2$  nS and  $u = 0.5$  nS.

In both cases the distances seem to be exponentially distributed and the estimated rates are 2.67 Hz for the short events and 4.50 Hz for the long events. Hence, longer events occur more often.

## 6. Conclusion and outlook

### 6.1. Conclusion

This work provides two major contributions. First of all, we presented in Section 2 with H-SMUCE an estimator for heterogeneous multiscale change-point inference and, in particular, established a comprehensive theory for the heterogeneous change-point model. Secondly, we obtained in Section 3.5 with JULES for homogeneous observations and in Section 3.7 with JILTAD for heterogeneous observations, but also usable for homogeneous observations, two non-parametric methods for estimating the piecewise constant signal underlying ion channel recordings. Their importance was illustrated by an electrophysiology study in the severe context of multidrug-resistant bacteria, see Section 5.

H-SMUCE recovers the change-points well, while at the same time it is robust against heterogeneous noise and controls the probability to overestimate the number of change-points. More precisely, the number of change-points is estimated consistently, see Theorem 9, and vanishing signals are detected at the optimal minimax rate, see Theorems 12 and 13. In addition, simultaneous confidence intervals for the change-point locations, see Theorem 11, and a confidence band for the unknown signal, see the end of Section 2.3.2, are obtained. Simulations show that H-SMUCE outperforms state of the art methods as long as the constant segments are not too short, see Figure 2.2, and is robust against heterogeneous noise even on constant signal segments, see Table 2.4, and as a by-product also against more heavily tailed errors, see Table 2.5. Moreover, it can be computed efficiently by a pruned dynamic program, see Section 4.1. For instance if change-points occur frequently the computation time increases only linearly in the number of observations, see Section 4.5. Together with other multiscale regression estimators, H-SMUCE is implemented in the *R* package *stepR* (Pein et al., 2017a).

Patch clamp recordings to quantify the gating dynamics of ion channels are a major tool in electrophysiology and a first step of their analysis is often the estimation of the underlying signals from filtered observations. Our proposed non-parametric methods JULES and JILTAD estimate these signals by combining multiscale estimation with local deconvolution. Simulations, Section 3.9, and real data applications, Sections 3.10 and 5, confirm that their estimation is still accurate when events occur on very small temporal scales, in particular also on scales below the filter length, where the smoothing effect of the filter hinders estimation by common methods. Moreover, JILTAD estimates the underlying signals still well when open channel noise causes heterogeneous noise, a situation for which previously no non-parametric estimation method existed. At the same time, we showed

in Section 4.5 that both methods can be computed efficiently, usually its complexity increases only linearly in the number of observations. Hence, an estimation for hundred thousands up to few millions observations lasts typically only seconds up to few minutes on a standard laptop. An *R* package *clampSeg* (Pein et al., 2017b) specialized for the analysis of ion channel recordings provides an implementation of JULES, while JILTAD is currently only available in an internal version, but will be published soon.

All in all, the simulations and data applications establish JULES and JILTAD as reliable, universally applicable, fully-automatic and efficiently computable non-parametric estimators for the analysis of ion channel recordings, JILTAD even in the presence of heterogeneous noise. Only when events follow each other very quickly they are not applicable as proposed, but can be refined at computational expenses.

We used in Section 5 these methods to explore the interaction of the antibiotic ampicillin with the outermembrane porin PorB. The very short blockage times and the huge amount of observations and events required an automatic analysis of these recordings with high precision on small temporal scales. We found that the average residence time of ampicillin is statistically significantly longer for the PorB mutant G103K than for the wild type. In conjuncture with other findings this suggests that ampicillin passes the mutant less likely which explains that bacteria with this mutation have an increased resistance against antibiotics. Such explorations help to develop new drugs against resistant bacteria.

## 6.2. Outlook

### 6.2.1. Varying voltage

So far we only considered ion channel recordings with a constant applied voltage, but also experiments with a varying voltage are interesting. That's because not only the voltage effects the channel, some channels are also effected by voltage changes. This includes channels that show no gating when the voltage is constant, but can be activated by a varying voltage. For other channels different dynamics are observed when the voltage changes. One example is the protein channel Tim23 which shows severe *stress* effects. More precisely, the channel tends to close when the voltage is high for a longer time period, while when a varying voltage is applied the channel shows regular gating (for a constant voltage the same applies for a short time period after a voltage change). And even if a channel is not effected by voltage changes such experiments are interesting, since often the dynamics at different voltage levels are examined, see for instance the analysis in Section 5.1. While experiments with a constant voltage only allow to examine the gating dynamics at few voltage points (or require huge experimental effort), with varying voltage the dynamics can be analyzed for a whole range of voltages by a single experiment. However, the analysis of such experiments is more difficult. In fact, recordings with varying voltage might be of greater interest if better analysis routines are available.

To simplify the model we restrict ourself to recordings with a linear voltage. Since the

conductance is still assumed to change only abruptly, it follows from Ohm's law that the current can be modeled as a piecewise linear function. Thereby, when the conductance changes the slope changes, while the intercept is assumed to be a global constant. The intercept is under perfect conditions zero, but often the electronic and other effects cause a small intercept. An example of such a recording is shown in Figure 6.1.

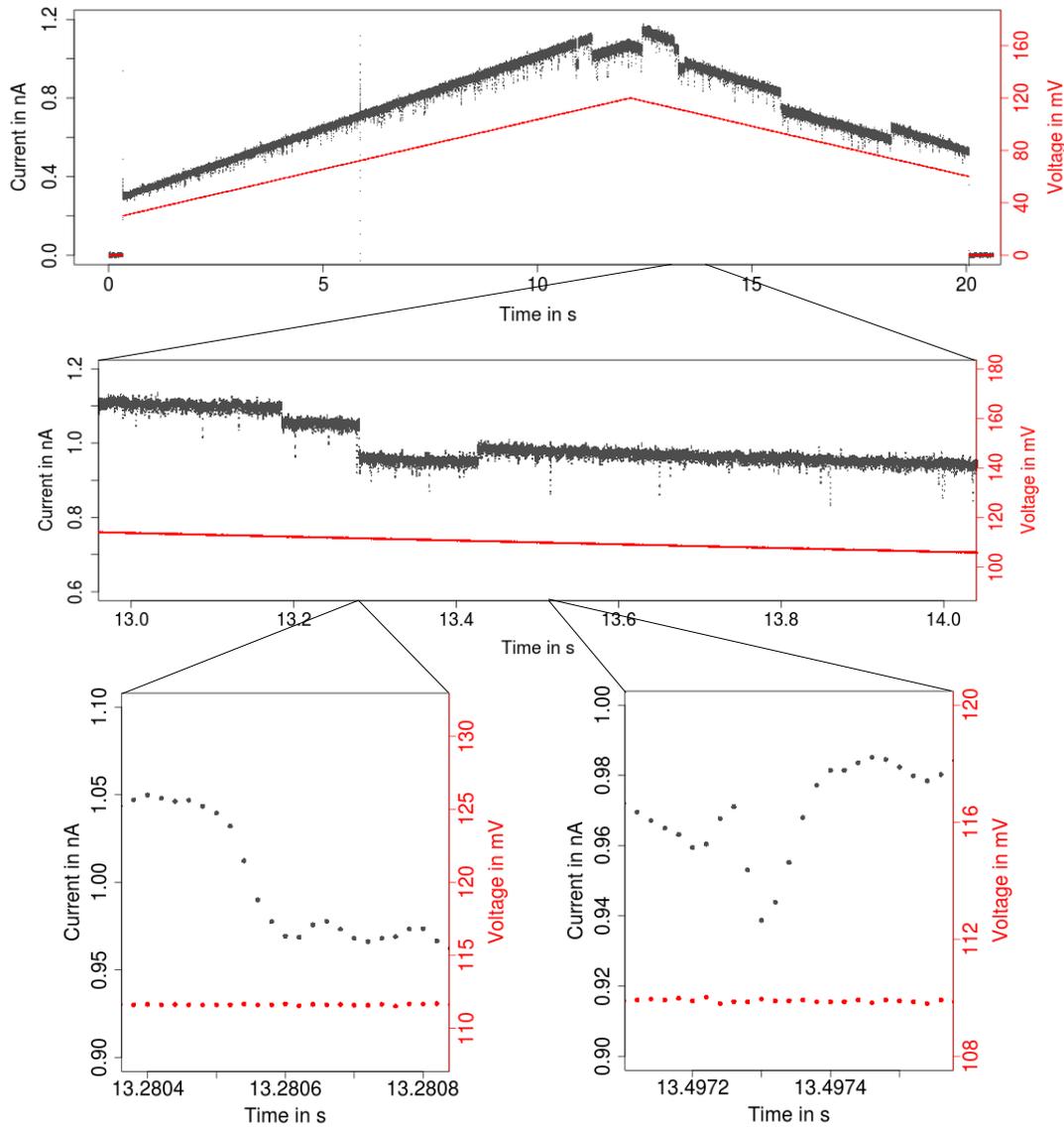


Figure 6.1.: Observations (grey points) of a representative current time series of PorB wild type with 1 mM ampicillin while a voltage ramp (red points) is applied. The time series is recorded by the patch clamp technique using black lipid membranes.

In addition to short blockage events by the ampicillin as in the recordings with constant voltage, we see gating events on larger time scales. Since such events are not observed when a constant voltage is applied, they are almost certainly caused by the voltage changes. The zooms reveal that the recording looks locally as the recordings with constant voltage. This

was expected, since the voltage varies only slowly.

An empirical data analysis confirms the model and suggests that the variance is quadratic in the applied voltage. The parameters of the variance can be assumed to be global, i.e., not affected by conductance changes. The global parameters can be determined by an empirical analysis of longer linear segments, i.e., parts in which the conductance does not change. Hence, we assume that the variance is known and without loss of generality that the intercept is zero, since the global intercept can be subtracted from the observations. Filtering and sampling is not influenced by the varying voltage. Thus, we model the recorded current observations by

$$Y_i = \left( F_L * (\mu \cdot U + \eta \sqrt{\sigma_0 + \sigma_1 \cdot U + \sigma_2 \cdot U^2}) \right) \left( \frac{i}{f_s} \right), \quad i = 1, \dots, n, \quad (6.1)$$

with  $F_L, \mu$  and  $\eta$  as in Section 3.1.1,  $U$  the voltage function in time and  $\sigma_0, \sigma_1, \sigma_2 \in \mathbb{R}$  such that  $\sigma_0 + \sigma_1 U(t) + \sigma_2 U(t)^2 > 0$  for all  $t \in \mathbb{R}$ . For given signal  $\mu$  and variance parameters  $\sigma_1, \sigma_2, \sigma_3$  the expectation and covariance of the observations can still be computed explicitly. For instance the convolution of a linear increasing function with the kernel is also a linear increasing function but with a smaller slope.

Although of its great practical importance, to the best of our knowledge no estimation method assumes this model or a similar one. Besides an analysis by manual chosen event times, the most common approach to analyze such recordings is to divide by the applied voltage. The resulting signal is piecewise constant, but the variance is not constant, which however is assumed by most approaches, and in particular very large when the voltage is close to zero. To overcome these issues, we aim to extend J-SMURF, JULES and JILTAD to model (6.1). In the following we outline exemplarily an extension of JULES, but an extension of J-SMURF and JILTAD should be possible as well.

For the local test statistics of the multiscale test the partial sum statistic can be modified to

$$T_i^j(\mathbf{Y}, m_{ij}) := \frac{\left( \sum_{l=i}^j Y_l - (F_L * (m_{ij} U))(l/f_s) \right)^2}{2\text{Var} \left[ \sum_{l=i}^j Y_l \right]}. \quad (6.2)$$

Note that, as mentioned before, the convolution and the variance can be computed explicitly. The required modifications for the computation of the critical values and of the multiscale regression estimator as well as for the postfilter step are straightforward. Also the local deconvolution approach can be used with slight adaptations. Here, maximizing the likelihood function is again equivalent to minimizing (3.22), but  $\mathbf{m}_{i,j}$  writes now as

$$\mathbf{m}_{i,j} = \left( (F_L * (\mu \cdot U)) \left( \frac{f_s \hat{\tau}_i - L + 1}{f_s} \right), \dots, (F_L * (\mu \cdot U)) \left( \frac{f_s \hat{\tau}_j + L - 1}{f_s} \right) \right)^t$$

and  $\Sigma_{i,j}$  is the known covariance matrix. Since only very few neighboring observations are involved, the covariance matrix is almost a Toeplitz matrix. Hence, Tikhonov regularization can still be used, with regularization parameter for instance the mean of the

variances of the involved observations.

### 6.2.2. Fixed levels

A small alphabet  $\{l_1, \dots, l_c\}$  of potential conductance levels is a common assumption to analyze ion channel recordings, in particular by hidden Markov models, see Section 3.8. If open channel noise causes heterogeneous noise, we assume an alphabet  $\{(l_1, v_1^2), \dots, (l_c, v_c^2)\}$  of expectation and variance pairs. So far we did not assume an alphabet of fixed levels for our non-parametric methods, in particular to be robust against artifacts. Consider exemplarily the situation that the conductance is for a second in the mid of two conductance levels due to an artifact. The approaches in Section 3 will estimate one change-point at the beginning and one at the end of this segment, which does not affect the whole analysis noticeably. In comparison, a method that assumes only the two conductance levels is forced to include many changes between the two conductance levels to obtain a convolution that is roughly at the observed conductance level. Hence, a single artifact can disturb the whole analysis.

Despite these difficulties, such an approach is interesting, since it allows a better estimation of very short events if artifacts does not disturb their estimation. To this end, artifacts can be removed by data cleaning or be taken into account by postfiltering. In the following we focus on the setting from Section 5.2 and on JILTAD assuming heterogeneous noise. Recall, that we found only two different levels in each measurement, but heterogeneous noise. Consequently, we assume the alphabet  $\{(l_1, v_1^2), (l_2, v_2^2)\}$ . Note that this setting is simpler as a setting with more levels, but at the same time it also allows the largest improvement in comparison to not assuming fixed levels. The levels can either be pre-estimated by long segments or by using JILTAD as described in Section 3.7.

We now outline how JILTAD can be adapted to this model. First of all, we apply J-SMURF without any adaptation, but project afterwards the estimated function values to the assumed alphabet. A simple approach is to use the alphabet value that has the closest Euclidean distance between the estimated function and the alphabet value for the expectation. Alternatively, more elaborate projections can be used, they may include an variance estimation on the constant segments. Note that after the projection some jump sizes are zero. We remove these change-points and assume for the following that  $\hat{m}_{k-1} \neq \hat{m}_k$  for all  $k = 1, \dots, \hat{K}$ . Next, we reestimate the change-point locations by local deconvolution, which is in this setting more accurate, since only the locations, but not the function values have to be estimated. Also for this estimation problem the corresponding likelihood function from the homogeneous case can be used. And we still suggest to use an iterative grid search to compute the estimator, since despite the simplification the optimization problem is still highly non-convex. Finally and most important, for the tests on small scales we take into account the assumed levels. We stress that this improves detection power severely, since the null hypothesis and the alternative are now both point hypothesis. Hence, a test similar to the one in (Enikeeva et al., 2016) can be used. More precisely, the filtering

has to be taken into account, for instance as in Section 3.7.4 by ignoring the correlation structure. The final change-point locations can again be estimated by deconvolution as described previously.

### 6.2.3. Parallel programming

Modern servers allow computation on multiple cores at the same time to speed up computations. In the following we want to outline how our algorithms can be parallelized. For Monte-Carlo simulations this is trivial, since  $M$  times the same steps have to be performed independently of each other. Here it can be expected that the computation time reduces almost exactly by the number of available cores. Almost the same applies for the local deconvolution. The computation time is dominated by the iterative grid search which has to be performed for each single event individually and hence parallelization is straightforward. Also the local tests in JILTAD are of a similar structure. Thus, in the following we focus on the dynamic program as described in Section 4.1.1.

The computation of the left limit  $L_k$  requires the computation of the intersected bounds  $\underline{B}_{r,L_{k+1}-1}$  and  $\overline{B}_{r,L_{k+1}-1}$  for all  $r = L_{k+1} - 1, \dots, 1$  until  $\underline{B}_{r,L_{k+1}-1} > \overline{B}_{r,L_{k+1}-1}$ . To simplify notation we ignore the upper bounds and define  $L := L_{k+1} - 1$ . The intersected bounds can be computed iteratively by

$$\underline{B}_{r,L} = \max\{\underline{B}_{r+1,L}, \underline{B}_r\}, \text{ with } \underline{B}_r := \max_{\substack{s=r,\dots,L \\ [r/n,s/n] \in \mathcal{I}}} \underline{b}_{r,s}. \quad (6.3)$$

The maxima  $\underline{B}_r$ ,  $r = L, \dots, 1$ , can be computed by different processes at the same time. The process that computes  $\underline{B}_r$  needs usually a little bit more time than the process that computes  $\underline{B}_{r+1}$ , since it has to compute one bound more. Otherwise, the process that computes  $\underline{B}_r$  has to wait until  $\underline{B}_{r+1}$  is computed to compute  $\underline{B}_{r,L}$  by the recursion formula (6.3). And if  $\underline{B}_{r,L} > \overline{B}_{r,L}$  for one  $r$  all processes can be stopped and the computation of  $L_{k-1}$  can begin.

A similar approach is possible to compute the cumulated costs  $c_p$ ,  $p = L_k, \dots, R_k - 1$ , at the same time. The computation of  $c_{p+1}$  requires to compute  $c_r + c_{r+1,p+1}$  for  $r = R_{k-1} - 1, \dots, L_{k-1}$ . The computation of  $c_{r+1,p+1}$  require the intersected bound  $\underline{B}_{r+1,p+1} = \max\{\underline{B}_{r+1,p}, \underline{B}_{r+2,p+1}, \underline{b}_{r+1,p+1}\}$ , but the intersected bound  $\underline{B}_{r+1,p}$  have also to be computed to compute  $c_{r+1,p}$ . Hence, the process that computes  $c_{p+1}$  computes  $\underline{b}_{r+1,p+1}$  and waits until the process that computes  $c_p$  has computed  $\underline{B}_{r+1,p}$ . Note that the process that computes  $c_{p+1}$  has the intersected bound  $\underline{B}_{r+2,p+1}$  already computed in the previous step, since it was required to compute  $c_{r+1} + c_{r+2,p+1}$ . Also note that the computation of  $\underline{B}_{r+1,p}$  only requires the computation of  $\underline{b}_{r+1,p}$  and to wait for  $\underline{B}_{r+1,p-1}$ , since a similar recursion applies. Hence, the waiting times are usually small. The stopping conditions are similar to before. This shows that the dynamic program can be parallelized. Since at certain steps some processes might have to wait until another process finished required computations, the overall reduction will be less than by the number of cores, but still significant.

# A. Appendix

In this section we collect the proofs together with some auxiliary statements.

## A.1. Proof of Lemma 1

*Proof of Lemma 1.* A single statistic  $T_i^j(\mathbf{Z}, 0)$  has the c.d.f.  $F_{1,j-i}(\cdot)$  of an F-distribution with  $(1, j - i)$  degrees of freedom. Thus,  $F_k(\cdot) = F_{1,2^k-1}(\cdot)^{|\mathcal{D}_k|}$  is continuous and strictly monotonically increasing for positive arguments. Now, it follows from equation (2.14) that

$$q_k = F_k^{-1} \left( 1 - \frac{\beta_k}{\beta_1} (1 - F_1(q_1)) \right) \text{ for } k = 2, \dots, d_n. \quad (\text{A.1})$$

This together with equation (2.12) yields

$$G(q_1) := F \left( q_1, F_2^{-1} \left( 1 - \frac{\beta_2}{\beta_1} (1 - F_1(q_1)) \right), \dots, F_{d_n}^{-1} \left( 1 - \frac{\beta_{d_n}}{\beta_1} (1 - F_1(q_1)) \right) \right) = 1 - \alpha.$$

Note, that  $F$  is continuous and  $\lim_{q_k \rightarrow 0} F(q_1, \dots, q_{d_n}) = 0$  for all  $k = 1, \dots, d_n$  as well as  $\lim_{q_1, \dots, q_{d_n} \rightarrow \infty} F(q_1, \dots, q_{d_n}) = 1$ . Thus, the function  $G$  is continuous, strictly monotonically increasing on  $[0, \infty)$  and attains all values in  $[0, 1)$ . Therefore, the existence of the vector of critical values follows from the intermediate value theorem and the vector is also unique.  $\square$

## A.2. Proof of Lemma 3

First of all, recall from the proof of Lemma 1 that the statistic  $T_k$  has c.d.f.  $F_{1,2^k-1}(\cdot)^{|\mathcal{D}_k|}$ . For every  $k = 1, \dots, d_n$  we use the transformation

$$U_k := F_{1,2^k-1}(T_k)^{|\mathcal{D}_k|}$$

and the identity

$$T_k = F_{1,2^k-1}^{-1} \left( U_k^{|\mathcal{D}_k|^{-1}} \right).$$

Here,  $F_{1,2^k-1}^{-1}(\cdot)$  denotes the quantile function of an F-distribution with  $(1, 2^k - 1)$  degrees of freedom. Analogously, we define

$$q_{k,U} := F_{1,2^k-1}(q_k)^{|\mathcal{D}_k|}$$

and have the identity

$$q_k = F_{1,2^{k-1}}^{-1} \left( q_{k,U}^{|\mathcal{D}_k|^{-1}} \right). \quad (\text{A.2})$$

Then, the events  $U_k > q_{k,U}$  and  $T_k > q_k$  are equivalent and therefore the vector  $\mathbf{q}_U = (q_{1,U}, \dots, q_{d_n,U})$  satisfies similar conditions to the equations (2.12) and (2.14), i.e.,

$$1 - \mathbb{P}(U_1 \leq q_{1,U}, \dots, U_{d_n} \leq q_{d_n,U}) = \alpha \quad (\text{A.3})$$

and

$$\frac{1 - \mathbb{P}(U_1 \leq q_{1,U})}{\beta_1} = \dots = \frac{1 - \mathbb{P}(U_{d_n} \leq q_{d_n,U})}{\beta_{d_n}}. \quad (\text{A.4})$$

The following bounds can be interpreted as a weighted version of the Bonferroni-inequality.

**Lemma 15.**  $q_{k,U} \leq 1 - \alpha\beta_k$  for  $k = 1, \dots, d_n$ .

*Proof.* We have  $\mathbb{P}(U_j \leq q_{j,U}) = q_{j,U}$  for  $j = 1, \dots, d_n$ , since  $U_j$  is uniformly distributed. Moreover, it follows from condition (A.4) that  $1 - q_{j,U} = (1 - q_{k,U})\beta_j/\beta_k$ . Combining this with equation (A.3) and  $\sum_{j=1}^{d_n} \beta_j = 1$  yields

$$\begin{aligned} \alpha &= 1 - \mathbb{P}(U_1 \leq q_{1,U}, \dots, U_{d_n} \leq q_{d_n,U}) \\ &\leq \sum_{j=1}^{d_n} \mathbb{P}(U_j > q_{j,U}) = \sum_{j=1}^{d_n} (1 - q_{j,U}) = \sum_{j=1}^{d_n} (1 - q_{k,U}) \frac{\beta_j}{\beta_k} = \frac{1 - q_{k,U}}{\beta_k}, \end{aligned}$$

which proves the assertion.  $\square$

Lemma 16 bounds the quantile function of an F-distribution with  $(1, c)$  degrees of freedom.

**Lemma 16** (Bounds on the F-quantiles). *Let  $F_{1,c}^{-1}(y)$  be the quantile function of an F-distribution with  $(1, c)$  degrees of freedom, then*

$$c \left[ (1 - y^2)^{-\frac{1}{c}} - 1 \right] \leq F_{1,c}^{-1}(y) \leq c \left[ (1 - y^2)^{-\frac{2}{c-\frac{1}{2}}} - 1 \right].$$

*Proof.* We have from (Fujikoshi and Mukaihata, 1993, Theorem 4.2) that

$$c \left[ \exp \left( \frac{(\chi_1^2)^{-1}(y)}{c} \right) - 1 \right] \leq F_{1,c}^{-1}(y) \leq c \left[ \exp \left( \frac{(\chi_1^2)^{-1}(y)}{c - \frac{1}{2}} \right) - 1 \right],$$

with  $(\chi_1^2)^{-1}(y)$  the quantile function of the chi-squared distribution with one degree of freedom and  $Z$  a standard Gaussian distributed random variable. Moreover, we obtain for all  $y \geq 0$

$$\begin{aligned} \mathbb{P}(\chi_1^2 \leq y) &= \mathbb{P}(-\sqrt{y} \leq Z \leq \sqrt{y}) = 2\Phi(\sqrt{y}) - 1 \\ \iff (\chi_1^2)^{-1}(y) &= \Phi^{-1} \left( \frac{y+1}{2} \right)^2, \end{aligned}$$

where  $\Phi^{-1}(y)$  is the quantile of the standard Gaussian distribution. Furthermore, we have from (Johnson et al., 1994, (13.48), p. 115) that

$$\frac{1}{2} \left[ 1 + \left( 1 - \exp \left( -\frac{x^2}{2} \right) \right)^{\frac{1}{2}} \right] \leq \Phi(x) \leq \frac{1}{2} \left[ 1 + \left( 1 - \exp \left( -x^2 \right) \right)^{\frac{1}{2}} \right]$$

and so for the quantile function one finds

$$\sqrt{-\log(1 - (2y - 1)^2)} \leq \Phi^{-1}(y) \leq \sqrt{-2 \log(1 - (2y - 1)^2)}.$$

Combining the formulas proves the assertion.  $\square$

*Proof of Lemma 3.* First of all, (A.2) and the equation  $|\mathcal{D}_k| = \lfloor n2^{-k} \rfloor$  yields

$$q_k = F_{1,2^{k-1}}^{-1} \left( q_{k,U}^{|\mathcal{D}_k|^{-1}} \right) = F_{1,2^{k-1}}^{-1} \left( q_{k,U}^{\lfloor n2^{-k} \rfloor^{-1}} \right) \leq F_{1,2^{k-1}}^{-1} \left( q_{k,U}^{2^k/n} \right).$$

Moreover, it follows from the Lemmas 15 and 16 that

$$\begin{aligned} q_k &\leq F_{1,2^{k-1}}^{-1} \left( q_{k,U}^{2^k/n} \right) \leq F_{1,2^{k-1}}^{-1} \left( (1 - \alpha\beta_k)^{2^k/n} \right) \\ &\leq (2^k - 1) \left[ \left( 1 - \left( (1 - \alpha\beta_k)^{2^k/n} \right)^2 \right)^{-\frac{2}{(2^k-1)-\frac{1}{2}}} - 1 \right] \\ &\leq 2^k \left[ \left( 1 - (1 - \alpha\beta_k)^{2^k/n} \right)^{-\frac{4}{2^{k+1}-3}} - 1 \right]. \end{aligned}$$

Applying Bernoulli's inequality  $(1 - x)^c \leq 1 - cx$  gives

$$q_k \leq 2^k \left[ \left( 1 - (1 - \alpha\beta_k)^{2^k/n} \right)^{-\frac{4}{2^{k+1}-3}} - 1 \right] \leq 2^k \left[ \left( \frac{2^k \alpha\beta_k}{n} \right)^{-\frac{4}{2^{k+1}-3}} - 1 \right].$$

Moreover, for  $x, c > 0$  the inequality  $c^x \leq 1 + 2x \log(c)$  holds whenever  $x \log(c) \leq 1$ . Together with the assumption  $k \geq 2$  we finally obtain

$$q_k \leq 2^k \left[ \left( \frac{2^k \alpha\beta_k}{n} \right)^{-\frac{4}{2^{k+1}-3}} - 1 \right] \leq 4 \frac{2^{k+1}}{2^{k+1}-3} \log \left( \frac{n}{2^k \alpha\beta_k} \right) \leq 8 \log \left( \frac{n}{2^k \alpha\beta_k} \right)$$

if

$$2^{-k} \log \left( \frac{n}{2^k \alpha\beta_k} \right) \leq \frac{1}{2} \frac{2^{k+1} - 3}{2^{k+1}} \leq \frac{1}{2}.$$

$\square$

### A.3. Exponential deviation bounds

For the subsequent proofs we have to bound the distribution function of a single test statistic  $T_i^j$  (1.7) under the alternative. This is for the setting in Section 2 a bound for

the c.d.f. of a non-central F-distribution.

**Lemma 17.** *Let  $\mathbf{Y} = (Y_1, \dots, Y_n)$  be a vector of i.i.d. Gaussian distributed random variables with expectation  $m \in \mathbb{R}$  and variance  $s^2 > 0$ . Let  $x_+ := \max(x, 0)$ . Then, for any  $\delta \neq 0$ ,  $q > 0$*

$$\begin{aligned} & \mathbb{P}(T_1^n(\mathbf{Y}, m + \delta) \leq q) \\ & \leq \min_{z \geq 0} \left\{ \exp \left( -\frac{1}{2} \left( \frac{\Delta\sqrt{n}}{2} - \frac{q(1+z)}{\Delta\sqrt{n}} \right)_+^2 \right) + \exp \left( -(n-1) \frac{z - \log(1+z)}{2} \right) \right\}, \end{aligned} \quad (\text{A.5})$$

where  $\Delta := |\delta|/s$ .

*Proof.* Let  $\tilde{T}_i^j(\mathbf{Y}, m) := (j - i + 1) (\bar{Y}_{ij} - m)^2 / s^2$ . Then,

$$T_1^n(\mathbf{Y}, m + \delta) = \frac{\tilde{T}_1^n(\mathbf{Y}, m + \delta)}{\hat{s}_{1n}^2/s^2}.$$

The statistics  $\hat{s}_{1n}^2/s^2$  and  $\tilde{T}_1^n(\mathbf{Y}, m + \delta)$  are independent, since  $\tilde{T}_1^n(\mathbf{Y}, m + \delta)$  depends only on the mean  $\bar{Y}_{1n}$ . Hence, for all  $z \geq 0$

$$\begin{aligned} \mathbb{P}(T_1^n(\mathbf{Y}, m + \delta) \leq q) &= \mathbb{P} \left( \tilde{T}_1^n(\mathbf{Y}, m + \delta) \leq q \frac{\hat{s}_{1n}^2}{s^2} \right) \\ &= \mathbb{P} \left( \tilde{T}_1^n(\mathbf{Y}, m + \delta) \leq q \frac{\hat{s}_{1n}^2}{s^2} \mid \frac{\hat{s}_{1n}^2}{s^2} \leq 1+z \right) \mathbb{P} \left( \frac{\hat{s}_{1n}^2}{s^2} \leq 1+z \right) \\ &\quad + \mathbb{P} \left( \tilde{T}_1^n(\mathbf{Y}, m + \delta) \leq q \frac{\hat{s}_{1n}^2}{s^2} \mid \frac{\hat{s}_{1n}^2}{s^2} > 1+z \right) \mathbb{P} \left( \frac{\hat{s}_{1n}^2}{s^2} > 1+z \right) \\ &\leq \mathbb{P} \left( \tilde{T}_1^n(\mathbf{Y}, m + \delta) \leq q(1+z) \right) + \mathbb{P} \left( \frac{\hat{s}_{1n}^2}{s^2} > 1+z \right) \\ &\leq \exp \left( -\frac{1}{2} \left( \frac{\Delta\sqrt{n}}{2} - \frac{q(1+z)}{\Delta\sqrt{n}} \right)_+^2 \right) + \exp \left( -(n-1) \frac{z - \log(1+z)}{2} \right). \end{aligned}$$

The first term of the last inequality follows from (Frick et al., 2014a, Lemma 7.3 and the proof) and the second from (Spokoiny and Zhilova, 2013, Theorem 2.1), since  $(n-1)\hat{s}_{1n}^2/s^2 \sim \chi_{n-1}^2$ .

It remains to show that the minimum in (A.5) is attained for some  $z \geq 0$ . The function  $(\Delta\sqrt{n}/2 - q(1+z)/(\Delta\sqrt{n}))_+^2$  is strictly monotonically decreasing for  $z > 0$  until the function value zero is attained for some finite  $z$ . The function  $(n-1)(z - \log(1+z))$  is zero for  $z = 0$  and strictly monotonically increasing on  $[0, \infty)$ . Therefore, the two continuous functions intersect and the minimum is attained for some  $z \geq 0$ .  $\square$

The minimum in Lemma 17 cannot be determined analytically, but it can be computed numerically. In Lemma 18 we estimate the right hand side further to obtain an explicit exponential bound.

**Lemma 18.** Let  $\mathbf{Y} = (Y_1, \dots, Y_n)$ ,  $n \geq 4$ , be a vector of i.i.d. Gaussian distributed random variables with expectation  $m \in \mathbb{R}$  and variance  $s^2 > 0$ , then we have for all  $q > 0$  with

$$\frac{q}{n} \leq \frac{1}{8} \quad (\text{A.6})$$

as well as for all  $\delta \neq 0$  and  $\Delta := |\delta|/s$  the bound

$$\mathbb{P}(T_1^n(\mathbf{Y}, m + \delta) \leq q) \leq 2 \exp\left(-\frac{1}{48} \left(\sqrt{n}\Delta - \sqrt{2q}\right)_+^2\right). \quad (\text{A.7})$$

*Proof.* Let  $z > 0$  be arbitrary, but fixed. Then, it follows from Lemma 17 that

$$\begin{aligned} \mathbb{P}(T_1^n(\mathbf{Y}, m + \delta) \leq q) &\leq \exp\left(-\frac{1}{2} \left(\frac{\Delta\sqrt{n}}{2} - \frac{q(1+z)}{\Delta\sqrt{n}}\right)_+^2\right) + \exp\left(-\frac{(n-1)(z - \log(1+z))}{2}\right) \\ &\leq 2 \exp\left(-\min\left[\frac{1}{2} \left(\frac{\Delta\sqrt{n}}{2} - \frac{q(1+z)}{\Delta\sqrt{n}}\right)_+^2, (n-1)\frac{z - \log(1+z)}{2}\right]\right). \end{aligned}$$

The inequality

$$z - \log(1+z) \geq \frac{1}{2} \frac{z^2}{1+z} \geq \frac{1}{4} \min(z^2, z)$$

yields

$$\begin{aligned} &\mathbb{P}(T_1^n(\mathbf{Y}, m + \delta) \leq q) \\ &\leq 2 \exp\left(-\min\left[\frac{1}{8}n \left(\Delta - \frac{2q(1+z)}{\Delta n}\right)_+^2, \frac{1}{8}(n-1) \min(z^2, z)\right]\right) \\ &\leq 2 \exp\left(-\frac{1}{8}(n-1) \min\left[\min\left[\left(\Delta - \frac{2q(1+z)}{\Delta n}\right)_+^2, z^2\right], \min\left[\left(\Delta - \frac{2q(1+z)}{\Delta n}\right)_+^2, z\right]\right]\right). \end{aligned}$$

Now, we minimize the r.h.s. in  $z \geq 0$ . The functions  $z$  and  $z^2$  are both increasing, the function  $(\Delta - 2q(1+z)/(\Delta n))_+^2$  in contrast is decreasing in  $z$ . Therefore, both inner minima are attained and by solving the corresponding quadratic equations (note that we have to take the solution with  $\Delta - 2q(1+z)/(\Delta n) \geq 0$ ) we get

$$\begin{aligned} &\mathbb{P}(T_1^n(\mathbf{Y}, m + \delta) \leq q) \\ &\leq 2 \exp\left(-\frac{1}{8}(n-1) \min\left[\left(\frac{\Delta - \frac{2q}{\Delta n}}{1 + \frac{2q}{\Delta n}}\right)_+^2, \frac{1 + 2\frac{2q}{\Delta n} \left(\Delta - \frac{2q}{\Delta n}\right)_+ - \sqrt{1 + 4\frac{2q}{\Delta n} \left(\Delta - \frac{2q}{\Delta n}\right)_+}}{2\left(\frac{2q}{\Delta n}\right)^2}\right]\right). \end{aligned}$$

Using the inequality  $\sqrt{1+4x} \leq 1+2x-2x^2+4x^3$  for all  $x > -1/4$  with  $x = 2q/(\Delta n)(\Delta -$

$2q/(\Delta n)_+$  we find

$$\begin{aligned} & \mathbb{P}(T_1^n(\mathbf{Y}\mathbf{Y}, m + \delta) \leq q) \\ & \leq 2 \exp \left( -\frac{1}{8} \frac{n-1}{n} n \min \left[ \left( \frac{\Delta - \frac{2q}{\Delta n}}{1 + \frac{2q}{\Delta n}} \right)_+^2, \left( \Delta - \frac{2q}{\Delta n} \right)_+ \left[ 1 - 2 \frac{2q}{\Delta n} \left( \Delta - \frac{2q}{\Delta n} \right)_+ \right] \right] \right). \end{aligned}$$

Next, we consider the two terms in the minimum separately. We assume w.l.o.g. that  $\sqrt{2q}/(\Delta\sqrt{n}) \leq 1$ , since otherwise the r.h.s. in (A.7) is two. For the first term we distinguish the cases  $2q > \Delta n$  and  $2q \leq \Delta n$ . If  $2q \leq \Delta n$  is satisfied, then

$$n \left( \frac{\Delta - \frac{2q}{\Delta n}}{1 + \frac{2q}{\Delta n}} \right)_+^2 \geq \frac{1}{4} n \left( \Delta - \frac{2q}{\Delta n} \right)_+^2 = \frac{1}{4} \left( \sqrt{n}\Delta - \frac{2q}{\Delta\sqrt{n}} \right)_+^2 \geq \frac{1}{4} \left( \sqrt{n}\Delta - \sqrt{2q} \right)_+^2.$$

For the other case, when  $2q > \Delta n$  holds, we obtain with  $q/n \leq 1/8$

$$\begin{aligned} n \left( \frac{\Delta - \frac{2q}{\Delta n}}{1 + \frac{2q}{\Delta n}} \right)_+^2 & \geq \frac{1}{4} n \left( \frac{\Delta - \frac{2q}{\Delta n}}{\frac{2q}{\Delta n}} \right)_+^2 = \frac{1}{4} n \left( \frac{n\Delta^2}{2q} - 1 \right)_+^2 \\ & = \frac{1}{4} \frac{n}{2q} \left( \frac{(\sqrt{n}\Delta)^2}{\sqrt{2q}} - \sqrt{2q} \right)_+^2 \geq \left( \sqrt{n}\Delta - \sqrt{2q} \right)_+^2. \end{aligned}$$

For the second term it follows with  $q/n \leq 1/8$  that

$$\begin{aligned} n \left( \Delta - \frac{2q}{\Delta n} \right)_+ \left[ 1 - 2 \frac{2q}{\Delta n} \left( \Delta - \frac{2q}{\Delta n} \right)_+ \right] & = \left( \sqrt{n}\Delta - \frac{2q}{\Delta\sqrt{n}} \right)_+ \left[ 1 - 4 \frac{q}{n} \left( 1 - \frac{2q}{\Delta^2 n} \right)_+ \right] \\ & \geq \left( \sqrt{n}\Delta - \sqrt{2q} \right)_+^2 \frac{1}{2}. \end{aligned}$$

This yields

$$\begin{aligned} \mathbb{P}(T_1^n(\mathbf{Y}, m + \delta) \leq q) & \leq 2 \exp \left( -\frac{1}{32} \frac{n-1}{n} \left( \sqrt{n}\Delta - \sqrt{2q} \right)_+^2 \right) \\ & \leq 2 \exp \left( -\frac{1}{48} \left( \sqrt{n}\Delta - \sqrt{2q} \right)_+^2 \right). \end{aligned}$$

□

## A.4. Proofs of Section 2.3

*Proof of Theorem 5.* The estimated number of change-points  $\hat{K}$  is by its definition in (2.18) equal to the minimal number of change-points of all feasible functions. Therefore, all functions with the true number of change-points (or less change-points) have to be

infeasible, if the number of change-points is overestimated. Hence, by (2.12)

$$\begin{aligned} \sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K} > K \right) &\leq \sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{P}_{(\mu, \sigma^2)} \left( \max_{\left[ \frac{i}{n}, \frac{j}{n} \right] \in \mathcal{D}(\mu)} \left[ T_i^j(\mathbf{Y}, \mu([i/n, j/n])) - q_{ij} \right] > 0 \right) \\ &\leq \mathbb{P}_{(0,1)} \left( \max_{\left[ \frac{i}{n}, \frac{j}{n} \right] \in \mathcal{D}} \left[ T_i^j(\mathbf{Y}, 0) - q_{ij} \right] > 0 \right) = \alpha, \end{aligned}$$

where the last inequality follows from  $\mathcal{D}(\mu) \subset \mathcal{D}$  and that the distribution of  $T_i^j(\mathbf{Y}, \mu([i/n, j/n]))$  does not depend on  $\mu(\cdot)$  and  $\sigma(\cdot)$ , as they are constant on intervals in  $\mathcal{D}(\mu)$ .  $\square$

*Proof of Theorem 6.* First of all, we show that it is enough to prove the result for  $\mu \equiv 0$  and  $\sigma^2 \equiv 1$  and hence  $K = 0$ . We have

$$\begin{aligned} &\sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K} > K + 2k \right) \\ &= \sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{P}_{(\mu, \sigma^2)} \left( \max_{\left[ \frac{i}{n}, \frac{j}{n} \right] \in \mathcal{D}(\tilde{\mu})} \left[ T_i^j(\mathbf{Y}, \tilde{\mu}([i/n, j/n])) - q_{ij} \right] > 0 \forall \tilde{\mu} \in \mathcal{M} \text{ s.t. } |\mathcal{J}(\tilde{\mu})| \leq K + 2k \right) \\ &\leq \sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{P}_{(\mu, \sigma^2)} \left( \max_{\left[ \frac{i}{n}, \frac{j}{n} \right] \in \mathcal{D}(\tilde{\mu})} \left[ T_i^j(\mathbf{Y}, \tilde{\mu}([i/n, j/n])) - q_{ij} \right] > 0 \right. \\ &\quad \left. \forall \tilde{\mu} \in \mathcal{M} \text{ s.t. } \mathcal{J}(\mu) \subset \mathcal{J}(\tilde{\mu}), |\mathcal{J}(\tilde{\mu})| \leq K + 2k \right) \\ &\leq \mathbb{P}_{(0,1)} \left( \max_{\left[ \frac{i}{n}, \frac{j}{n} \right] \in \mathcal{D}(\tilde{\mu})} \left[ T_i^j(\mathbf{Y}, \tilde{\mu}([i/n, j/n])) - q_{ij} \right] > 0 \forall \tilde{\mu} \in \mathcal{M} \text{ s.t. } |\mathcal{J}(\tilde{\mu})| \leq 2k \right) \\ &= \mathbb{P}_{(0,1)} \left( \hat{K} > 2k \right), \end{aligned}$$

where the last inequality follows from the same argument as in the proof of Theorem 5. Now, we define  $R_0 := 0$  and iteratively

$$R_{k+1} := \min \{ t > R_k : \exists s \text{ s.t. } R_k < s < t \text{ and } [s/n, t/n] \in \mathcal{D}, T_s^t(\mathbf{Y}, 0) > q_{\log_2(t-s+1)} \},$$

with the convention  $\min \emptyset = \infty$ . Then,

$$\mathbb{P}_{0,1}(R_{k+1} \leq n | R_1 = t) \leq \mathbb{P}_{0,1}(R_k \leq n) \text{ for all } t \in \{1, \dots, n\},$$

since for the l.h.s. the remaining  $k$  rejections  $R_2, \dots, R_{k+1}$  have to be in  $\{t+1, \dots, n\}$  instead of  $\{1, \dots, n\}$ . It follows

$$\begin{aligned} \mathbb{P}_{0,1}(\hat{K} > 2k) &\leq \mathbb{P}_{0,1}(R_{k+1} \leq n) = \sum_{t=1}^n \mathbb{P}_{0,1}(R_{k+1} \leq n | R_1 = t) \mathbb{P}_{0,1}(R_1 = t) \\ &\leq \mathbb{P}_{0,1}(R_1 \leq n) \mathbb{P}_{0,1}(R_k \leq n) \leq \dots \leq \mathbb{P}_{0,1}(R_1 \leq n)^{k+1} \leq \alpha^{k+1}, \end{aligned}$$

where the last inequality is given by Theorem 5. It follows

$$\begin{aligned} \sup_{(\mu, \sigma^2) \in \mathcal{S}} \mathbb{E}_{(\mu, \sigma^2)} \left[ (\hat{K} - K)_+ \right] &= \sup_{(\mu, \sigma^2) \in \mathcal{S}} \sum_{k=0}^{\infty} \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K} - K > k \right) \\ &\leq \sup_{(\mu, \sigma^2) \in \mathcal{S}} 2 \sum_{k=0}^{\infty} \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K} - K > 2k \right) \leq 2 \sum_{k=0}^{\infty} \alpha^{k+1} = \frac{2\alpha}{1-\alpha}. \end{aligned}$$

□

The following theorem is a sharper version of Theorem 7 that shows different probabilities for the detection of the change-points.

**Theorem 19** (Underestimation control II). *Let  $\lambda_j := \tau_{j+1} - \tau_j$  and  $k_{n,j} := \lfloor \log_2(n\lambda_j/4) \rfloor$ ,  $j = 0, \dots, K$ , as well as  $\delta_j := |m_j - m_{j-1}|$  and*

$$\eta_j := \left[ 1 - 3 \exp \left( -\frac{1}{48} \left( \sqrt{\frac{n\lambda_{j-1}\delta_j^2}{32\sigma_{j-1}^2}} - \sqrt{16 \log \left( \frac{8}{\lambda_j \alpha \beta_{k_{n,j-1}}} \right)} \right)_+^2 \right) \right]_+ \times \left[ 1 - 3 \exp \left( -\frac{1}{48} \left( \sqrt{\frac{n\lambda_j\delta_j^2}{32\sigma_j^2}} - \sqrt{16 \log \left( \frac{8}{\lambda_j \alpha \beta_{k_{n,j}}} \right)} \right)_+^2 \right) \right]_+,$$

$j = 1, \dots, K$ . Under the assumptions of Theorem 5 and if  $n\lambda_j \geq 32$  and

$$(n\lambda_j)^{-1} \log \left( \frac{8}{\lambda_j \alpha \beta_{k_{n,j}}} \right) \leq \frac{1}{512}$$

are satisfied for all  $j = 1, \dots, K$ , then

$$\mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K} < K \right) \leq 1 - \prod_{j=1}^K \eta_j \text{ and } \mathbb{E}_{(\mu, \sigma^2)} \left[ (K - \hat{K})_+ \right] \leq \sum_{j=1}^K (1 - \eta_j).$$

*Proof.* For each  $j = 1, \dots, K$  we consider the disjoint intervals  $I_j := [\tau_j - \lambda_{j-1}/2, \tau_j + \lambda_j/2)$  and split them into disjoint intervals  $I_j^+ \cup I_j^- = I_j$  such that  $\mu(t) = \mu^+ \forall t \in I_j^+$  and  $\mu(t) = \mu^- \forall t \in I_j^-$ , with  $\mu^+ := \max(m_{j-1}, m_j)$  and  $\mu^- := \min(m_{j-1}, m_j)$ . Without loss of generality we assume  $\mu^+ = m_{j-1}$  and  $\mu^- = m_j$  in the following. Then, there exists subintervals  $J_j^+ \subset I_j^+$  and  $J_j^- \subset I_j^-$  with  $J_j^+, J_j^- \in \mathcal{D}$  that have length  $\lambda_{j-1}^* := n^{-1}2^{\lfloor \log_2(n\lambda_{j-1}/4) \rfloor} = n^{-1}2^{k_{n,j-1}} \geq \lambda_{j-1}/8$ , since  $n|I_j^+| = n\lambda_{j-1}/2 \geq 3$ , and  $\lambda_j^* := n^{-1}2^{\lfloor \log_2(n\lambda_j/4) \rfloor} = n^{-1}2^{k_{n,j}} \geq \lambda_j/8$ , since  $n|I_j^-| = n\lambda_j/2 \geq 3$ , respectively. It

follows

$$\begin{aligned}
\mathbb{P}_{(\mu, \sigma^2)}(\hat{K} < K) &= 1 - \mathbb{P}_{(\mu, \sigma^2)}(\hat{K} \geq K) \\
&\leq 1 - \mathbb{P}_{(\mu, \sigma^2)}(\nexists \hat{\mu} \in C(\mathbf{Y}, \mathbf{q}), j \in \{1, \dots, K\} : \hat{\mu} \text{ is constant on } I_j) \\
&\leq 1 - \mathbb{P}_{(\mu, \sigma^2)}\left(\forall j \in \{1, \dots, K\} : \nexists \hat{m} \leq (m_{j-1} + m_j)/2 : T_{J_j^+}(\mathbf{Y}, \hat{m}) \leq q_{k_{n,j-1}} \text{ and} \right. \\
&\quad \left. \nexists \hat{m} \geq (m_{j-1} + m_j)/2 : T_{J_j^-}(\mathbf{Y}, \hat{m}) \leq q_{k_{n,j}}\right) \\
&\leq 1 - \prod_{j=1}^K \mathbb{P}_{(\mu, \sigma^2)}\left(\nexists \hat{m} \leq (m_{j-1} + m_j)/2 : T_{J_j^+}(\mathbf{Y}, \hat{m}) \leq q_{k_{n,j-1}} \text{ and} \right. \\
&\quad \left. \nexists \hat{m} \geq (m_{j-1} + m_j)/2 : T_{J_j^-}(\mathbf{Y}, \hat{m}) \leq q_{k_{n,j}}\right),
\end{aligned}$$

where we used in the last inequality that the events are independent, since all intervals are disjoint. We denote by  $\mathbf{Z} = (Z_1, \dots, Z_n)$  a vector of i.i.d. standard normally distributed random variables. It follows from once again from the independence due to disjoint intervals and from the Lemmas 7.1 in (Frick et al., 2014a), 3 and 18 that

$$\begin{aligned}
&\mathbb{P}_{(\mu, \sigma^2)}\left(\nexists \hat{m} \leq (m_{j-1} + m_j)/2 : T_{J_j^+}(\mathbf{Y}, \hat{m}) \leq q_{k_{n,j-1}} \text{ and} \right. \\
&\quad \left. \nexists \hat{m} \geq (m_{j-1} + m_j)/2 : T_{J_j^-}(\mathbf{Y}, \hat{m}) \leq q_{k_{n,j}}\right) \\
&\geq \left[1 - \mathbb{P}_{(\mu, \sigma^2)}\left(\exists \hat{m} \leq (m_{j-1} + m_j)/2 : T_{J_j^+}(\mathbf{Y}, \hat{m}) \leq q_{k_{n,j-1}}\right)\right] \times \\
&\quad \left[1 - \mathbb{P}_{(\mu, \sigma^2)}\left(\exists \hat{m} \geq (m_{j-1} + m_j)/2 : T_{J_j^-}(\mathbf{Y}, \hat{m}) \leq q_{k_{n,j}}\right)\right] \geq \eta_j,
\end{aligned}$$

since

$$\begin{aligned}
&\mathbb{P}_{(\mu, \sigma^2)}\left(\exists \hat{m} \leq (m_{j-1} + m_j)/2 : T_{J_j^+}(\mathbf{Y}, \hat{m}) \leq q_{k_{n,j-1}}\right) \\
&\leq \mathbb{P}_{(\mu, \sigma^2)}\left(\bar{Y}_{J_j^+} \leq (m_{j-1} + m_j)/2 \text{ or } T_{J_j^+}(\mathbf{Y}, (m_{j-1} + m_j)/2) \leq q_{k_{n,j-1}}\right) \\
&\leq \mathbb{P}_{(\mu, \sigma^2)}\left(\bar{Y}_{J_j^+} \leq (m_{j-1} + m_j)/2\right) + \mathbb{P}_{(\mu, \sigma^2)}\left(T_{J_j^+}(\mathbf{Y}, (m_{j-1} + m_j)/2) \leq q_{k_{n,j-1}}\right) \\
&\leq \mathbb{P}\left(\bar{Z}_{[0, \lambda_{j-1}^*]} \geq \frac{\delta_j}{2\sigma_{j-1}}\right) + \mathbb{P}\left(T_{[0, \lambda_{j-1}^*]}(\mathbf{Z}, \frac{\delta_j}{2\sigma_{j-1}}) \leq q_{k_{n,j-1}}\right) \\
&\leq \exp\left(-\frac{1}{64} \frac{n\lambda_{j-1}\delta_j^2}{\sigma_{j-1}^2}\right) + 2 \exp\left(-\frac{1}{48} \left(\sqrt{\frac{n\lambda_{j-1}\delta_j^2}{32\sigma_{j-1}^2}} - \sqrt{2q_{k_{n,j-1}}}\right)_+^2\right) \\
&\leq 3 \exp\left(-\frac{1}{48} \left(\sqrt{\frac{n\lambda_{j-1}\delta_j^2}{32\sigma_{j-1}^2}} - \sqrt{16 \log\left(\frac{8}{\lambda_{j-1}\alpha\beta_{k_{n,j-1}}}\right)}\right)_+^2\right)
\end{aligned}$$

and the second term by symmetry arguments. Moreover, it follows

$$\begin{aligned}
& \mathbb{E}_{(\mu, \sigma^2)} \left[ \left( K - \hat{K} \right)_+ \right] \\
& \leq \mathbb{E}_{(\mu, \sigma^2)} \left[ \sum_{j=1}^K \mathbb{1}_{\exists \hat{m} \leq (m_{j-1} + m_j)/2: T_{J_j^+}(\mathbf{Y}, \hat{m}) \leq q_{k_n, j-1} \text{ or } \exists \hat{m} \geq (m_{j-1} + m_j)/2: T_{J_j^-}(\mathbf{Y}, \hat{m}) \leq q_{k_n, j}} \right] \\
& \leq \sum_{j=1}^K (1 - \eta_j).
\end{aligned}$$

□

*Proof of Theorem 7.* The proof is analogue to the proof of Theorem 19, but with  $I_j = [\tau_j - \lambda/2, \tau_j + \lambda/2)$ . □

*Proof of Theorem 9.* We prove the theorem with the Borel-Cantelli lemma. It follows from Theorems 5 and 7 that

$$\begin{aligned}
& \sup_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K}_n \neq K \right) \\
& = \sup_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K}_n > K \right) + \sup_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K}_n < K \right) \\
& \leq \alpha_n + 1 - \left[ 1 - 3 \exp \left( -\frac{1}{48} \left( \sqrt{\frac{n\lambda\Delta^2}{32}} - \sqrt{16 \log \left( \frac{8}{\lambda\alpha_n\beta_{k_n, n}} \right)} \right)_+^2 \right) \right]_+^{2K} \\
& \leq \alpha_n + 6K \exp \left( -\frac{1}{48} \left( \sqrt{\frac{n\lambda\Delta^2}{32}} - \sqrt{16 \log \left( \frac{8}{\lambda\alpha_n\beta_{k_n, n}} \right)} \right)_+^2 \right),
\end{aligned}$$

since under the given assumptions the conditions of Theorem 7 are satisfied. The upper bounds for the error probabilities are summable if (2.21) is satisfied. □

**Lemma 20** (Confidence set). *Assume the setting and assumptions of Theorem 7 and let  $C(\mathbf{Y}, \mathbf{q})$  be as in (2.5) with significance level  $\alpha$  and weights  $\beta_1, \dots, \beta_{d_n}$ . Let  $\mathcal{S}_{\Delta, \lambda}$  be as in (2.19) with  $\Delta, \lambda > 0$  arbitrary, but fixed, and  $k_n := \lfloor \log_2(n\lambda/4) \rfloor$ . If  $n\lambda \geq 32$  and*

$$\frac{\log \left( \frac{8}{\lambda\alpha_n\beta_{k_n}} \right)}{n\lambda} \leq \frac{1}{512}$$

hold, then uniformly in  $\mathcal{S}_{\Delta, \lambda}$

$$\mathbb{P}_{(\mu, \sigma^2)} \left( \mu \in C(\mathbf{Y}, \mathbf{q}) \right) \geq 1 - \alpha - (1 - \eta^K),$$

with  $\eta$  like in Theorem 7.

*Proof.* It follows from the definition of  $C(\mathbf{Y}, \mathbf{q})$  in (2.5) as well as from Theorems 5 and

7 that

$$\begin{aligned}
& \inf_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} (\mu \in C(\mathbf{Y}, \mathbf{q})) \\
&= \inf_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} \left( \max_{[\frac{i}{n}, \frac{j}{n}] \in \mathcal{D}(\mu)} \left[ T_i^j(\mathbf{Y}, \mu([i/n, j/n])) - q_{ij} \right] \leq 0, \hat{K} = K \right) \\
&= \inf_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} \left( \max_{[\frac{i}{n}, \frac{j}{n}] \in \mathcal{D}(\mu)} \left[ T_i^j(\mathbf{Y}, \mu([i/n, j/n])) - q_{ij} \right] \leq 0, \hat{K} \geq K \right) \\
&\geq \inf_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} \left( \max_{[\frac{i}{n}, \frac{j}{n}] \in \mathcal{D}(\mu)} \left[ T_i^j(\mathbf{Y}, \mu([i/n, j/n])) - q_{ij} \right] \leq 0 \right) - \sup_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} (\hat{K} < K) \\
&\geq 1 - \alpha - (1 - \eta^K).
\end{aligned}$$

□

*Proof of Theorem 10.* The statement is a direct consequence of Lemma 20. □

**Lemma 21** (Change-point locations). *Assume the setting of Lemma 20. If  $c_n$  is a sequence with  $0 < c_n \leq \lambda/2$  and  $k_n := \lfloor \log_2(nc_n/2) \rfloor$  such that  $nc_n \geq 16$  and*

$$\frac{\log\left(\frac{4}{c_n \alpha \beta_{k_n}}\right)}{nc_n} \leq \frac{1}{256} \tag{A.8}$$

hold, then uniformly in  $\mathcal{S}_{\Delta, \lambda}$

$$\begin{aligned}
& \mathbb{P}_{(\mu, \sigma^2)} \left( \sup_{\hat{\mu} \in C(\mathbf{Y}, \mathbf{q}_n)} \max_{\tau \in \mathcal{J}(\mu)} \min_{\hat{\tau} \in \mathcal{J}(\hat{\mu})} |\hat{\tau} - \tau| > c_n \right) \\
&\leq 1 - \left[ 1 - 3 \exp \left( -\frac{1}{48} \left( \sqrt{\frac{nc_n \Delta^2}{16}} - \sqrt{16 \log \left( \frac{4}{c_n \alpha \beta_{k_n}} \right)} \right)_+^2 \right) \right]_+^{2K}.
\end{aligned}$$

*Proof.* Analogously to the proof of Theorem 19 we have

$$\begin{aligned}
& \sup_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} \left( \sup_{\hat{\mu} \in C(\mathbf{Y}, \mathbf{q}_n)} \max_{\tau \in \mathcal{J}(\mu)} \min_{\hat{\tau} \in \mathcal{J}(\hat{\mu})} |\hat{\tau} - \tau| > c_n \right) \\
&\leq \sup_{(\mu, \sigma^2) \in \mathcal{S}_{\Delta, \lambda}} \mathbb{P}_{(\mu, \sigma^2)} (\exists j \in \{1, \dots, K\} \text{ and } \hat{\mu} \in C(\mathbf{Y}, \mathbf{q}_n) : \hat{\mu} \text{ is constant on } [\tau_j - c_n, \tau_j + c_n]) \\
&\leq 1 - \left[ 1 - 3 \exp \left( -\frac{1}{48} \left( \sqrt{\frac{nc_n \Delta^2}{16}} - \sqrt{16 \log \left( \frac{4}{c_n \alpha \beta_{k_n}} \right)} \right)_+^2 \right) \right]_+^{2K}.
\end{aligned}$$

□

*Proof of Theorem 11.* For  $n$  large enough such that (2.23) guarantees the assumptions of

Lemma 21 it follows

$$\begin{aligned}
& \mathbb{P}_{(\mu, \sigma^2)} \left( \sup_{\hat{\mu} \in C(\mathbf{Y}, \mathbf{q}_n)} \max_{j=1, \dots, K} c_n^{-1} |\tau_j - \hat{\tau}_j| > 1 \right) \\
& \leq \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K} > K \text{ or } \exists \hat{\mu} \in C(\mathbf{Y}, \mathbf{q}_n), j \in \{1, \dots, K\} : \hat{\mu} \text{ is constant on } [\tau_j - c_n, \tau_j + c_n] \right) \\
& \leq \mathbb{P}_{(\mu, \sigma^2)} \left( \hat{K} > K \right) + \mathbb{P}_{(\mu, \sigma^2)} \left( \exists \hat{\mu} \in C(\mathbf{Y}, \mathbf{q}_n), j \in \{1, \dots, K\} : \hat{\mu} \text{ is constant on } [\tau_j - c_n, \tau_j + c_n] \right) \\
& \leq \alpha_n + \left( 1 - \left[ 1 - 3 \exp \left( -\frac{1}{48} \left( \sqrt{\frac{nc_n \Delta^2}{16}} - \sqrt{16 \log \left( \frac{4}{c_n \alpha_n \beta_{k_n, n}} \right)} \right)_+^2 \right) \right]_+^{2K} \right).
\end{aligned}$$

The assertion follows from  $\alpha_n \rightarrow 0$  and

$$\lim_{n \rightarrow \infty} \sqrt{\frac{nc_n \Delta^2}{16}} - \sqrt{16 \log \left( \frac{4}{c_n \alpha_n \beta_{k_n, n}} \right)} = \infty,$$

whereby latter one is direct consequence of (2.23).  $\square$

*Proof of Theorem 12.* We denote by  $J_n$  the longest subinterval  $J_n \subset I_n^C$  which is part of the dyadic partition. Such an interval exists (at least for  $n$  large enough) always, since  $|I_n| \rightarrow 0$ , and has at least length  $|I_n^C|/8$ . Moreover, let  $k_n := \log_2(n|I_n|)$  and  $l_n := \log_2(n|J_n|)$ . Then, the Lemmas 7.1 in (Frick et al., 2014a) and 18 yield for any  $\theta_n > 0$

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \mathbb{P}_{(\mu_n, \sigma_n^2)} \left( \hat{K}_n > 0 \right) \\
& = \lim_{n \rightarrow \infty} 1 - \mathbb{P}_{(\mu_n, \sigma_n^2)} \left( \hat{\mu} \text{ is constant} \right) \\
& \geq \lim_{n \rightarrow \infty} 1 - \mathbb{P}_{(\mu_n, \sigma_n^2)} \left( \exists \hat{m} \leq m_0 + \theta_n : T_{I_n}(\mathbf{Y}, \hat{m}) \leq q_{k_n} \text{ or } \exists \hat{m} \geq m_0 + \theta_n : T_{J_n}(\mathbf{Y}, \hat{m}) \leq q_{l_n} \right) \\
& \geq \lim_{n \rightarrow \infty} 1 - \mathbb{P}_{(\mu_n, \sigma_n^2)} \left( \exists \hat{m} \leq m_0 + \theta_n : T_{I_n}(\mathbf{Y}, \hat{m}) \leq q_{k_n} \right) \\
& \quad - \mathbb{P}_{(\mu_n, \sigma_n^2)} \left( \exists \hat{m} \geq m_0 + \theta_n : T_{J_n}(\mathbf{Y}, \hat{m}) \leq q_{l_n} \right) \\
& \geq \lim_{n \rightarrow \infty} 1 - \mathbb{P}_{(\mu_n, \sigma_n^2)} \left( \bar{Y}_{I_n} \leq m_0 + \theta_n \right) - \mathbb{P}_{(\mu_n, \sigma_n^2)} \left( T_{I_n}(\mathbf{Y}, m_0 + \theta_n) \leq q_{k_n} \right) \\
& \quad - \mathbb{P}_{(\mu_n, \sigma_n^2)} \left( \bar{Y}_{J_n} \geq m_0 + \theta_n \right) - \mathbb{P}_{(\mu_n, \sigma_n^2)} \left( T_{J_n}(\mathbf{Y}, m_0 + \theta_n) \leq q_{l_n} \right) \\
& \geq \lim_{n \rightarrow \infty} 1 - 2\mathbb{P}_{(\mu_n, \sigma_n^2)} \left( T_{I_n}(\mathbf{Y}, m_0 + \theta_n) \leq q_{k_n} \right) - 2\mathbb{P}_{(\mu_n, \sigma_n^2)} \left( T_{J_n}(\mathbf{Y}, m_0 + \theta_n) \leq q_{l_n} \right) \\
& \geq \lim_{n \rightarrow \infty} 1 - 4 \exp \left( -\frac{1}{48} (\Gamma_{I_n})_+^2 \right) - 4 \exp \left( -\frac{1}{48} (\Gamma_{J_n})_+^2 \right) = 1,
\end{aligned}$$

if

$$\Gamma_{I_n} := \sqrt{n|I_n|} \frac{\delta_n - \theta_n}{s_n} - \sqrt{2q_{k_n}} \rightarrow \infty \text{ and } \Gamma_{J_n} := \sqrt{n|J_n|} \theta_n - \sqrt{2q_{l_n}} \rightarrow \infty,$$

and if the conditions of Lemma 18 are satisfied. This is the case, since  $n|I_n| \rightarrow \infty$  and  $n|J_n| \rightarrow \infty$ , because of (2.26) and  $|I_n| \rightarrow 0$ , as well as  $q_{k_n}/(n|I_n|) \leq 1/8$  and  $q_{l_n}/(n|J_n|) \leq 1/8$  hold at least for  $n$  large enough: The first one is a direct consequence of Lemma 3

and (2.26)

$$\frac{q_{k_n}}{n|I_n|} \leq \frac{8 \log \left( \frac{1}{|I_n| \alpha_n \beta_{k_n, n}} \right)}{n|I_n|} \leq \frac{1}{8},$$

since then the assumptions of Lemma 3 are also fulfilled. The second inequality follows from Lemma 3, (2.26) and (2.28) as well as the fact that  $|I_n|/|J_n| \rightarrow 0$

$$\lim_{n \rightarrow \infty} \frac{q_{l_n}}{n|J_n|} \leq \lim_{n \rightarrow \infty} \frac{8 \log \left( \frac{1}{|J_n| \alpha_n \beta_{l_n, n}} \right)}{n|J_n|} \leq \lim_{n \rightarrow \infty} \frac{8 \log \left( \frac{1}{|J_n| \alpha_n \beta_{l_n, n}} \right)}{8 \log \left( \frac{1}{|I_n| \alpha_n \beta_{k_n, n}} \right)} \frac{|I_n|}{|J_n|} \frac{8 \log \left( \frac{1}{|I_n| \alpha_n \beta_{k_n, n}} \right)}{n|I_n|} \rightarrow 0,$$

since then the assumptions of Lemma 3 are also fulfilled.

We define now  $\theta_n = \sqrt{\gamma_n/n}$  via the equation

$$\sqrt{\frac{\gamma_n |I_n|}{s_n^2}} = c \epsilon_n \sqrt{\log \left( \frac{1}{|I_n|} \right)}$$

for  $0 < c < 1$ . Then, it follows from Lemma 3 and from  $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$  for  $x, y > 0$  together with the assumptions of the theorem that

$$\begin{aligned} \Gamma_{I_n} &= \sqrt{n|I_n|} \frac{\delta_n - \theta_n}{s_n} - \sqrt{2q_{k_n}} \\ &= \sqrt{n|I_n|} \frac{\delta_n^2}{s_n^2} - \sqrt{\gamma_n |I_n| / s_n^2} - \sqrt{2q_{k_n}} \\ &\geq \sqrt{n|I_n|} \Delta_n^2 - \sqrt{\gamma_n |I_n| / s_n^2} - \sqrt{16 \log \left( \frac{1}{|I_n| \alpha_n \beta_{k_n, n}} \right)} \\ &\geq (4 + \epsilon_n) \sqrt{\log \left( \frac{1}{|I_n|} \right)} - c \epsilon_n \sqrt{\log \left( \frac{1}{|I_n|} \right)} - 4 \sqrt{\log \left( \frac{1}{|I_n|} \right)} - 4 \sqrt{\log \left( \frac{1}{\alpha_n \beta_{k_n, n}} \right)} \\ &\geq (1 - c) \epsilon_n \sqrt{\log \left( \frac{1}{|I_n|} \right)} - 4 \sqrt{\log \left( \frac{1}{\alpha_n \beta_{k_n, n}} \right)} \rightarrow \infty, \end{aligned}$$

since the conditions of Lemma 3 are satisfied, as shown above.

Moreover, we have  $\Gamma_{J_n} := \sqrt{n|J_n|} \theta_n - \sqrt{2q_{l_n}} = \sqrt{|J_n| \gamma_n} - \sqrt{2q_{l_n}} \rightarrow \infty$  if

$$\sqrt{\frac{|J_n| \gamma_n}{2q_{l_n}}} \geq \sqrt{\frac{|J_n| \gamma_n}{16 \log \left( \frac{1}{|J_n| \alpha_n \beta_{l_n, n}} \right)}} \rightarrow \infty,$$

where we used Lemma 3 again. Finally, it follows from the assumptions of the theorem

that  $\liminf_{n \rightarrow \infty} |J_n| \geq \liminf_{n \rightarrow \infty} |I_n^C|/8 > 0$  and thus

$$\begin{aligned} \sqrt{\frac{|J_n| \gamma_n}{\log\left(\frac{1}{\alpha_n \beta_{l_n, n}}\right)}} &= \frac{\sqrt{|I_n| \gamma_n}}{s_n \sqrt{\log\left(\frac{1}{\alpha_n \beta_{k_n, n}}\right)}} \frac{s_n \sqrt{|J_n|} \sqrt{\log\left(\frac{1}{\alpha_n \beta_{k_n, n}}\right)}}{\sqrt{|I_n|} \sqrt{\log\left(\frac{1}{\alpha_n \beta_{l_n, n}}\right)}} \\ &\geq \frac{\alpha \epsilon_n \sqrt{\log\left(\frac{1}{|I_n|}\right)}}{\sqrt{\log\left(\frac{1}{\alpha_n \beta_{k_n, n}}\right)}} \frac{s_n \sqrt{|I_n^C|/8}}{\sqrt{|I_n|}} \frac{\sqrt{\log\left(\frac{1}{\alpha_n \beta_{k_n, n}}\right)}}{\sqrt{\log\left(\frac{1}{\alpha_n \beta_{\min, n}}\right)}} \rightarrow \infty. \end{aligned}$$

□

*Proof of Theorem 13.* It follows from Theorem 7 that

$$\mathbb{P}_{(\mu_n, \sigma_n^2)}\left(\hat{K}_n < K_n\right) \leq 1 - \left[1 - 3 \exp\left(-\frac{1}{48} (\Gamma_n)_+^2\right)\right]_+^{2K_n} \leq 6K_n \exp\left(-\frac{1}{48} (\Gamma_n)_+^2\right),$$

with

$$\Gamma_n := \sqrt{\frac{n\lambda_n \Delta_n^2}{32}} - \sqrt{16 \log\left(\frac{8}{\lambda_n \alpha_n \beta_{k_n, n}}\right)},$$

since the assumptions of Theorem 7 are satisfied by (2.30).

In case (1) it is enough to show  $\Gamma_n \rightarrow \infty$ , because  $K_n$  is bounded. Finally,  $\Gamma_n \rightarrow \infty$  follows from

$$\frac{n\lambda_n \Delta_n^2}{\log\left(\frac{8}{\lambda_n \alpha_n \beta_{k_n, n}}\right)} \rightarrow \infty.$$

In case (2) for bounded  $K_n$ ,  $\Gamma_n \rightarrow \infty$  follows from

$$\begin{aligned} \Gamma_n &= \sqrt{\frac{n\lambda_n \Delta_n^2}{32}} - \sqrt{16 \log\left(\frac{8}{\lambda_n \alpha_n \beta_{k_n, n}}\right)} \\ &\geq \left(\frac{\sqrt{512}}{\sqrt{32}} + \frac{\epsilon_n}{\sqrt{32}}\right) \sqrt{\log\left(\frac{1}{\lambda_n}\right)} - \sqrt{16} \sqrt{\log\left(\frac{1}{\lambda_n}\right)} - \sqrt{16} \sqrt{\log\left(\frac{8}{\alpha_n \beta_{k_n, n}}\right)} \\ &= \frac{1}{\sqrt{32}} \left(\epsilon_n \sqrt{\log\left(\frac{1}{\lambda_n}\right)} - \sqrt{512} \sqrt{\log\left(\frac{8}{\alpha_n \beta_{k_n, n}}\right)}\right) \rightarrow \infty. \end{aligned}$$

For unbounded  $K_n$  we have  $K_n \leq 1/\lambda_n$ . It follows

$$\begin{aligned} &K_n \exp\left(-\frac{1}{48} (\Gamma_n)_+^2\right) \\ &\leq \exp\left(\log\left(\frac{1}{\lambda_n}\right) - \frac{1}{48} \left(\frac{C}{\sqrt{32}} \sqrt{\log\left(\frac{1}{\lambda_n}\right)} + \frac{1}{\sqrt{32}} \epsilon_n \sqrt{\log\left(\frac{1}{\lambda_n}\right)} - \sqrt{16} \sqrt{\log\left(\frac{8}{\alpha_n \beta_{k_n, n}}\right)}\right)_+^2\right) \\ &\leq \exp\left(-\frac{1}{48} \left(\frac{1}{\sqrt{32}} \epsilon_n \sqrt{\log\left(\frac{1}{\lambda_n}\right)} - \sqrt{16} \sqrt{\log\left(\frac{8}{\alpha_n \beta_{k_n, n}}\right)}\right)_+^2\right) \rightarrow 0. \end{aligned}$$

□

## A.5. Proofs of Section 3

In this Section we collect few propositions used for the analysis of ion channel recordings in Section 3.

**Proposition 22.** *The covariance of two observations  $Y_i$  and  $Y_{i+j}$  in the heterogeneous ion channel model in Section 3.1.2 is given by*

$$\text{Cov}[Y_i, Y_{i+j}] = \begin{cases} \sum_{k=0}^K s_k^2 [\mathcal{A}(i/f_s - \tau_k, j/f_s) - \mathcal{A}(i/f_s - \tau_{k+1}, j/f_s)] & \text{for } |j| = 0, \dots, L, \\ 0 & \text{for } |j| > L, \end{cases}$$

for all  $i = 1, \dots, n$  and  $1 - i \leq j \leq n - i$ , with  $\mathcal{A}(t, l)$  as defined in (3.15).

*Proof.* Let  $\delta$  be Diracs delta function. By definition of the observations and the white noise process  $\eta$ , by the structure of the noise  $\sigma^2$  and by Fubini's theorem for the third equality it follows

$$\begin{aligned} & \text{Cov}[Y_i, Y_{i+j}] \\ &= \mathbb{E} \left[ \left( \int_{\mathbb{R}} F(t) \sigma(i/f_s - t) \eta(i/f_s - t) dt \right) \left( \int_{\mathbb{R}} F(s) \sigma((i+j)/f_s - s) \eta((i+j)/f_s - s) ds \right) \right] \\ &= \mathbb{E} \left[ \int_{\mathbb{R}} \int_{\mathbb{R}} F(t) \sigma(i/f_s - t) \eta(i/f_s - t) F(s) \sigma((i+j)/f_s - s) \eta((i+j)/f_s - s) dt ds \right] \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} F(t) F(s) \sigma(i/f_s - t) \sigma((i+j)/f_s - s) \mathbb{E}[\eta(i/f_s - t) \eta((i+j)/f_s - s)] dt ds \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} F(t) F(s) \sigma(i/f_s - t) \sigma((i+j)/f_s - s) \delta(t - s + j/f_s) dt ds \\ &= \int_{\mathbb{R}} F(t) F(t + j/f_s) \sigma(i/f_s - t)^2 dt \\ &= \sum_{k=0}^K s_k^2 \int_{i/f_s - \tau_{k+1}}^{i/f_s - \tau_k} F(t) F(t + j/f_s) dt \\ &= \sum_{k=0}^K s_k^2 [\mathcal{A}(i/f_s - \tau_k, j/f_s) - \mathcal{A}(i/f_s - \tau_{k+1}, j/f_s)]. \end{aligned}$$

The zero for  $|j| > L$  follows from truncating the autocorrelation. □

**Proposition 23.** *Let  $\tilde{O}_1, \tilde{O}_2, \dots$  be independent and exponentially distributed random variables, with rate  $\lambda_O$ . And let  $N$  be a random variable independent of  $\tilde{O}_1, \tilde{O}_2, \dots$  and geometrically distributed with probability  $p$ . Then,  $O := \sum_{i=1}^N \tilde{O}_i$  is exponentially distributed with rate  $p\lambda_O$ .*

*Proof.* It follows that the random variable  $O$  is with probability  $(1 - p)^{n-1}p$  equal to

$O(n) := \sum_{i=1}^n \tilde{O}_i$ . Hence,  $O$  has cumulative distribution function

$$\sum_{n=1}^{\infty} p(1-p)^{n-1} \mathbb{P}(O(n) \leq x).$$

Note that  $O(n)$  is the sum of  $n$  independent exponentially distributed random variables, and hence Gamma distributed with shape parameter  $n$  and rate parameter  $\lambda_O$ . Hence,  $O$  has density

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{p(1-p)^{n-1} \lambda_O^n x^{n-1} \exp(-\lambda_O x)}{\Gamma(n)} &= p \lambda_O \exp(-\lambda_O x) \sum_{n=1}^{\infty} \frac{(1-p)^{n-1} \lambda_O^{n-1} x^{n-1}}{(n-1)!} \\ &= p \lambda_O \exp(-p \lambda_O x). \end{aligned}$$

This is the density of an exponential distribution with rate  $p \lambda_O$ . □

**Proposition 24.** *The sequence  $\sqrt{n}(\hat{\lambda}_C - \lambda_C)$  is asymptotically normal distributed with mean zero and variance  $\sigma^2 := \text{Var}[C_1]/V_{\lambda_C}^2$ , with*

$$\text{Var}_{\lambda_C} [C_1] = \frac{1}{\lambda_C^2} + \frac{(\lambda_C a + 1)^2 / \lambda_C^2 \exp(-\lambda_C a) - (\lambda_C b + 1)^2 / \lambda_C^2 \exp(-\lambda_C b)}{\exp(-\lambda_C a) - \exp(-\lambda_C b)} - \mathbb{E}_{\lambda_C} [C_1]^2,$$

$$\mathbb{E}_{\lambda_C} [C_1] = \frac{1}{\lambda_C} + \frac{a \exp(-\lambda_C a) - b \exp(-\lambda_C b)}{\exp(-\lambda_C a) - \exp(-\lambda_C b)}$$

and

$$V_{\lambda_C} := \frac{(b-a)^2 \exp(-\lambda_C(a+b))}{(\exp(-\lambda_C a) - \exp(-\lambda_C b))^2} - \frac{1}{\lambda_C^2}.$$

Therefore,  $[\hat{\lambda}_C - z_{1-\alpha/2} \sigma / \sqrt{n}, \hat{\lambda}_C + z_{1-\alpha/2} \sigma / \sqrt{n}]$ , with  $z_a$  the  $a$ -quantil of the standard normal distribution, is an asymptotic  $1 - \alpha$  confidence interval for  $\lambda_C$ .

*Proof.* We proof the theorem with (van der Vaart, 2007, Section 5). We also use a similar notation. The expectation and variance follow by direct calculations using the density  $f_{C_1}(x)$  in (3.34). To prove consistency of  $\hat{\lambda}_C$  we use (van der Vaart, 2007, Lemma 5.10). We have parameter space  $\Lambda := [0, \infty)$ . Moreover, we define

$$\Psi_n(\lambda) := n \frac{a \exp(-\lambda a) - b \exp(-\lambda b)}{\exp(-\lambda a) - \exp(-\lambda b)} + \frac{n}{\lambda} - n \bar{C}$$

and

$$\Psi(\lambda) := n \frac{a \exp(-\lambda a) - b \exp(-\lambda b)}{\exp(-\lambda a) - \exp(-\lambda b)} + \frac{n}{\lambda} - n \frac{a \exp(-\lambda_C a) - b \exp(-\lambda_C b)}{\exp(-\lambda_C a) - \exp(-\lambda_C b)} - \frac{n}{\lambda_C}.$$

For each  $n$ ,  $\Psi_n(\lambda)$  is continuous and its only root is  $\hat{\lambda}_C$ , since  $\hat{\lambda}_C$  is defined as the minimizer of  $\Psi_n(\lambda)$ . Also  $\Psi(\lambda)$  is continuous and its only root is  $\lambda_C$ . Finally, it follows from the law of large numbers that  $\Psi_n(\lambda) \rightarrow \Psi(\lambda)$  in probability for every  $\lambda \in \Lambda$ . This was to show to prove consistency.

To show asymptotic normality we use (van der Vaart, 2007, Theorem 5.21). Let

$$\psi_\lambda(x) := \frac{a \exp(-\lambda a) - b \exp(-\lambda b)}{\exp(-\lambda a) - \exp(-\lambda b)} + \frac{1}{\lambda} - x.$$

This function is measurable for all  $\lambda$ .  $\psi_{\lambda_1}(x) - \psi_{\lambda_2}(x)$  is independent of  $x$  and continuous. Hence,  $\|\psi_{\lambda_1}(x) - \psi_{\lambda_2}(x)\| \leq \dot{\psi}(x) \|\lambda_1 - \lambda_2\|$  is satisfied for a constant  $\dot{\psi}(x)$ . Consequently,  $\dot{\psi}(x)$  is measurable and has existing second moment. Moreover,  $\mathcal{P}\|\psi_{\lambda_C}\|^2 = \mathcal{P}\psi_{\lambda_C}\psi_{\lambda_C}^T = \text{Var}[C_1] < \infty$  exists. Also  $V_{\lambda_C}$  exists and our definition in the proposition matches the definition in (van der Vaart, 2007, Theorem 5.21). By definition of  $\hat{\lambda}_C$ ,  $\mathbb{P}_n\psi_{\hat{\lambda}_C} = 0 = o_P(n^{-1/2})$ . This was to show and the assertion follows from (van der Vaart, 2007, Theorem 5.21).  $\square$

## A.6. Proofs of Section 4

*Proof of Theorem 14.* We prove the assertion with (van der Vaart, 2007, Theorem 5.9) which states three conditions for the convergence of a Z-estimator. Note that the convergence in probability can be replaced by almost sure convergence, if the assumptions hold almost surely. We define

$$\Psi(\theta) := |F(\theta) - (1 - \alpha)| + \sum_{k=2}^n \left| \frac{1 - F_1(\theta_1)}{\beta_1} - \frac{1 - F_k(\theta_k)}{\beta_k} \right|$$

and

$$\Psi_M(\theta) := |F_M(\theta) - (1 - \alpha)| + \sum_{k=2}^n \left| \frac{1 - F_{M,1}(\theta_1)}{\beta_1} - \frac{1 - F_{M,k}(\theta_k)}{\beta_k} \right|$$

as well as  $\Theta := [0, \infty)^n$ ,  $\theta_0 := \mathbf{q}$  and  $\hat{\theta}_M := \hat{\mathbf{q}}_M$ . Now, (4.8) and (4.9) yield

$$\Psi_M(\hat{\mathbf{q}}_M) \leq \frac{1}{M} \left( 1 + \frac{n-1}{\min\{\beta_1, \dots, \beta_n\}} \right) = o(1)$$

almost surely. In addition, Lemma 1 shows that the vector of critical values  $\mathbf{q}$  is unique. Moreover,  $\sup_{\theta \in [0, \infty)^n} \|F_M(\theta) - F(\theta)\|$  and  $\sup_{\theta_k \geq 0} \|F_{M,k}(\theta_k) - F_k(\theta_k)\|$  for all  $k \in \{1, \dots, n\}$  converge to zero almost surely. Thus, all assumptions of (van der Vaart, 2007, Theorem 5.9) are satisfied and the assertion follows.  $\square$

## Bibliography

- Acred, P., Brown, D. M., Turner, D. H., and Wilson, M. J. (1962). Pharmacology and chemotherapy of ampicillin - a new broad-spectrum penicillin. *Br. J. Pharmacol.*, 18(2):356–369.
- Alberts, B., Johnson, A., Walter, P., Lewis, J., Raff, M., and Roberts, K. (2007). *Molecular Biology of the Cell*. Taylor & Francis, 5. edition.
- Andersen, O. S., Koeppe, R. E., and B. Roux (2005). Gramicidin channels. *IEEE Trans. Nanobioscience*, 4(1):10–20.
- Andersen, O. S., Koeppe, R. E., and Roux, B. (2007). Gramicidin channels: Versatile tools. In *Biological Membrane Ion Channels*, pages 33–80. Springer.
- Arlot, S. and Celisse, A. (2011). Segmentation of the mean of heteroscedastic data via cross-validation. *Statist. Comput.*, 21(4):613–632.
- Arlot, S., Celisse, A., and Harchaoui, Z. (2012). A kernel multiple change-point algorithm via model selection. arXiv:1202.3878.
- Armstrong, K. M. and Cukierman, S. (2002). On the origin of closing flickers in gramicidin channels: a new hypothesis. *Biophys. J.*, 82(3):1329–1337.
- Babu, G. J. and Rao, C. R. (2004). Goodness-of-fit tests when parameters are estimated. *Sankhya*, 66(1):63–74.
- Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.
- Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *J. Appl. Econometrics*, 18:1–22.
- Bakirov, N. K. and Szekely, G. J. (2006). Student’s t-test for Gaussian scale mixtures. *J. Math. Sci.*, 139(3):6497–6505.
- Ball, F. G. and Rice, J. A. (1992). Stochastic models for ion channels: introduction and bibliography. *Math. Biosci.*, 112(2):189–206.
- Bartsch, A., Kattner, C., Pein, F., Llabrés, S., Schön, M., Diehn, M., Zachariae, U., Tanabe, M., Munk, A., and Steinem, C. (2017). Time-resolved interaction of single ampicillin molecules with *Neisseria meningitidis* PorB. *submitted*.
- Basseville, M. and Benveniste, A. (1983). Design and comparative study of some sequential jump detection algorithms for digital signals. *IEEE Trans. Acoust.*, 31(3):521–535.
- Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.*, 3(3):203–268.
- Boysen, L., Bruns, S., and Munk, A. (2009a). Jump estimation in inverse regression. *Electron. J. Stat.*, 3:1322–1359.
- Boysen, L., Kempe, A., Liebscher, V., Munk, A., and Wittich, O. (2009b). Consistencies

- and rates of convergence of jump-penalized least squares estimators. *Ann. Statist.*, 37(1):157–183.
- Braun, J. V., Braun, R. K., and Müller, H.-G. (2000). Multiple changepoint fitting via quasilielihood, with application to DNA sequence segmentation. *Biometrika*, 87(2):301–314.
- Brockwell, P. and Davis, R. (2006). *Time Series: Theory and Methods*. Springer, New York.
- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms: 2. the new algorithm. *IMA J. Appl. Math.*, 6(3):222–231.
- Chan, H. P. and Walther, G. (2013). Detection with the scan and the average likelihood ratio. *Statist. Sinica*, 23:409–428.
- Chung, S.-H., Anderson, O. S., and Krishnamurthy, V. (2007). *Biological membrane ion channels: dynamics, structure, and applications*. Springer.
- Colquhoun, D. (1987). *Practical analysis of single channel records*. Microelectrode techniques. The Plymouth workshop handbook. Cambridge: Company of Biologists.
- Colquhoun, D., Hawkes, A. G., and Srodzinski, K. (1996). Joint distributions of apparent open and shut times of single-ion channels and maximum likelihood fitting of mechanisms. *Philos. Trans. A Math. Phys. Eng. Sci.*, 354(1718):2555–2590.
- Colquhoun, D. and Sigworth, F. J. (1995). Fitting and statistical analysis of single-channel records. In *Single-channel recording*, pages 483–587. Springer.
- Cosgrove, S. E. and Carmeli, Y. (2003). The impact of antimicrobial resistance on health and economic outcomes. *Clin. Infect. Dis.*, 36(11):1433–1437.
- Csörgö, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley.
- Danelon, C., Nestorovich, E. M., Winterhalter, M., Ceccarelli, M., and Bezrukov, S. M. (2006). Interaction of zwitterionic penicillins with the OmpF channel facilitates their translocation. *Biophys. J.*, 90(5):1617–1627.
- de Gunst, M. C. M., Künsch, H. R., and Schouten, J. G. (2001). Statistical analysis of ion channel data using hidden Markov models with correlated state-dependent noise and filtering. *J. Am. Stat. Assoc.*, 96(455):805–815.
- Delcour, A. H. (2003). Solute uptake through general porins. *Front. Biosci.*, 8:d1055–71.
- Delcour, A. H. (2009). Outer membrane permeability and antibiotic resistance. *BBA Proteins Proteom.*, 1794(5):808–816.
- Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.
- Du, C., Kao, C.-L. M., and Kou, S. C. (2016). Stepwise signal extraction via marginal likelihood. *J. Amer. Statist. Assoc.*, 111(513):314–330.
- Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.*, 29(1):124–152.
- Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.*, 36(4):1758–1785.

- Dykstra, R. L. (1983). An algorithm for restricted least squares regression. *J. Am. Stat. Assoc.*, 78(384):837–842.
- Eichinger, B. and Kirch, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli*, 24(1):526–564.
- Engl, H.-W., Hanke, M., and Neubauer, A. (2000). *Regularization of inverse problems*. Kluwer Academic Publishers Group.
- Enikeeva, F., Munk, A., and Werner, F. (2016). Bump detection in a heterogeneous Gaussian model. *Bernoulli*, to appear, *arXiv:1504.07390*.
- Epstein, M., Calderhead, B., Girolami, M. A., and Sivilotti, L. G. (2016). Bayesian statistical inference in ion-channel models with exact missed event correction. *Biophys. J.*, 111(2):333–348.
- Fang, X., Li, J., and Siegmund, D. (2016). Segmentation and estimation of change-point models. *arXiv:1608.03032*.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *Comput. J.*, 13(3):317–322.
- Frick, K., Munk, A., and Sieling, H. (2014a). Multiscale change point inference (with discussion and rejoinder by the authors). *J. Roy. Statist. Soc. Ser. B*, 76(3):495–580.
- Frick, S., Hohage, T., and Munk, A. (2014b). Asymptotic laws for change point estimation in inverse regression. *Stat. Sin.*, 24(2):555–575.
- Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, 42(6):2243–2281.
- Fujikoshi, Y. and Mukaihata, S. (1993). Approximations for the quantiles of Student’s  $t$  and  $F$  distributions and their error bounds. *Hiroshima Math. J.*, 23(3):557–564.
- Fuliński, A., Grzywna, Z., Mellor, I., Siwy, Z., and Usherwood, P. N. R. (1998). Non-Markovian character of ionic current fluctuations in membrane channels. *Phys. Rev. E*, 58(1):919–924.
- Futschik, A., Hotz, T., Munk, A., and Sieling, H. (2014). Multiscale DNA partitioning: Statistical evidence for segments. *Bioinformatics*, 30(16):2255–2262.
- Giné, E. and Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models*, volume 40. Cambridge University Press.
- Gnanasambandam, R., Nielsen, M. S., Nicolai, C., Sachs, F., Hofgaard, J. P., and Dreyer, J. K. (2017). Unsupervised idealization of ion channel recordings by minimum description length: Application to human PIEZO1-channels. *Front. Neuroinform.*, 11.
- Goldenshluger, A., Tsybakov, A., and Zeevi, A. (2006). Optimal change-point estimation from indirect observations. *Ann. Statist.*, 34(1):350–372.
- Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Math. Comput.*, 24(109):23–26.
- Goychuk, I., Hänggi, P., Vega, J. L., and Miret-Artés, S. (2005). Non-Markovian stochastic resonance: Three-state model of ion channel gating. *Phys. Rev. E*, 71(6):061906.
- Guillaume, P., Pintelon, R., and Schoukens, J. (1990). On the use of signals with a

- constant signal-to-noise ration in the frequency domain. *IEEE Trans. Instrum. Meas.*, 39(6):835–842.
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B., and Sigworth, F. J. (1981). Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.*, 391(2):85–100.
- Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.*, 105:1480–1493.
- Hawkes, A. G., Jalali, A., and Colquhoun, D. (1990). The distributions of the apparent open times and shut times in a single channel record when brief events cannot be detected. *Philos. Trans. A Math. Phys. Eng. Sci.*, 332(1627):511–538.
- Haynes, K., Fearnhead, P., and Eckley, I. A. (2017). A computationally efficient nonparametric approach for changepoint detection. *Stat. Comput.*, 27(5):1293–1305.
- Heinemann, S. H. and Sigworth, F. J. (1988). Open channel noise. IV. Estimation of rapid kinetics of formamide block in gramicidin A channels. *Biophys. J.*, 54(4):757–764.
- Heinemann, S. H. and Sigworth, F. J. (1990). Open channel noise. V. Fluctuating barriers to ion entry in gramicidin A channels. *Biophys. J.*, 57(3):499–514.
- Heinemann, S. H. and Sigworth, F. J. (1991). Open channel noise. VI. Analysis of amplitude histograms to determine rapid kinetic parameters. *Biophys. J.*, 60(3):577–587.
- Hille, B. (2001). *Ion channels of excitable membranes*. Sinauer Associates, 3rd edition.
- Hotz, T., Schütte, O. M., Sieling, H., Polupanow, T., Diederichsen, U., Steinem, C., and Munk, A. (2013). Idealizing ion channel recordings by a jump segmentation multiresolution filter. *IEEE Trans. Nanobioscience*, 12(4):376–386.
- Hotz, T. and Sieling, H. (2015). *stepR: Fitting Step-Functions*. R package version 1.0-1.
- Ibragimov, R. and Müller, U. K. (2010). t-test based correlation and heterogeneity robust inference. *J. Bus. Econom. Statist.*, 28(4):453–468.
- James, C. E., Mahendran, K. R., Molitor, A., Bolla, J.-M., Bessonov, A. N., Winterhalter, M., and Pagès, J.-M. (2009). How  $\beta$ -lactam antibiotics enter bacteria: a dialogue with the porins. *PLoS One*, 4(5):e5453.
- Jeng, X. J., Cai, T. T., and Li, H. (2010). Optimal sparse segment identification with application in copy number variation analysis. *J. Amer. Statist. Assoc.*, 105(491):1156–1166.
- Johannsson, A., Keightley, C. A., Smith, G. A., Richards, C. D., Hesketh, T. R., and Metcalfe, J. C. (1981). The effect of bilayer thickness and n-alkanes on the activity of the  $(\text{Ca}^{2+} + \text{Mg}^{2+})$ -dependent ATPase of sarcoplasmic reticulum. *J. Biol. Chem.*, 256(4):1643–1650.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, volume 1. Wiley, 2. edition.
- Kass, R. S. (2005). The channelopathies: novel insights into molecular and genetic mechanisms of human disease. *J. Clin. Invest.*, 115(8):1986–1989.
- Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.*, 107(500):1590–1598.

- Kim, J.-B. (2014). Channelopathies. *Korean J. Pediatr.*, 57(1):1–18.
- Kolaczyk, E. D. and Nowak, R. D. (2005). Multiscale generalised linear models for non-parametric function estimation. *Biometrika*, 92(1):119–133.
- Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics, 3. edition.
- Li, H., Guo, Q., and Munk, A. (2017). Multiscale change-point segmentation: Beyond step functions. *arXiv:1708.03942*.
- Li, H., Munk, A., and Sieling, H. (2016). FDR-control in multiscale change-point segmentation. *Electron. J. Stat.*, 10(1):918–959.
- Li, H. and Sieling, H. (2015). *FDRSeg: FDR-Control in Multiscale Change-point Segmentation*. R package version 1.0-1.
- Lilliefors, H. W. (1969). On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *J. Am. Stat. Assoc.*, 64(325):387–389.
- Mahendran, K. R., Hajjar, E., Mach, T., Lovelle, M., Kumar, A., Sousa, I., Spiga, E., Weingart, H., Gameiro, P., and Winterhalter, M. (2010). Molecular basis of enrofloxacin translocation through OmpF, an outer membrane channel of *Escherichia coli*-when binding does not imply translocation. *J. Phys. Chem. B*, 114(15):5170–5179.
- Maidstone, R., Hocking, T., Rigai, G., and Fearnhead, P. (2017). On optimal multiple changepoint algorithms for large data. *Stat. Comput.*, 27(2):519–533.
- Maidstone, R. and Pickering, B. (2014). Discussion of "Multiscale change-point inference" by K. Frick, A. Munk and H. Sieling. *J. Roy. Statist. Soc. Ser. B*, 76(3):560–561.
- Matteson, D. S. and James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *J. Amer. Statist. Assoc.*, 109(505):334–345.
- McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.
- Mercik, S. and Weron, K. (2001). Stochastic origins of the long-range correlations of ionic current fluctuations in membrane channels. *Phys. Rev. E*, 63(5):051910.
- Molecular Devices (2008). *The Axon Guide: A Guide to Electrophysiology and Biophysics Laboratory Techniques*, 3rd edition.
- Muggeo, V. M. R. and Adelfio, G. (2011). Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics*, 27(2):161–166.
- Neher, E. and Eibl, H. (1977). The influence of phospholipid polar groups on gramicidin channels. *BBA Biomembranes*, 464(1):37–44.
- Neher, E. and Sakmann, B. (1976). Single-channel currents recorded from membrane of denervated frog muscle fibers. *Nature*, 260(5554):799–802.
- Nestorovich, E. M., Danelon, C., Winterhalter, M., and Bezrukov, S. M. (2002). Designed to penetrate: time-resolved interaction of single antibiotic molecules with bacterial pores. *Proc. Natl. Acad. Sci. U.S.A.*, 99(15):9789–9794.
- Nicolai, C. and Sachs, F. (2013). Solving ion channel kinetics with the QuB software. *Biophys. Rev. Lett.*, 8(03n04):191–211.
- Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited.

- Microbiol. Mol. Biol. R.*, 67(4):593–656.
- O’Connell, A. M., Koeppe, R. E., and Andersen, O. S. (1990). Kinetics of gramicidin channel formation in lipid bilayers: transmembrane monomer association. *Science*, 250(4985):1256–1259.
- Olesky, M., Hobbs, M., and Nicholas, R. A. (2002). Identification and analysis of amino acid mutations in porin IB that mediate intermediate-level resistance to penicillin and tetracycline in *Neisseria gonorrhoeae*. *Antimicrob. Agents. Ch.*, 46(9):2811–2820.
- Oppenheim, B. A. (1997). Antibiotic resistance in *Neisseria meningitidis*. *Clin. Infect. Dis.*, 24(Supplement\_1):S98–S101.
- Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006). How many drug targets are there? *Nat. Rev. Drug. Discov.*, 5(12):993–996.
- Pein, F. (2016). *HSMUCE: Implementation of H-SMUCE*. R package version 0.0.0.9001.
- Pein, F., Hotz, T., Sieling, H., and Aspelmeier, T. (2017a). *stepR: Multiscale change-point inference*. R package version 2.0-1.
- Pein, F., Hotz, T., Tecuapetla-Gómez, I., and Aspelmeier, T. (2017b). *clampSeg: Idealisation of Patch Clamp Recordings*. R package version 1.0-1.
- Pein, F. and Munk, A. (2017). *R package stepR: Multiscale change-point inference*. Vignette of R package stepR.
- Pein, F., Sieling, H., and Munk, A. (2017c). Heterogeneous change point inference. *J. Roy. Statist. Soc. Ser. B*, 79(4):1207–1227.
- Pein, F., Tecuapetla-Gómez, I., Schütte, O. M., Steinem, C., and Munk, A. (2017d). Fully-automatic multiresolution idealization for filtered ion channel recordings: flickering event detection. *arXiv:1706.03671*, submitted.
- Proakis, J. G. and Manolakis, D. G. (1996). *Digital Signal Processing: Principles, Algorithms, and Applications*. Prentice Hall.
- Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., McNamara, J. O., and White, L. E. (2008). *Neuroscience*. Sinauer Associates, 4th edition.
- Qin, F., Auerbach, A., and Sachs, F. (1996). Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.*, 70(1):264–280.
- Qin, F., Auerbach, A., and Sachs, F. (2000). Hidden Markov modeling for single channel kinetics with filtering and correlated noise. *Biophys. J.*, 79(4):1928–1944.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raj Singh, P., Ceccarelli, M., Lovelle, M., Winterhalter, M., and Mahendran, K. R. (2012). Antibiotic permeation across the OmpF channel: modulation of the affinity site in the presence of magnesium. *J. Phys. Chem. B*, 116(15):4433–4438.
- Rice, L. B. (2007). Emerging issues in the management of infections caused by multidrug-resistant gram-negative bacteria. *Clev. Clin. J. Med.*, 74:S12–20.
- Rigollet, P. and Tsybakov, A. (2012). Sparse estimation by exponential weighting. *Statist. Sci.*, 27(4):558–575.

- Ring, A. (1986). Brief closures of gramicidin A channels in lipid bilayer membranes. *BBA Biomembranes*, 856(3):646–653.
- Robertson, T. and Cryer, J. D. (1974). An iterative procedure for estimating the mode. *J. Am. Stat. Assoc.*, 69(348):1012–1016.
- Rufibach, K. and Walther, G. (2010). The block criterion for multiscale inference about a density, with applications to other multiscale problems. *J. Comput. Graph. Statist.*, 19(1):175–190.
- Sakmann, B. and Neher, E. (1995). *Single-Channel Recording*. Springer, 2nd. edition.
- Schervish, M. J. (1996). P values: what they are and what they are not. *Am. Stat.*, 50(3):203–206.
- Schroeder, I. (2015). How to resolve microsecond current fluctuations in single ion channels: The power of beta distributions. *Channels*, 9(5):262–280.
- Schulz, G. E. (2002). The structure of bacterial outer membrane proteins. *BBA Biomembranes*, 1565(2):308–317.
- Scott, A. J. and Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 30(3):507–512.
- Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Math. Comput.*, 24(111):647–656.
- Shelley, C., Niu, X., Geng, Y., and Magleby, K. L. (2010). Coupling and cooperativity in voltage activation of a limited-state BK channel gating in saturating  $\text{Ca}^{2+}$ . *J. Gen. Physiol.*, 135(5):461–480.
- Siekmann, I., Wagner, L. E., Yule, D., Fox, C., Bryant, D., Crampin, E. J., and Sneyd, J. (2011). MCMC estimation of Markov models for ion channels. *Biophys. J.*, 100(8):1919–1929.
- Sieling, H. (2013). *Statistical Multiscale Segmentation: Inference, Algorithms and Applications*. PhD thesis, Georg-August University Göttingen.
- Sigworth, F. J. (1985). Open channel noise. I. Noise in acetylcholine receptor currents suggests conformational fluctuations. *Biophys. J.*, 47(5):709–720.
- Sigworth, F. J. (1986). Open channel noise. II. A test for coupling between current fluctuations and conformational transitions in the acetylcholine receptor. *Biophys. J.*, 49(5):1041–1046.
- Sigworth, F. J., Urry, D. W., and Prasad, K. U. (1987). Open channel noise. III. High-resolution recordings show rapid current fluctuations in gramicidin A and four chemical analogues. *Biophys. J.*, 52(6):1055–1064.
- Song, J., Minetti, C. A. S. A., Blake, M. S., and Colombini, M. (1998). Successful recovery of the normal electrophysiological properties of PorB (class 3) porin from *Neisseria meningitidis* after expression in *Escherichia coli* and renaturation. *BBA Biomembranes*, 1370(2):289–298.
- Spokoiny, V. and Zhilova, M. (2013). Sharp deviation bounds for quadratic forms. *Math. Methods Statist.*, 22(2):100–113.
- Suginta, W., Winterhalter, M., and Smith, M. F. (2016). Correlated trapping of

- sugar molecules by the trimeric protein channel chitoporin. *BBA Biomembranes*, 1858(12):3032–3040.
- Tanabe, M., Nimigean, C. M., and Iverson, T. M. (2010). Structural basis for solute transport, nucleotide regulation, and immunological recognition of *Neisseria meningitidis* PorB. *Proc. Natl. Acad. Sci. U.S.A.*, 107(15):6811–6816.
- Tecuapetla-Gómez, I. and Munk, A. (2017). Autocovariance estimation in regression with a discontinuous signal and  $m$ -dependent errors: A difference-based approach. *Scand. J. Stat.*, 44(2):346–368.
- van der Vaart, A. W. (2007). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, 8th printing edition.
- VanDongen, A. M. (1996). A new algorithm for idealizing single ion channel data containing multiple unknown conductance levels. *Biophys. J.*, 70(3):1303–1315.
- Venkataramanan, L., Kuc, R., and Sigworth, F. J. (2000). Identification of hidden Markov models for ion channel currents. III. Bandlimited, sampled data. *IEEE Trans. Signal Process.*, 48(2):376–385.
- Venkatraman, E. S. and Olshen, A. B. (2007). A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*, 23(6):657–663.
- Virji, M. (2009). Pathogenic neisseriae: surface modulation, pathogenesis and infection control. *Nat. Rev. Microbiol.*, 7(4):274.
- Vogel, C. R. (2002). *Computational Methods for Inverse Problems*. Siam.
- Vostrikova, L. J. (1981). Detecting "disorder" in multidimensional random processes. *Soviet Math. Dokl.*, 24(1):55–59.
- Walther, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.*, 38(2):1010–1033.
- Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*. CRC Press.
- World Health Organization (2014). WHO's first global report on antibiotic resistance reveals serious, worldwide threat to public health. In *Antimicrobial resistance—global surveillance report. Virtual Press Conference*, volume 30.
- Yao, Y. (1988). Estimating the number of change-points via Schwarz criterion. *Statist. Probab. Lett.*, 6:181–189.
- Yau, C. Y. and Zhao, Z. (2016). Inference for multiple change points in time series via likelihood ratio scan statistics. *J. Roy. Statist. Soc. Ser. B*, 78(4):895–916.
- Yellen, G. (1984). Ionic permeation and blockade in  $Ca^{2+}$ -activated  $K^{+}$  channels of bovine chromaffin cells. *J. Gen. Physiol.*, 84(2):157–186.
- Zhang, N. R. and Siegmund, D. O. (2012). Model selection for high-dimensional, multi-sequence change-point problems. *Statist. Sinica*, 22:1507–1538.
- Zhou, Z. (2014). Discussion of "Multiscale change-point inference" by K. Frick, A. Munk and H. Sieling. *J. Roy. Statist. Soc. Ser. B*, 76(3):566–567.
- Zou, C., Yin, G., Feng, L., and Wang, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems. *Ann. Statist.*, 42(3):970–1002.

# Curriculum Vitae

## Florian Pein

Auf dem Lohberge 18, 37085 Göttingen; 0551 39 172134; fpein@uni-goettingen.de

## Education

02/03/1989:	born in Frankenberg (Eder)
05/2008:	General Qualification for University Entrance, Edertalschule Frankenberg
10/2008-10/2011:	B.Sc. in Mathematics, Georg-August-University of Göttingen
10/2011-10/2013:	M.Sc. in Mathematics, Georg-August-University of Göttingen
since 10/2013:	Research Assistant (funded by DFG CRC 803) and Ph.D. Student in Mathematics, Georg-August-University of Göttingen

## Publications

Pein, F., Sieling, H., and Munk, A. (**2017**). Heterogeneous change point inference. *Journal of the Royal Statistical Society, Series B*, 79(4):1207-1227.

Pein, F., Tecuapetla-Gómez, I., Schütte, O. M., Steinem, C., and Munk, A. (**2017**). Fully-automatic multiresolution idealization for filtered ion channel recordings: flickering event detection. arXiv:1706.03671, submitted.

## Software

R package *stepR*: Multiscale change-point inference, <https://CRAN.R-project.org/package=stepR>, version 2.0-1, **2017**.

R package *clampSeg*: Idealisation of Patch Clamp Recordings, <https://CRAN.R-project.org/package=clampSeg>, version 1.0-1, **2017**.