

Real-Time detection of state transitions in stochastic signals from biological systems

Citation for published version (APA):

Bergkamp, M. H., van IJzendoorn, L. J., & Prins, M. W. J. (2021). Real-Time detection of state transitions in stochastic signals from biological systems. *ACS Omega*, 6(27), 17726-17733.
<https://doi.org/10.1021/acsomega.1c02498>

Document license:
CC BY

DOI:
[10.1021/acsomega.1c02498](https://doi.org/10.1021/acsomega.1c02498)

Document status and date:
Published: 13/07/2021

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Real-Time Detection of State Transitions in Stochastic Signals from Biological Systems

Max H. Bergkamp, Leo J. van IJzendoorn, and Menno W.J. Prins*



Cite This: *ACS Omega* 2021, 6, 17726–17733



Read Online

ACCESS |



Metrics & More



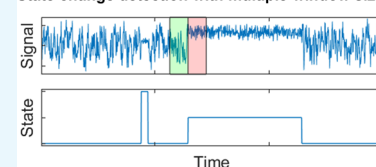
Article Recommendations



Supporting Information

ABSTRACT: Robust analysis of signals from stochastic biomolecular processes is critical for understanding the dynamics of biological systems. Measured signals typically show multiple states with heterogeneities and a wide range of state lifetimes. Here, we present an algorithm for robust detection of state transitions in experimental time traces where the properties of the underlying states are *a priori* unknown. The method implements a maximum-likelihood approach to fit models in neighboring windows of data points. Multiple windows are combined to achieve a high sensitivity for state transitions with a wide range of lifetimes. The proposed maximum-likelihood multiple-windows change point detection (MM-CPD) algorithm is computationally extremely efficient and enables real-time signal analysis. By analyzing both simulated and experimental data, we demonstrate that the algorithm provides accurate change point detection in time traces with multiple heterogeneous states that are *a priori* unknown. A high sensitivity for a wide range of state lifetimes is achieved.

State change detection with multiple window sizes



1. INTRODUCTION

Time-dependencies observed in biological systems are at the most basic level controlled by the dynamics of biomolecular processes. Examples are the binding and unbinding of ligands, conformational switching of proteins, addition and removal of chemical groups, modulation of enzyme activity, etc. At the level of individual molecules, transitions between states are stochastic and described by transition probabilities. The stochastic nature of molecules and their interactions transpires into stochastic properties at higher biological levels. For example, intracellular transport by the molecular motor protein kinesin is dependent on ATP binding and conformational changes of kinesin.¹ Another example is bacterial chemotaxis where run-and-tumble motion leads to a biased random walk in a preferred chemical gradient direction. The bacterial motion, controlled by flagellar motors, depends on single-molecule processes including stochastic ligand binding to receptors.²

Dedicated experimental techniques have been developed to study the dynamics of stochastic biological systems, such as single-particle tracking,³ optical tweezers,⁴ magnetic tweezers,⁴ fluorescence resonance energy transfer,⁵ super-resolution microscopy,⁶ and nanopores.⁷ A critical aspect of the experiments is the data analysis. Here, discrete change points and corresponding state transitions need to be accurately and reliably detected in noisy time-dependent data. Furthermore, the experiments typically exhibit multiple heterogeneous states of which the properties are not *a priori* known, caused by the complexity of biological systems and time-dependencies of the constituent biomolecules. Furthermore, change points are intrinsically difficult to recognize when states have time-correlated properties that interfere with state change detection. Several algorithms have been developed for change point detection (CPD), e.g., thresholding,⁸ hidden Markov models,⁹

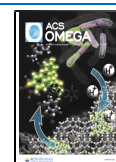
and maximum-likelihood-based methods.^{10–12} However, the algorithms typically require input parameters that sensitively depend on the experimental conditions, and the implementations of the algorithms are generally computationally demanding. Thus, there is a need to develop CPD algorithms that require minimal input parameters for generalizability and robustness and that are computationally efficient for enabling real-time data analysis.

In this work, we present a CPD algorithm that can accurately detect state transitions in experimental time traces from multistate biological systems with *a priori* unknown state properties. The algorithm is generalizable and computationally efficient. Change points are detected by calculating the change in log-likelihood of time-shifted neighboring windows, where multiple window sizes are combined to achieve sensitivity for a wide range of state lifetimes. The performance of the maximum-likelihood multiple-windows change point detection (MM-CPD) is demonstrated using simulated and experimental time traces from a biomolecular sensing technology with single-molecule resolution. Finally, we demonstrate the improved sensitivity and speed of MM-CPD compared to alternative methods reported in the scientific literature.

Received: May 12, 2021

Accepted: June 16, 2021

Published: July 1, 2021



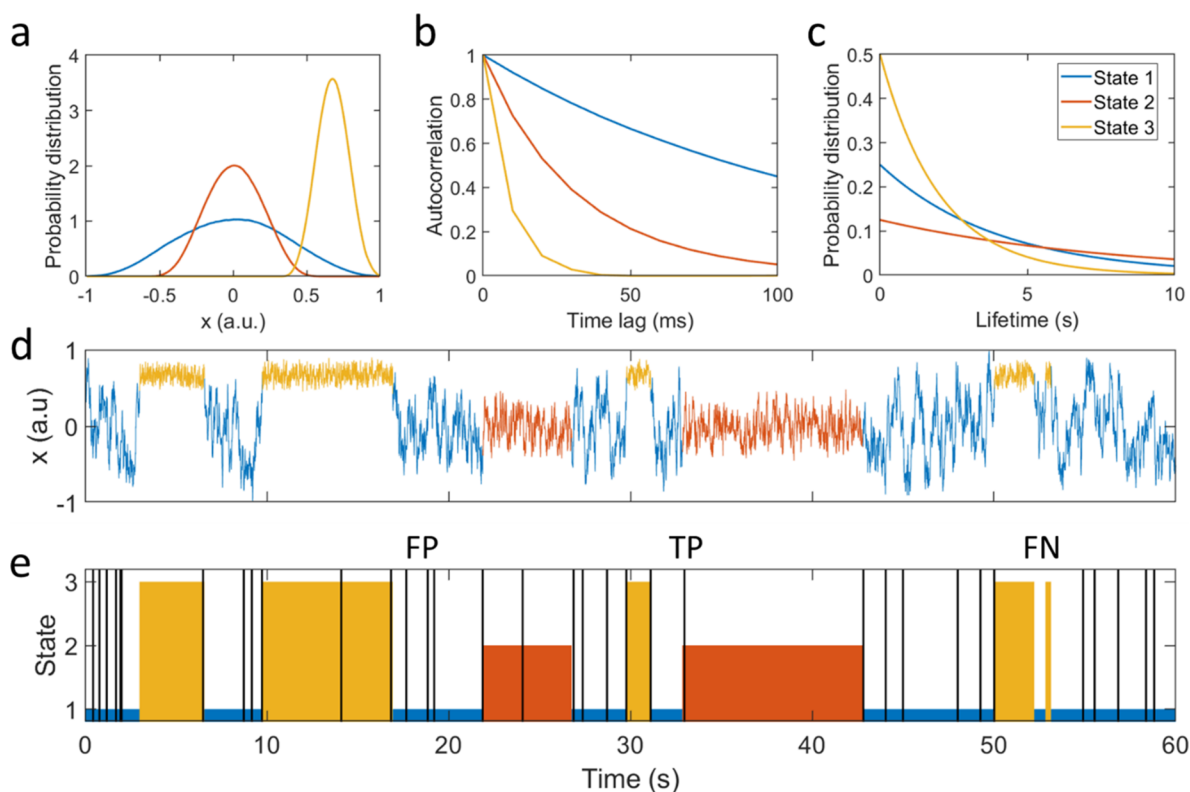


Figure 1. Change point detection (CPD) in a biological system with multiple heterogeneous states. Illustration of the scientific problem using simulated states and time traces of a model system with three mutually exclusive states. (a) The states have overlapping probability distributions of coordinate x . (b) The traces $x(t)$ of each state have different autocorrelation properties (0 to 10 time lag data points). (c) The states have different lifetime distributions, represented as single-exponential distributions with different characteristic lifetimes. (d) Simulated time trace $x(t)$, where the color indicates the state of the system. The sampling rate of the time trace is 100 Hz. (e) State transitions detected by a CPD algorithm (example). The black vertical lines indicate the detected state transitions. Not all state transitions are true positives (TP), i.e., correctly detected CPs, since both false positives (FP) and false negatives (FN) are observed.

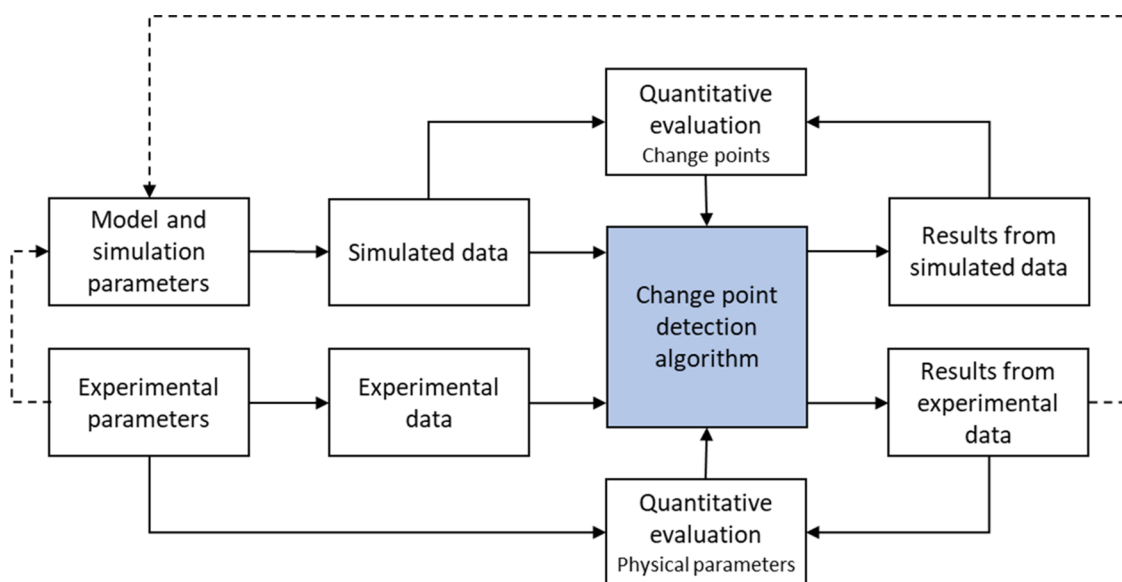


Figure 2. Flow chart for CPD algorithm development using simulated data and experimental data. The arrows indicate the information flow. Analysis of simulated data allows a quantitative evaluation of the algorithm by comparing detected change points with the true change points in the simulation. The CPD algorithm is evaluated on experimental data by comparing scaling relationships between experimental input parameters and physical output parameters. The dashed arrows indicate how the simulation parameters are adapted based on experimental parameters and results.

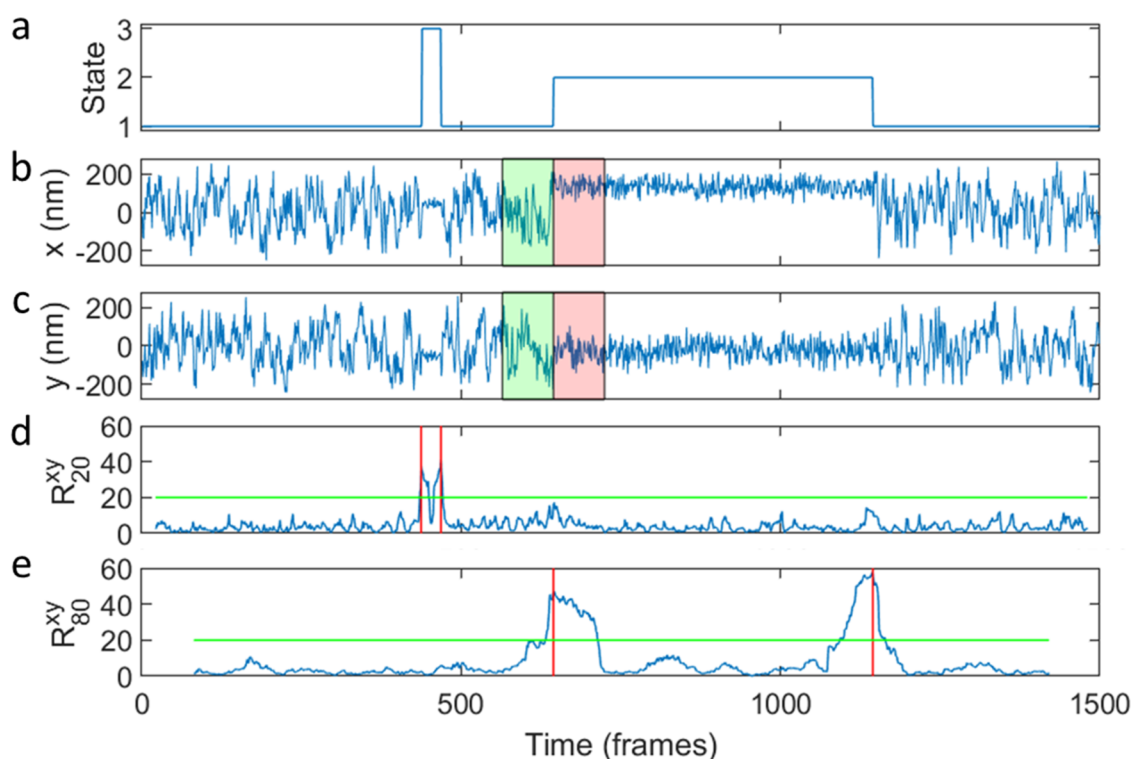


Figure 3. Response time traces in a simulated two-dimensional system with multiple states. (a) Simulated state as a function of time. (b, c) Simulated x and y time traces (a sampling rate of 100 Hz). The green and red boxes indicate two neighboring windows A and B (eq 5) with a window size of 80 data points. (d, e) Response time traces for window sizes of 20 and 80 data points, respectively. The red vertical lines indicate the detected change points, which are identified by applying a threshold (green line).

2. CHANGE POINT DETECTION METHODS

2.1. Data Analysis Challenge. Figure 1 illustrates the data analysis challenge for a biological system with multiple heterogeneous states. A simulated time trace is shown of a system with three states, where the states have overlapping probability distributions, different autocorrelations, and different lifetime distributions. The overlap complicates state identification, the time-correlation influences the time that is required to recognize a state, and the presence of a range of lifetimes complicates the detection of short-lived states, making the data analysis prone to misidentifications of states and state transitions.

2.2. CPD Algorithm Development. Figure 2 shows a flow chart that illustrates how a CPD algorithm is developed and validated using both simulated and experimental data. Simulated data gives the opportunity to quantitatively evaluate a CPD algorithm by comparing detected change points with the known change points in the simulation. In order to generate simulated data that is a good representation of an experimental system, a simulation model needs to be developed, based on knowledge of the system, biophysical equations, and parameters derived from experimental data. By performing the quantitative evaluation, algorithm parameters can be tuned to optimize the CPD algorithm. In parallel, the algorithm is tested on experimental data. Here, the locations of true change points are not known. Therefore, quantitative evaluations do not focus on individual change points but rather on scaling relationships between experimental input parameters and extracted physical output parameters, such as lifetime distributions.

2.3. Quantitative Evaluation of CPD with Simulated Data. Figure 1e shows the detected state transitions after CPD.

Detected CPs can be either true positives (TP) i.e., detected change points that correspond to a true change point or can be false positives (FP). Missed state transitions are categorized as false negatives (FN). The CPD performance can be evaluated by calculating the F1-score,¹³ which is equal to the maximum value of 1 if no false positives and no false negatives are present.

$$\text{F1-score} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})} \quad (1)$$

2.4. Maximum-Likelihood Multiple-Windows Change Point Detection (MM-CPD). The main principle of MM-CPD is to calculate the probability of a change in distribution as a function of time. First, a distribution is assumed, and a maximum-likelihood approach is applied to calculate the distribution parameters. A distribution with a mean μ , standard deviation σ , and nearest-neighbor coupling parameter ϵ is assumed. This model is an approximation of an Ornstein–Uhlenbeck process.¹²

For convenience, we introduce θ , which describes the model parameters:

$$\theta \equiv (\mu, \sigma, \epsilon) \quad (2)$$

The probability density function of a variable x_i is given by

$$f(x_i; \theta) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left[-\frac{(x_i - \epsilon x_{i-1} - (1 - \epsilon)\mu)^2}{2\sigma^2}\right] \quad (3)$$

The log-likelihood of a time trace of N data points is given by

$$g(x_i; \theta) = \log\left(\prod_{i=1}^N f(x_i; \theta)\right) = \sum_{i=1}^N \log(f(x_i; \theta)) \quad (4)$$

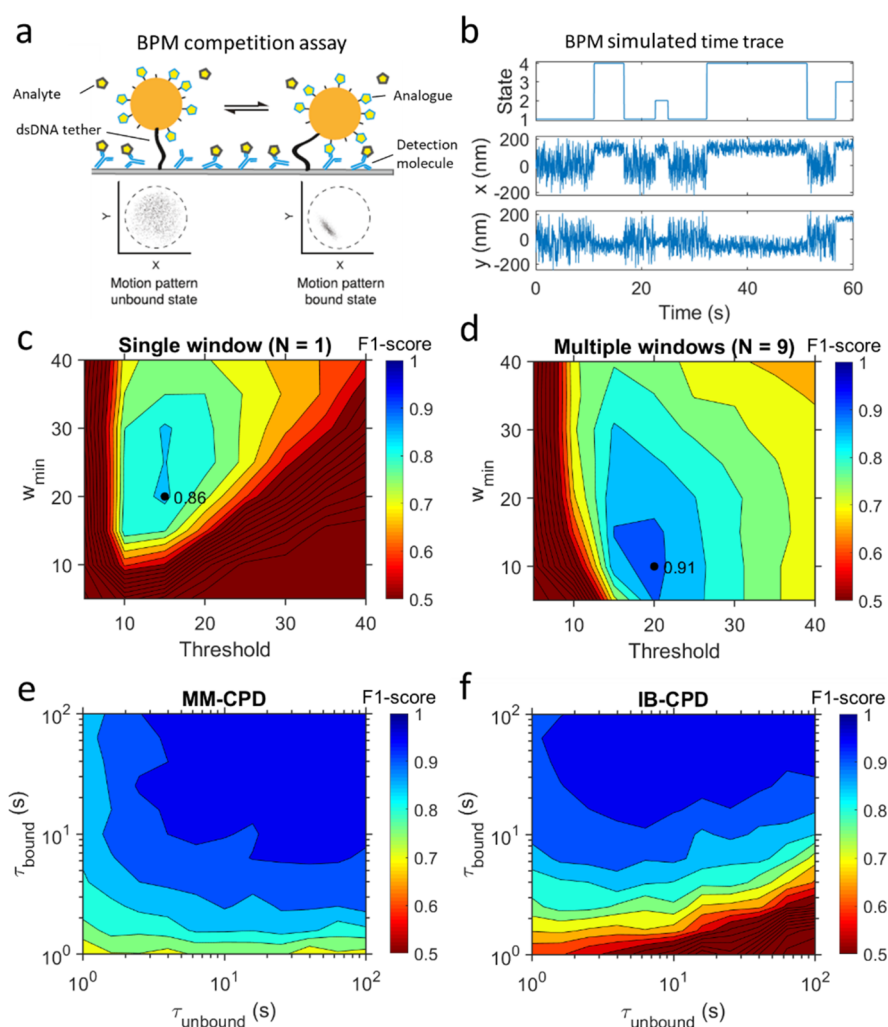


Figure 4. MM-CPD algorithm performance studied on simulated BPM data. (a) BPM competition assay system with a dsDNA tether, analogue molecules on the particle, and detection molecules on the surface.¹⁵ Switching between bound and unbound states is induced by reversible interactions between the analogue molecules and detection molecules. Analyte molecules can also bind to the detecting molecules. Therefore, the number of switches between bound and unbound states is reduced as a function of analyte concentration. Observed motion patterns in the unbound state are circular in shape. In contrast, motion patterns in the bound state can have different shapes and sizes. Images are obtained from Yan *et al.*¹⁵ (b) Simulated state and x and y time traces of a single BPM particle. State 1 refers to the unbound state in (a) and states 2, 3, and 4 are bound states with different distributions. (c) Performance of MM-CPD with a single window. The F1-score is shown as a function of the threshold and window size w_{\min} . (d) MM-CPD with multiple windows ($N = 9$) gives an improved F1-score compared to the single-window approach. (e, f) F1-score as a function of the mean bound state lifetime τ_B and mean unbound state lifetime τ_U for the MM-CPD and IB-CPD.

Maximization of eq 4 with respect to the model parameters provides expressions for the maximum-likelihood estimators (MLE), $\hat{\mu}$, $\hat{\epsilon}$, and $\hat{\sigma}$. Supporting Information Section 1 provides a detailed derivation of the MLEs.

The null hypothesis assumes that the data points in two neighboring windows A and B of equal window size w originate from a single distribution:

$$H_0: \theta_A = \theta_B = \theta_{AB} \quad (5)$$

The alternative hypothesis assumes that the data points in windows A and B come from different distributions:

$$H_1: \theta_A \neq \theta_B \quad (6)$$

For both hypotheses the log-likelihood can be calculated with eq 4 and the MLEs corresponding to the hypothesis. Calculating the log-likelihood ratio between the two hypotheses provides an expression for the likelihood of a change point:

$$R_w = \log \left[\frac{\prod_{i=1}^w f(x_i; \hat{\theta}_A) \prod_{i=w+1}^{2w} f(x_i; \hat{\theta}_B)}{\prod_{i=1}^{2w} f(x_i; \hat{\theta}_{AB})} \right] \quad (7)$$

which can be expressed in terms of individual log-likelihood functions:

$$R_w = g(x_{1,\dots,w}; \theta_A) + g(x_{w+1,\dots,2w}; \theta_B) - g(x_{1,\dots,2w}; \theta_{AB}) \quad (8)$$

We refer to R_w as the response that relates to the probability of a change in distribution. In multidimensional systems, responses from different traces can be added. For example, the response of a two-dimensional system with traces of variables x and y is given by

$$R_w^{xy} = R_w^x + R_w^y \quad (9)$$

Figure 3 shows the time trace of the response for simulated two-dimensional time traces with multiple states. Panels 3a–c

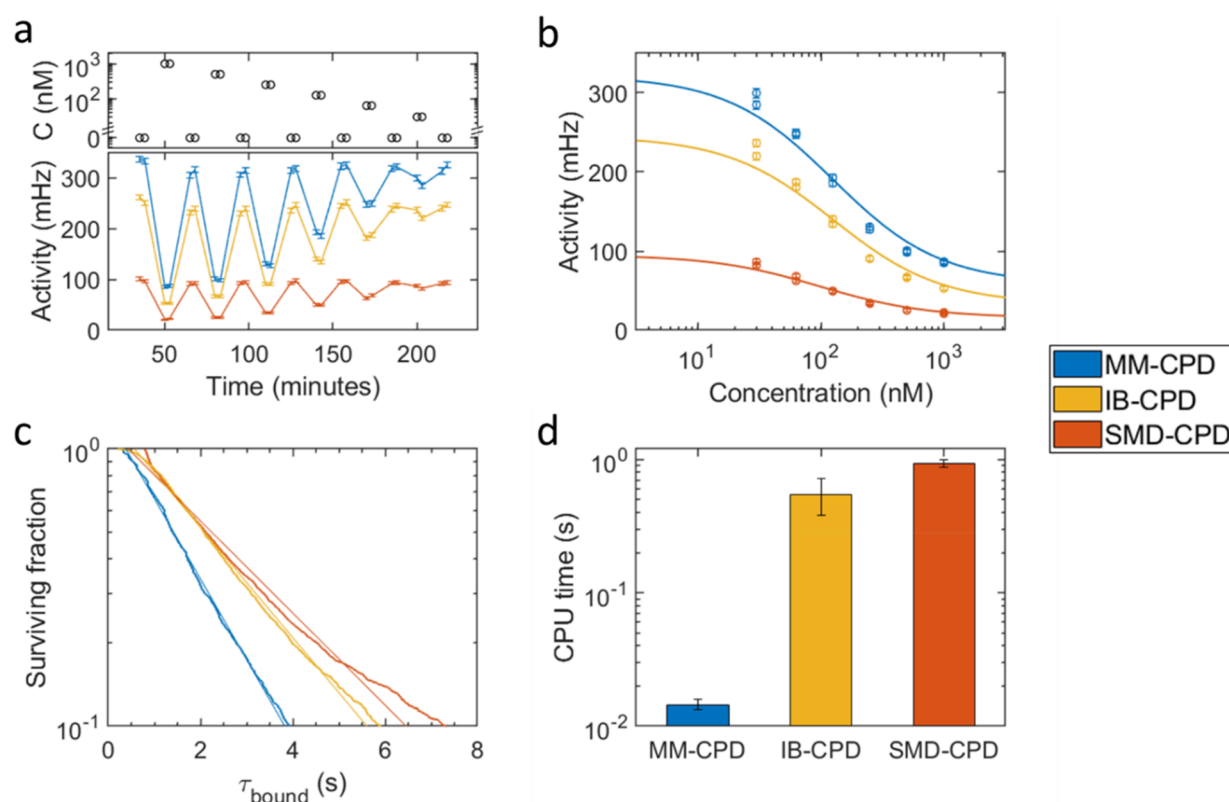


Figure 5. CPD algorithms applied to experimental BPM data for continuous monitoring of single-stranded DNA. Comparison between MM-CPD, IB-CPD, and SMD-CPD. (a) Top panel shows the concentration of the ssDNA analyte as a function of time. The bottom panel shows the detected switching activity as a function of time with MM-CPD, IB-CPD, and SMD-CPD. The error bars indicate the 99% confidence intervals of the mean of the fitted Poisson distribution. (b) Dose–response curves of ssDNA determined with the different CPD algorithms. The curves were fitted with a Hill equation with a Hill coefficient 1. (c) Survival curves of the bound state lifetimes determined from the detected change points with the different CPD algorithms. (d) Average CPU time required to analyze an experimental time trace of 180 s (a sampling rate of 33.7 Hz) for the different CPD algorithms implemented in MATLAB.

show the state and x and y time traces of the simulated data, respectively. The response is plotted as a function of time for window sizes of 20 and 80 data points (Figure 3d,e). The response time trace is a measure for the probability of a change point as a function of time.

The red vertical lines in Figure 3 indicate the change points that are identified from the response signals by applying a threshold. Change points that are detected with a small window size typically correspond to large changes in the distribution. Applying a larger window size allows detecting smaller changes in distribution, since a larger number of data points give a more precise estimation of the model parameters. However, states with a short lifetime, typically shorter than the window size, will be missed with the larger window size. The threshold is an important parameter that can be tuned to minimize the number of FP and FN. Typically, a higher threshold results in an increased number of FN, whereas a lower threshold gives more FP.

The MM-CPD algorithm combines the detected change points from multiple windows. The minimum window size w_{min} and number of windows N are input parameters that define the list of window sizes. The sizes of the respective window are chosen as follows:

$$w_1 = w_{\text{min}}, w_2 = \text{round}(w_{\text{min}} \cdot \sqrt{2}) \quad (10)$$

The sizes of the windows from 3 to N are given by

$$w_i = 2 \cdot w_{i-2} \quad (11)$$

The increment by a factor of two is chosen for the reason of computational efficiency, since the log-likelihood of the null hypothesis (eq 5) for a window size w_i can be used for calculating the log-likelihood of the alternative hypothesis (eq 6) for a window size of $2 \cdot w_i$. Combining the change points (CPs) from multiple windows is a sequential procedure in which the CPs of w_1 are all accepted. Change points from the next window size w_i are accepted only if the distance to already accepted CPs is larger than w_{i-1} . The process of combining change points from multiple window sizes is performed sequentially in an ascending order, i.e., from the smallest window size to the largest window size. The final step of the MM-CPD algorithm is to perform a test on accepted CPs with a minimum distance of w_N to neighboring CPs. These CPs are rejected if R_{w_N} is lower than the threshold. This step is important to reduce the number of FP in systems with long state lifetimes.

3. RESULTS

The MM-CPD algorithm discussed in the previous paragraph is generally applicable for CPD in multistate biological systems. Here, we will apply the algorithm to analyze simulated and experimental time traces of a continuous biomolecular sensing technology with single-molecule resolution, called biosensing by particle mobility (BPM).^{14–16} The biophysical sensing methodology relies on detecting changes in the motion of tethered particles induced by reversible affinity-based interactions between a biofunctionalized particle and surface. Different

molecular sensor designs have been demonstrated, all exhibiting motion pattern changes due to state switching between bound and unbound states. In this paper, we will analyze experimental data of a BPM competition assay, which is schematically drawn in Figure 4a. Here, the surface is provided with antibodies and the particles with analyte analogues. When no analyte is present in solution, then particles switch frequently between unbound and bound states due to the affinity between antibodies on the surface and analyte analogues on the particle. When analyte molecules are present in solution, then these bind to the antibodies, thereby inhibiting binding of the particle to the surface and increasing the mean unbound state lifetime of the particles. BPM experiments provide time traces with the heterogeneities presented in Figure 1b–d, namely, multiple states, states with different autocorrelations, and a wide range of state lifetimes. The Supporting Information Section 2 provides a more detailed explanation of the BPM system and the BPM time trace simulations.

3.1. Application to Simulated BPM Data. BPM time traces with single-exponential distributed state lifetimes were simulated and analyzed with the MM-CPD algorithm. Figure 4b shows the x , y , and state time traces for a typical BPM particle with a single unbound state and multiple bound states. Figure 4c shows the F1-score as a function of the algorithm parameters, for the case of a single window size ($N = 1$). The F1-score clearly depends on the threshold and on the minimum window size w_{min} . Given a certain w_{min} , an optimal threshold can be found, where a lower threshold would lead to more FP and a higher threshold to more FN. The same data set was analyzed with a multiple-windows approach ($N = 9$), showing that a higher F1-score is achieved by combining the change points from multiple windows (Figure 4d).

To test the performance of MM-CPD for a wide range of state lifetimes, time traces with mean bound state lifetime τ_B and mean unbound state lifetime τ_U ranging from 1 to 100 s were simulated. In order to be effective for a wide range of state lifetimes, the following algorithm settings were chosen: $w_{min} = 10$, $N = 9$, and $threshold = 20$. The simulations were used to compare the performance of the MM-CPD algorithm to the information-based CPD (IB-CPD) algorithm developed by Wiggins.¹² The latter is a CPD algorithm for biophysical systems with multiple heterogeneous states. Figure 4e,f shows the F1-score as a function of τ_B and τ_U for both algorithms. Both algorithms show a high F1-score for bound states with a long lifetime. For traces with short bound state lifetimes and long unbound state lifetimes, the MM-CPD algorithm gives a significantly higher F1-score compared to the IB-CPD algorithm.

3.2. Application to Experimental BPM Data. In order to demonstrate the performance of the algorithm on experimental data, we analyzed BPM data from Yan *et al.*¹⁵ This study describes a competition assay for continuous biosensing of single-stranded DNA (ssDNA) and small molecules. In short, addition of analyte molecules decreases the probability for a particle to transition from an unbound to a bound state. The average number of switching events between bound and unbound states per particle per unit of time is referred to as the activity. The activity is obtained by fitting a Poisson distribution on the number of detected change points per particle. Figure 5a shows the activity as a function of time in a BPM assay with varying ssDNA concentrations. The system behaves as expected, since an increase in analyte concentration leads to a decrease in activity.

The activity was calculated with MM-CPD, IB-CPD, and with the second-moment divergence CPD (SMD-CPD) algorithm that was applied by Yan *et al.* The latter SMD-CPD algorithm was developed specifically for BPM.¹⁴ Clearly, more change points are detected with MM-CPD and IB-CPD compared to SMD-CPD. Figure 5b shows the dose–response curves for ssDNA determined with the three CPD algorithms.

A further comparison of the CPD algorithms was performed by extracting the distributions of state lifetimes from the measured data. In a BPM assay, the bound state lifetime is determined by the dissociation properties of the molecular interactions between the particle and the surface. In the regime of single-molecular interactions, the dissociation behavior should obey a first-order rate equation with a well-defined dissociation rate constant (k_{off}) and single-exponential distributed bound state lifetimes.

Here, we studied to what extent the algorithms reproduce single-exponential distributed bound state lifetimes. We used experimental BPM traces, applied the three algorithms, and classified states between consecutive change points. States were classified as a bound state, unbound state, or an undefined state by analysis of the standard deviation in x and y (see Supporting Information Section 3). Figure 5c shows cumulative distribution functions, also called survival plots, of the bound state lifetimes determined with the three different CPD algorithms. A straight line in a survival plot with a linear x -axis and a logarithmic y -axis indicates a single-exponential distribution. The figure shows that MM-CPD and IB-CPD give straight lines, while the data from SMD-CPD deviates from a straight line. This indicates that SMD-CPD detects short and long state lifetimes with nonuniform sensitivity, resulting in an under-representation of short state lifetimes in the distribution. A comparative analysis of the bound and unbound state lifetime distributions for different concentrations can be found in the Supporting Information Section 5.1.

The computational efficiencies of the three CPD algorithms are reported in Figure 5d. The data shows the average CPU time required for the analysis of experimental time traces, for CPD algorithm implementations in MATLAB. The results show that the MM-CPD is between one and two orders of magnitude faster than the IB-CPD and SMD-CPD algorithms. The increased CPU time of IB-CPD can be attributed to the segmentation algorithm that is applied in IB-CPD, which is a reiterative procedure. This leads to an analysis time that is dependent on the number of detected change points; data in the Supporting Information Section 4.2 shows that the CPU time in IB-CPD depends on the state lifetimes.

4. DISCUSSION

From a statistical perspective, a minimum number of data points is required to detect a specific state transition with a certain reliability. The required number of data points is dependent on the significance of the change in distribution but is also influenced by time-correlation properties of the states. Detection of short-lived states is therefore theoretically limited. High sensitivity for short-lived states is desirable and can be achieved by applying a window size close to this theoretical limit. However, in time traces with multiple heterogeneous states, transitions are unique and for each transition, a different window size might be optimal. By combining detected change points from multiple windows, a high sensitivity is achieved for a wide range of state lifetimes. In particular, small window sizes are sensitive for significant state transitions between short-lived

states. On the other hand, larger window sizes are sensitive for less significant state transitions between states with longer lifetimes. This explains why the multiple-windows CPD approach performs significantly better compared to the single-window approach (Figure 4c,d).

Supporting Information Section 6 gives an overview of frequently applied methods for CPD in biological time traces. These include half-amplitude thresholding,⁸ maximum-likelihood-based methods,¹¹ and hidden Markov models.⁹ Most methods rely on a defined number of states as input and therefore are not suitable for systems with *a priori* undefined states as presented in this paper. The IB-CPD algorithm of Wiggins¹² can be applied without *a priori* knowledge of the states. In comparison to the IB-CPD algorithm, the MM-CPD algorithm is more likely to detect short-lived bound states, especially when neighboring unbound states are long. This effect was clearly visible in the F1-scores of the simulated data (Figure 4e,f). In addition, the lifetime analysis in experimental BPM data showed shorter bound state lifetimes for the MM-CPD approach compared to the IB-CPD approach (Figure 5c). The difference is most likely due to the binary segmentation algorithm that is applied in IB-CPD. The algorithm starts with determining the most likely change point in a time trace. If the candidate change point is significant, it is accepted, and the trace is split into two segments. The process is repeated until no new candidate change points are accepted. This method might have difficulties to detect a short-lived bound state between two long identical unbound states, since the distributions on both sides of a candidate change point are dominated by the unbound state. In contrast, the MM-CPD approach compares two windows of data points. Thus, a short-lived bound state will result in two peaks in the response trace of a window size close to the state lifetime. In general, short-lived states that are hidden in the distribution of neighboring identical states are more likely to be detected with MM-CPD compared to IB-CPD.

The MM-CPD algorithm is designed to be robust for traces with time-correlated data points, such as BPM. Similar to the IB-CPD, the MM-CPD assumes a Gaussian distribution with a nearest-neighbor coupling parameter. Including the nearest-neighbor coupling parameter in the distribution model significantly increases the robustness of CPD in traces with time-correlated data points (Supporting Information Section 4.1). The MM-CPD has only three algorithm parameters. The following settings were found to give good results for BPM data with two input time traces: $w_{min} = 10$, $N = 9$, and $threshold = 20$. In some cases, it might be beneficial to further tune the algorithm parameters. For example, the minimum window size can be increased in systems with only long state lifetimes. Also, the number of windows could be decreased if higher computational efficiency is desired. The threshold level might also be tuned, especially if the number of input time traces is changed, since responses are added linearly (eq 9). Furthermore, it should be considered that the choice of the threshold can have a large influence on the extracted physical parameters (Supporting Information Section 5.2). Therefore, performing a quantitative evaluation with both simulated and experimental data as shown in Figure 2 is important for validation when applying the MM-CPD in other biological applications.

The performance of the MM-CPD was discussed for BPM experiments. In BPM, the detection of short-lived states is theoretically limited by the autocorrelation times of the states. Further increasing the sampling rate does not significantly improve the CPD performance, but it does decrease the

computational efficiency. In other biological processes with shorter autocorrelation times, the sampling rate can be increased to improve the sensitivity for short-lived states. When autocorrelation times are long in comparison to the intersampling time, it might be beneficial to increase the minimum window size.

One of the main advantages of the MM-CPD is that the algorithm is computationally very efficient. In contrast to alternative methods, the analysis time is independent of the state lifetimes (Supporting Information Section 4.2). This leads to a predictable total analysis time, which is crucial in real-time applications. We showed that the time to analyze a two-dimensional BPM time trace of 180 s (~6000 data points) is only ~0.02 s in MATLAB (Figure 5d). This speed enables real-time CPD of ~10⁴ BPM time traces in parallel. Even faster biological processes, having shorter autocorrelation times measured, e.g., at a sampling rate of 1 kHz, can be analyzed in real time with MM-CPD for hundreds of time traces in parallel. We expect that the MM-CPD algorithm will allow real-time analysis with a time lag that is of the order of magnitude of the largest window size.

5. CONCLUSIONS

We developed a CPD algorithm for rapid and reliable detection of state transitions in experimental time traces from multistate biological systems with *a priori* unknown state properties and a wide range of state lifetimes. The maximum-likelihood multiple-windows change point detection (MM-CPD) algorithm is controlled by three input parameters with a clear significance. The algorithm was validated using both simulated and experimental data of a biosensing method with single-molecule resolution. The MM-CPD shows an increased sensitivity for short-lived states that are hidden in the distribution of neighboring identical states, and the algorithm is between one and two orders of magnitude faster compared to alternative methods. The computational efficiency allows CPD in thousands of time traces in parallel, for a real-time readout and statistical analysis of stochastic signals from dynamic biological systems.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.1c02498>.

Derivations of the MLEs, BPM time trace simulations, procedure for extracting lifetimes from BPM time traces with detected CPs, additional results on simulated and experimental data, and overview of CPD methods in the literature (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Menno W.J. Prins – Department of Biomedical Engineering and Institute for Complex Molecular Systems (ICMS), Eindhoven University of Technology, Eindhoven 5612, The Netherlands; Department of Applied Physics, Eindhoven University of Technology, Eindhoven 5612, The Netherlands; Helia BioMonitoring, Eindhoven 5612, The Netherlands; orcid.org/0000-0002-9788-7298; Email: m.w.j.prins@tue.nl

Authors

Max H. Bergkamp – Department of Biomedical Engineering and Institute for Complex Molecular Systems (ICMS), Eindhoven University of Technology, Eindhoven 5612, The Netherlands; orcid.org/0000-0001-8904-4322

Leo J. van IJzendoorn – Department of Applied Physics, Eindhoven University of Technology, Eindhoven 5612, The Netherlands; Institute for Complex Molecular Systems (ICMS), Eindhoven University of Technology, Eindhoven 5612, The Netherlands

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.1c02498>

Author Contributions

M.B. developed the MM-CPD algorithm. M.B. performed the simulations and the data analysis. All authors interpreted data, discussed results, and co-wrote the paper. All authors approved the submitted version of the manuscript.

Notes

The authors declare the following competing financial interest(s): M.P. is cofounder of Helia BioMonitoring. All authors declare no further competing interests.

The data and software of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

Part of this work was funded by the Dutch Research Council (NWO), section Applied and Engineering Sciences, under grant number 16255. Part of this work was funded by the Safe-N-Medtech H2020 project under grant agreement no. 814607.

REFERENCES

- (1) Carter, N. J.; Cross, R. A. Mechanics of the kinesin step. *Nature* **2005**, 308.
- (2) Sourjik, V.; Wingreen, N. S. Responding to chemical gradients: Bacterial chemotaxis. *Curr. Opin. Cell Biol.* **2012**, 24, 262–268.
- (3) Manzo, C.; Garcia-Parajo, M. F. A review of progress in single particle tracking: From methods to biophysical insights. *Rep. Prog. Phys.* **2015**, 124601.
- (4) Neuman, K. C.; Nagy, A. Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* **2008**, 491.
- (5) Schuler, B.; Eaton, W. A. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **2008**, 16.
- (6) Jungmann, R.; et al. Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **2014**, 313.
- (7) Shi, W.; Friedman, A. K.; Baker, L. A. Nanopore Sensing. *Anal. Chem.* **2017**, 89, 157–188.
- (8) Finzi, L.; Gelles, J. Measurement of lactose repressor-mediated loop formation and breakdown in single DNA molecules. *Science* **1995**, 378.
- (9) Beausang, J. F.; Nelson, P. C. Diffusive hidden Markov model characterization of DNA looping dynamics in tethered particle experiments. *Phys. Biol.* **2007**, 205.
- (10) Watkins, L. P.; Yang, H. Detection of intensity change points in time-resolved single-molecule measurements. *J. Phys. Chem. B* **2005**, 109, 617–628.
- (11) Manzo, C.; Finzi, L. Quantitative Analysis of DNA-Looping Kinetics from Tethered Particle Motion Experiments. *Methods Enzymol.* **2010**, 475, 199–220.
- (12) Wiggins, P. A. An Information-Based Approach to Change-Point Analysis with Applications to Biophysics and Cell Biology. *Biophys. J.* **2015**, 346.

(13) Aminikhanghahi, S.; Cook, D. J. A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **2017**, 339.

(14) Visser, E. W. A.; Yan, J.; van IJzendoorn, L. J.; Prins, M. W. J. Continuous biomarker monitoring by particle mobility sensing with single molecule resolution. *Nat. Commun.* **2018**, 2541.

(15) Yan, J.; van Smeden, L.; Merkx, M.; Zijlstra, P.; Prins, M. W. J. Continuous Small-Molecule Monitoring with a Digital Single-Particle Switch. *ACS Sens.* **2020**, 5, 1168–1176.

(16) Lubken, R. M.; de Jong, A. M.; Prins, M. W. J. Multiplexed Continuous Biosensing by Single-Molecule Encoded Nanoswitches. *Nano Lett.* **2020**, 2296.