Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v11n1p208

Comparison of classic and novel change point detection methods for time series with changes in variance

By Breitenberger et al.

Published: 26 April 2018

Electronic Journal of Applied Statistical Analysis
Vol. 11, Issue 01, April 2018, 208-234
DOI: 10.1285/i20705948v11n1p208

Comparison of classic and novel change point detection methods for time series with changes in variance

Sandra Breitenberger*[a], Dmitry Efrosinin[b,c], Nicole Hofmann[c], and Wolfgang Auer[d]

[a] *Linz Center of Mechatronics GmbH (LCM), 69 Altenberger Straße, Linz, 4040, Austria*
[b] *Johannes Kepler University, 69 Altenberger Straße, Linz, 4040, Austria*
[c] *RUDN University, 6 Miklukho-Maklaya st, Moscow, 117198, Russia*
[d] *Smartbow GmbH, 3 Jutogasse, Weibern, 4675, Austria*

Published: 26 April 2018

Segmentation or change point detection is a very common topic in time series analysis, anomaly detection and pattern recognition. In Breitenberger et al. (2017) the time series generated by sensors with 3D accelerometers were analysed. It was noticed that such series consist of segments of independent and correlated observations. Hence the appropriate methods for change point detection for both data types must be implemented simultaneously. This paper provides an auxiliary comparison analysis which we intend to implement later for the above mentioned acceleration data. The available methods require usually a long execution time, so that it is time-consuming if several methods should be compared. In the framework of the present publication we want to give additional help for detecting a suitable change point detection method and for finding a good parameter setting. Our analysis is performed on simulated time series, that are normally distributed with constant but unknown mean and changes in variance.

keywords: change point, time series segmentation, non-stationary time series, binary classification, CUSUM method.

*Corresponding author: dmitry.efrosinin@jku.at, sandra.breitenberger@lcm.at.

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1 Introduction

In application areas like biometrics, neurology, speech recognition, agriculture, climatology, finance, telecommunication systems, etc., where data samples are represented in form of non-stationary time series, it is often necessary to segment the given series into stationary blocks. Such a segmentation includes evaluation of the periods of stationarity and homogeneity, identification of the change point times of some specified characteristic values like mean, variance, correlation and allows to understand the similarity inside each recognized block. The main problem of the segmentation or change point analysis consists in defining statistical tests for detection and estimation of the change point position. The special task considered here is a multiple change point case which requires appropriate recursive algorithms with a small evaluation time. The present change point detection methods can be classified into two groups: retrospective (offline) and real-time (online) methods. Real-time change point detection is very useful in robot control, where immediate responses are required. Retrospective methods, in contrast, tend to give more robust and accurate detections. Nowadays exist many different approaches for finding change points. Some of these approaches use the cumulative sums (Badagián et al., 2009; Inclán and Tiao, 1994), the ratios of probability densities (Liu et al., 2013), the likelihood-ratio methods (Ibrahim et al., 2003; Rohrbeck, 2013), the Bayesian methods (Ryan et al., 2007) or the standard F-Test (Tsay, 1988; Ureche-Rangau and Speeg, 2011).

This paper represents the following new results. In this study, five offline and three online methods are compared on the basis of normally distributed time series data. All methods – except one that is called BOCPD later – use sliding time windows, such that only parts of the time series data have to be used for analysis. For this purpose, we choose the window lengths 50, 200 and 500. As soon as the variances within these intervals become too different, an alarm for a change point is given. One out of the eight methods is new and is called HeurMeth. It should be marked that the usually retrospective methods have been implemented for sliding time windows too. So, the retrospective analyses are applied for each time segment independently. In our study, we assume that we do not know the number and type of change points in the data streams. We only know, that there are some changes in variance and we want to find them in an online manner.

The paper is organized as follows. In section 2, we start with the definition of a change point problem. In section 3, we will present the different change point detection methods used in our study. In section 4, we will introduce the data generation process for time series with one change point. Afterwards, we will display our simulation results regarding the data sequences defined in this section. In section 5, we will make a further case study for simulated time series with four change points instead of one change point. After the presentation of the new data generation process, we will again show some results concerning the comparison of different change point detection methods – for two online methods also with an additional variance stabilizing transformation. In section 6, we confirm our results by a further simulation study that contains much more time series with one and without any change point respectively. In section 7, we give a short summary of the most important results. For preserving the text flow, additional figures can be found in the appendix.

2 The Change Point Problem

For the definition of a change point problem, we follow an earlier definition found in Badagián (2015). Let $x = \{x_1, x_2, \dots, x_n\}$ be a given time series with m change points at the time points $t_1^*, \dots, t_m^*$ with $1 < t_1^* < \dots < t_m^* \leq n$. The probability density function $f(x_t|\theta)$ depending on the parameter θ is of the form

$$f(x_t|\theta) = \begin{cases} f(x_t|\theta_1), & \text{for } 1 \leq t \leq t_1^* - 1, \\ f(x_t|\theta_2), & \text{for } t_1^* \leq t \leq t_2^* - 1, \\ \dots & \dots \\ f(x_t|\theta_{m+1}), & \text{for } t_m^* \leq t \leq n, \end{cases}$$

where $\theta_1 \neq \theta_2 \neq \dots \neq \theta_{m+1}$ are the variances for each block of observations. In general, the aim consists in detection and identification of the positions $t_1^*, t_2^*, \dots, t_m^*$ and estimation of the unknown parameters $\theta_i, i = 1, 2, \dots, m+1$.

The problem of the multiple change point detection can be formally rewritten as hypothesis testing:

$$\begin{aligned} H_0: \quad & X_t \sim f(x_t|\theta_0),\ 1 \leq t \leq n, \\ H_1: \quad & X_t \sim f(x_t|\theta_1),\ 1 \leq t \leq t_1^* - 1,\ X_t \sim f(x_t|\theta_2),\ t_1^* \leq t \leq t_2^* - 1, \dots, \\ & X_t \sim f(x_t|\theta_{m+1}),\ t_m^* \leq t \leq n,\ \theta_1 \neq \theta_2 \neq \dots \neq \theta_{m+1}. \end{aligned} \tag{1}$$

Now, θ corresponds to the variance of observation t.

In the present paper, the density functions $f(x_1|\theta_1), f(x_2|\theta_2), \dots, f(x_t|\theta_{m+1})$ belong to a normal distribution family $\mathcal{N}(0, \sigma^2)$ with zero mean, i.e. $\theta_i = \sigma_i^2$ and

$$f(x_t|\sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{x_t^2}{2\sigma_i^2}},\ i = 1, 2, \dots, m+1,$$

hence, the change point detection problem (1) is equivalent to the following hypothesis testing:

$$\begin{aligned} H_0: \quad & \sigma_1^2 = \sigma_2^2 = \dots = \sigma_{m+1}^2 = \sigma^2 \\ H_1: \quad & \sigma_1^2 = \dots = \sigma_{t_1^*-1}^2 \neq \sigma_{t_1^*}^2 = \dots = \sigma_{t_2^*-1}^2 \neq \dots \\ & \neq \sigma_{t_{m-1}^*}^2 = \dots = \sigma_{t_m^*-1}^2 \neq \sigma_{t_m^*}^2 = \dots = \sigma_n^2. \end{aligned} \tag{2}$$

Note that within the paper we provide comparison analysis of the different change point detection methods applied to the observations which will be regarded as independent. Nevertheless, we intend to use additionally the methods which are normally applicable in case of correlated data. As it was noticed in Breitenberger et al. (2017), the time series generated by the sensors with 3D accelerometers may consist of segments with independent and correlated observations. Therefore, the change point detection for both data types must be implemented simultaneously. Hence we want to propose an auxiliary comparison analysis in order to check the quality of the different types of methods under the assumption of independence. But to complete the picture we give some examples for segmentation of correlated data in remark 1.

3 Change Point Detection Methods

In table 1, the different change point detection methods are listed. The methods ICM and SIC are normally used for dependent observations. The abbreviations BSOP and HeurMeth are not common in literature and represent abbreviations used in this publication. In the following, we shortly present the ideas of each method; the last three methods are typical online methods.

- The first method, ICSS, uses cumulative sums of squares, which are centered and normalized. The test statistic

$$M(k) = \sqrt{n/2}\max_k \left\{ \frac{\sum_{t=1}^k x_t^2}{\sum_{t=1}^n x_t^2} - \frac{k}{n} \right\},\ 0 < k < n, \tag{3}$$

  behaves asymptotically like a Brownian bridge. This statistic oscillates around 0, but if a sudden change in variance happens, it will leave some specified boundaries. The null hypothesis is rejected when the maximum value of the function $M(k)$ is greater than the asymptotic critical value $D^*_{.05} = 1.358$ and the change point is located at $\hat{k}$ for which $M(k) > D^*_{.05}$ and $M(\hat{k}) = \max_k M(k)$.

- ICM algorithm searches for changes in the parameters of a RCA(1)-model (first order random coefficient autoregressive model) defined as

$$x_t = (\phi + b_t)x_{t-1} + \varepsilon_t, \text{ where } \begin{pmatrix} b_t \\ \varepsilon_t \end{pmatrix} \sim \text{iid}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right). \tag{4}$$

  This iterative cusum method is based on ideas of the ICSS method, where

$$M(k) = \max_k \frac{k^2}{n}(\hat{\theta}_k - \hat{\theta}_n)\Gamma^{-1}(\hat{\theta}_k - \hat{\theta}_n) \tag{5}$$

  with $\hat{\theta}_k = (\hat{\phi}_k, \hat{\omega}_k^2, \hat{\sigma}_k^2)'$, where the estimators are obtained by minimization of $\sum_{t=1}^k (x_t - \phi x_{t-1})^2$ and $\sum_{t=1}^k ((x_t - \hat{\phi}_k x_{t-1})^2 - \omega^2 x_{t-1}^2 - \sigma^2)^2$ and matrix Γ can be estimated adequately. If $M(k) > D^*$, where D^* is a critical value with $1-p$ level of significance, there is a shift in the time series at time $\hat{k}$. The critical value is computed for the 0.01 significance level using simulated time series.

- The SIC method is normally applicable for the autoregressive process, e.g. for the order 1 and one change point:

$$x_t = \begin{cases} c_1 + \phi_1 x_{t-1} + \varepsilon_t, & 1 \le t \le t_1^* - 1, \\ c_2 + \phi_2 x_{t-1} + \varepsilon_t, & t_1^* \le t \le n. \end{cases}$$

The method combines the binary segmentation procedure with a test statistic that is based on Schwarz Information Criterion. In general, SIC is defined as

$$\mathrm{SIC} = -2\log L(\hat{\theta}) + k\log n, \tag{6}$$

where $L(\hat{\theta})$ is the maximum-likelihood function of the model, k the number of parameters to be estimated (free parameters) and n the sample size. Let $\mathrm{SIC}_{H_0}(n)$ the SIC under H_0 in (2) where no change point in a data sample exists and $\mathrm{SIC}_{H_1}(k)$ the criterion under the assumption of a change point at $t = k, 1 \leq k \leq n$. The rejection of H_0 is based on the principle of minimum information criterion, namely H_0 is not rejected if $\mathrm{SIC}_{H_0}(n) < \min_k \mathrm{SIC}_{H_1}(k)$ and is rejected otherwise. In this case, the position of the change point t^* is defined by $\hat{k}$ in such a way that

$$\mathrm{SIC}(\hat{k}) = \min_{1<k<n} \mathrm{SIC}_{H_1}(k). \tag{7}$$

Table 1: Variance change detection methods and their introduction in literature

Short-form of method	Long-form of method
ICSS	Iterated cumulative sums of squares (Inclán and Tiao, 1994)
ICM	Iterative CUSUM method (Badagián et al., 2009)
SIC	Autoregressive model using Schwarz Information Criterion (Ibrahim et al., 2003)
BSOP	Binary segmentation and optimal partitioning (Rohrbeck, 2013)
RuLSIF	Relative unconstrained least-squares importance fitting (Liu et al., 2013)
BOCPD	Bayesian online change point detection (Ryan et al., 2007)
HeurMeth	Heuristic method based on moving variance differences
F-Test	Fisher ratio statistic (moving variance ratios) (Tsay, 1988; Ureche-Rangau and Speeg, 2011)

- The BSOP method is based on cost functions for different segments of the sequence and a penalty term that penalizes a high amount of change points in order to avoid overfitting. Binary segmentation is an iterative approach for minimizing an existing cost function by splitting into two cost functions. With optimal partitioning, a global minimum of the minimization problem can be found. The cost function for the given data sample is determined to

$$C(s,t) = \sum_{i=s}^{t} \Big[\frac{x_i^2}{\hat{\sigma}^2(s,t)} + \log(2\pi\hat{\sigma}^2(s,t))\Big], \tag{8}$$

where $\hat{\sigma}^2(s,t) = \frac{1}{t-s}\sum_{i=s}^{t} x_i^2$ and $s = 1$, $t = n$ at the beginning of the procedure. To get a change point, first the integer value k^* must be determined,

$$k^* = \arg\min_k [C(s,k) + C(k+1,t) + \log(n)], \tag{9}$$

and second, it is necessary to check whether

$$C(s,k^*) + C(k^*+1,t) + \log(n) < C(s,t). \tag{10}$$

- The method RuLSIF is based on non-parametric divergence estimation between two retrospective segments. This method uses the relative Pearson divergence (PE) as a divergence measure, which is estimated by a method of direct density-ratio estimation. This estimation is called unconstrained least-squares importance fitting (uLSIF) and it directly learns the density-ratio function in the least-squares fitting framework. The uLSIF-based method is further improved by considering relative density ratios. By using a density ratio estimator $\hat{g}(x_1, \ldots, x_n)$, the α-relative PE divergence is approximated as

$$\begin{aligned} \hat{\text{PE}}_\alpha = -\frac{\alpha}{2n}\sum_t \hat{g}(x_{t:t+k-1})^2 - \frac{1-\alpha}{2n}\sum_t \hat{g}(x_{t-k:t-1})^2 \\ +\frac{1}{n}\sum_t \hat{g}(x_{t:t+k-1}) - \frac{1}{2} \end{aligned} \tag{11}$$

  for two consecutive time series segments $x_{t-k:t-1}$ and $x_{t:t+k-1}$ with $k+1 \leq t \leq n-k+1$. To identify a change point the value (11) is compared with an optimal critical score value which can be evaluated by simulated time series.

- BOCPD is a Bayesian online method, that computes the probability distribution $\mathbb{P}[r_t|x_{1:t}]$ of the length of the current run r_t (time since the last change point). The method is based on causal predictive filtering – generating a distribution $\mathbb{P}[x_{t+1}|x_{1:t}]$ of the next unseen datum of the sequence out of the data already observed. The change point probability is defined as

$$\mathbb{P}[r_t = 0, x_{1:t}] = \sum_{j=0}^{t-1} \mathbb{P}[r_{t-1} = j|x_{1:t-1}]\mathbb{P}[x_{1:t-1}]\pi_t^{(j)} H(t), \tag{12}$$

  where $\mathbb{P}[x_{1:t-1}] = \sum_{j=0}^{t-1}\mathbb{P}[r_{t-1} = j, x_{1:t-1}]$, $\mathbb{P}[r_t = j|x_{1:t}] = \frac{\mathbb{P}[r_t=j,x_{1:t}]}{\mathbb{P}[x_{1:t}]}, \pi_t^{(j)} = f_{\mathcal{N}(\mu_t^{(j)},\tau_t^{2(j)})}(x_t)$, $j = 0, 1, \ldots, t-1$, and $H(t)$ is the hazard function.

- The heuristic method HeurMeth is derived from the Mann-Whitney statistic for medians and simply computes the absolute differences of the medians of variances of two subsequent sliding windows W_1 and W_2,

$$H(t_0, n) = |\tilde{y}_{W_1(t_0)} - \tilde{y}_{W_2(t_0)}|, \tag{13}$$

  where $\tilde{y}_{W_1(t_0)}$ and $\tilde{y}_{W_2(t_0)}$ are the empirical medians for the testing point t_0. If the maximum of this statistic $\max\limits_{1<t_0<n} H(t_0, n)$ takes a value which is greater as a predefined threshold c_n, a change point is found. To speed up execution time, we used instead of the calculation of one variance out of the data of the whole sliding window the calculation of several variances of length 10 and used the median for further comparisons. This is the only new method of this study.

- The last method is based on the maximum of the Fisher ratio statistic, which is commonly used for validating differences in the variances of two normally distributed samples. The test statistic is based on the moving variance ratio
$$F(t_0, n) = \frac{s^2_{W_1(t_0)}}{s^2_{W_2(t_0)}} \tag{14}$$
and is defined as
$$\tilde{F}(t_0, n) = \max\{F(t_0, n), 1/F(t_0, n)\}, \tag{15}$$
where $s^2_{W_1(t_0)}$ and $s^2_{W_2(t_0)}$ are the empiric variances for sliding windows W_1 and W_2 respectively. The change point is selected, if $\max\limits_{1<t_0<n} H(t_0, n) > c_n$.

All methods have been implemented in a way to be able to find multiple change points in a time series. The methods RuLSIF, HeurMeth and F-Test use two adjacent sliding time windows (with a shift of 1 time point), the methods ICSS, ICM, SIC and BSOP use only one sliding time window (with a shift that corresponds to the half of the window length to speed up the retrospective methods) and the method BOCPD does not need any sliding time window, as it is computed for each new data point.

4 Case Study I

4.1 Data Generation

For our study, time series with $n = 1000$ and one change point at position $t_0 = 501$ have been generated, where

$$X_t \sim \mathcal{N}(0, \sigma_1^2),\ 1 \le t \le t_0 - 1, \quad X_t \sim \mathcal{N}(0, \sigma_2^2),\ t_0 \le t \le n.$$

Let $E = \{1, 2, 3, 4, 6, 8, 10, 15\}$ be a set of values. Then, the standard deviations satisfy $\sigma_1 \in E$ and $\sigma_2 \in E \setminus \{\sigma_1\}$. With all possible combinations, we get $|E|(|E| - 1) = 56$ time series with different changes in variance.

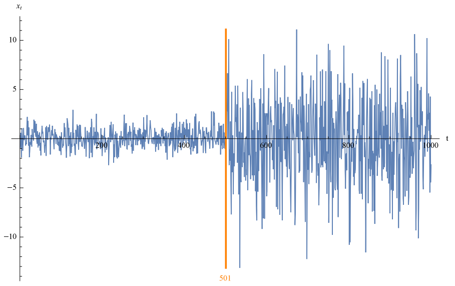


Figure 1: Simulated time series with one change point ($\sigma_1 = 1$ and $\sigma_2 = 4$)

Figure 1 shows an exemplary time series for $\sigma_1 = 1$ and $\sigma_2 = 4$ with a change point at $t_0 = 501$ and was realized in Mathematica by the repeated application of the command

```
RandomVariate[NormalDistribution[0,sigma]].
```

This gives us a sequence of pseudorandom variates from the symbolic distribution.

4.2 Simulation Results

For comparison purposes, we compute the F_1 score (F-measure), the balanced accuracy (mean of sensitivity and specificity), the AUC-values of ROC-curve (Area Under Curve value of Receiver Operating Characteristic curve) as well as the AUC-values of PR-curve (Precision Recall curve) of each method for different sliding window lengths (50, 200, 500) and for different passes in time (forward, backward or forward and backward computation). Forward computation in our context means, that the sliding time windows pass over the time series in ongoing manner from the first to the last sample, whereas in the backward computation case we start with the shifting process at the end of the time series and slide over it till we reach the beginning of the time series. For the combined forward and backward computation, we apply the union command for the found change point positions of each computation to get the set of new change points.

The just mentioned statistics generate values between 0 and 1, whereby value 1 is the best outcome. All computations are made with the help of the computer algebra system Mathematica (Version Number 10.3.1.0). To be able to compute sensitivities, specificities, recalls and precisions, we define that a change point is found correctly, if it is located in the interval $\{t_0 - 5, t_0 + 4\}$. If a change point lies outside this interval, we will detect a false positive. So, for binary classification, we count the number of intervals of length 10 (except at each edge we have length 15) that contain one or more change points (positives) versus the number of intervals that do not include a change point (negatives). The optimal case would lead to 56 true positives and 5488 (=98*56) true negatives. As the used classes are of very different sizes, it is important to have a look at indicators used in information retrieval – in our case the F_1 score and the AUC-values of the precision-recall-curve. More information on plotting ROC- and PR-curves can be found in Fawcett (2006) and Saito and Rehmsmeier (2015). In our plots we used vertical averaging for the occurrence of equal x-values and different y-values.

We start with some plots of the F_1 score and the balanced accuracy (see figure 5 till 8 in the appendix). These plots display the used parameter choices and are shown for forward and backward computation together. The darker an area is, the higher are these measures and we can find an optimal parameter setting for each method. The method BSOP reached the highest F_1 score for all different window lengths and the highest balanced accuracy for the window length 500. The method F-Test showed the highest balanced accuracy for window length 50 and 200.

Next we consider the ROC-curves and PR-curves of all methods for different window lengths. We want to mention, that the method BOCPD does not use window lengths – so these curves will remain the same (see figure 9 and 10 in the appendix). With the help of the areas under the curves, we want to find a tendency for the effectiveness of each change point detection method. Therefore, we want to refer to figure 11 and 12. If we consider the areas under the ROC-curve, the method F-Test is the best method for window length 50, whereas the method RuLSIF is for the window lengths 200 and 500 best. A different picture show the areas under PR-curve. Here, the method BSOP was best for all window lengths.

In table 2, we summarize the best methods for the different statistics and window lengths, and we also give the values of the statistics in brackets.

Table 2: Change point detection methods that reached the maximum of a special statistic for different window lengths (for time series with one change point and usage of forward and backward computation together)

	Window length 50	Window length 200	Window length 500
Max(F_1 score)	BSOP (0.66)	BSOP (0.73)	BSOP (0.78)
Max(Balanced accuracy)	F-Test (0.84)	F-Test (0.91)	BSOP (0.89)
AUC-ROC	F-Test (0.93)	RuLSIF (0.96)	RuLSIF (0.93)
AUC-PR	BSOP (0.59)	BSOP (0.70)	BSOP (0.66)

If we expand the interval length for finding an correct change point from 10 to 20, the method ICSS will provide better results, but it remains under the maximum values. Apart from that, the results keep almost the same. Only for higher window lengths, the statistical values increased, e.g. the method BSOP has now an AUC-PR value of 0.90 instead of 0.66 for window length 500 and balanced accuracy of 0.97 instead of 0.89 for the optimal parameter setting. In contrast to table 2, the method F-Test is slightly outbid by BSOP for the balanced accuracy with window length 200 and the method RuLSIF is outbid by BSOP for the AUC-ROC with window length 500.

Now we want to consider the changes in performance, if we only make a forward computation. First, we again consider the table of the best methods for different statistics and window lengths (see table 3).

Table 3: Change point detection methods that reached the maximum of a special statistic for different window lengths (for time series with one change point and usage of forward computation)

	Window length 50	Window length 200	Window length 500
Max(F_1 score)	BSOP (0.69)	BSOP (0.73)	BSOP (0.78)
Max(Balanced accuracy)	F-Test (0.84)	RuLSIF (0.91)	BSOP (0.89)
AUC-ROC	F-Test (0.93)	F-Test (0.95)	RuLSIF (0.93)
AUC-PR	BSOP (0.61)	BSOP (0.71)	BSOP (0.67)

The results of the best methods stay almost the same. Only two modifications regarding balanced accuracy and AUC-ROC for window length 200 can be considered – the methods RuLSIF and F-Test change their role. Except of this modification, also figure 11 remains very similar. Only for window length 50, method RuLSIF drops down to 0.66 points. In figure 12, BSOP is still the best method for all window lengths and everything remains equal.

Remarkable is the fact that the maxima of F_1 score and AUCPR-value of forward computation are at the same level as the maxima of forward and backward computation together (+0.012 and +0.009 respectively in the mean). Usually, one would expect much lower statistics. The maximum of balanced accuracy as well as the maximum of AUC-value of ROC-curve remain also at the same level (+0.001 and −0.005 respectively in the mean). As conclusion, only the use of forward computation must not be worse, and computation time is of course better.

Table 4: Performance indicators for forward computation and time series with one change point (note the following abbreviations: I1 = Max(F_1 score), I2 = Max(Balanced accuracy), I3 = AUC-ROC, I4 = AUC-PR, V+/- = all variance changes are considered, V+ = only variance increases are considered, V- = only variance decreases are considered)

		Window length 50				Window length 200				Window length 500			
		I1	I2	I3	I4	I1	I2	I3	I4	I1	I2	I3	I4
ICSS	V+/-	0.50	0.80	0.85	0.38	0.62	0.82	0.84	0.49	0.62	0.81	0.80	0.47
	V+	0.50	0.79	0.88	0.42	0.65	0.83	0.86	0.52	0.65	0.82	0.84	0.55
	V-	0.56	0.82	0.82	0.41	0.59	0.82	0.82	0.44	0.60	0.81	0.76	0.39
ICM	V+/-	0.04	0.60	0.60	0.02	0.05	0.61	0.62	0.02	0.05	0.61	0.62	0.02
	V+	0.04	0.61	0.61	0.02	0.05	0.63	0.61	0.02	0.04	0.58	0.56	0.02
	V-	0.04	0.61	0.59	0.02	0.04	0.63	0.62	0.02	0.07	0.67	0.67	0.03
SIC	V+/-	0.02	0.53	0.52	0.01	0.02	0.54	0.53	0.01	0.06	0.63	0.62	0.02
	V+	0.03	0.64	0.65	0.02	0.03	0.57	0.56	0.01	0.08	0.71	0.71	0.03
	V-	0.01	0.43	0.39	0.01	0.02	0.53	0.50	0.01	0.04	0.57	0.53	0.02
BSOP	V+/-	0.69	0.81	0.84	0.61	0.73	0.86	0.87	0.71	0.78	0.89	0.89	0.67
	V+	0.68	0.80	0.83	0.59	0.71	0.82	0.84	0.64	0.73	0.86	0.87	0.57
	V-	0.70	0.82	0.84	0.63	0.81	0.90	0.89	0.77	0.85	0.93	0.92	0.77
RuLSIF	V+/-	0.24	0.67	0.66	0.18	0.34	0.91	0.95	0.29	0.25	0.86	0.93	0.23
	V+	0.02	0.50	0.37	0.01	0.30	0.89	0.92	0.15	0.23	0.84	0.92	0.09
	V-	0.40	0.85	0.94	0.38	0.50	0.93	0.97	0.47	0.38	0.88	0.94	0.34
BOCPD	V+/-	0.26	0.62	0.62	0.11	0.26	0.62	0.62	0.11	0.26	0.62	0.62	0.11
	V+	0.42	0.76	0.77	0.22	0.42	0.76	0.77	0.22	0.42	0.76	0.77	0.22
	V-	0.00	0.50	0.47	0.01	0.00	0.50	0.47	0.01	0.00	0.50	0.47	0.01
HeurM.	V+/-	0.11	0.69	0.76	0.04	0.16	0.75	0.84	0.06	0.09	0.79	0.84	0.04
	V+	0.18	0.70	0.75	0.05	0.15	0.75	0.83	0.06	0.09	0.79	0.84	0.04
	V-	0.10	0.71	0.77	0.03	0.19	0.75	0.84	0.06	0.11	0.79	0.84	0.04
F-Test	V+/-	0.45	0.84	0.93	0.34	0.19	0.91	0.95	0.12	0.10	0.83	0.88	0.06
	V+	0.47	0.82	0.90	0.32	0.19	0.91	0.95	0.12	0.10	0.82	0.88	0.06
	V-	0.49	0.87	0.96	0.37	0.19	0.93	0.95	0.12	0.12	0.84	0.88	0.05

Another interesting point is the following: Are the methods equally good for variance increases and variance decreases respectively? To answer this question, we want to refer to table 4, where all performance indicators can be found for the whole dataset and also for the half of the dataset due to the splitting into variance changes that are going upwards versus the changes going downwards. The colored fields signalize for each method the better recognized kind of change. The methods ICSS, SIC and BOCPD show higher indicators for upward changes, whereas the methods ICM, BSOP, RuLSIF, HeurMeth and F-Test have higher rates for downward changes. Especially the methods RuLSIF and BOCPD display divergences that should be regarded.

Remark 1. Table 4 shows that the methods ICM and SIC intended for the correlated data sets have unsatisfactory efficiency for independent data. For correlated data the $F1$ score resulted a good segmentation quality. The next table confirms this observation for stationary processes AR(1) and MA(1) with parameter 0.8 changing to -0.8.

Table 5: Performance indicator I1 for V+

	Window length 50	Window length 200	Window length 500
MA(1) ICM	0.80	0.74	0.86
MA(1) SIC	0.65	0.76	0.91
AR(1) ICM	0.76	0.60	0.83
AR(1) SIC	0.69	0.83	0.89

5 Case Study II

5.1 Data Generation

In Case Study II, we consider time series x of length $n = 1000$ with four change points at the positions $t_1 = 256$, $t_2 = 466$, $t_3 = 536$ and $t_4 = 746$. Now the random variable X_t satisfies

$$\begin{array}{ll} X_t \sim \mathcal{N}(0, \sigma_1^2),\ 1 \leq t \leq t_1 - 1, & X_t \sim \mathcal{N}(0, \sigma_2^2),\ t_1 \leq t \leq t_2 - 1, \\ X_t \sim \mathcal{N}(0, \sigma_1^2),\ t_2 \leq t \leq t_3 - 1, & X_t \sim \mathcal{N}(0, \sigma_2^2),\ t_3 \leq t \leq t_4 - 1, \\ X_t \sim \mathcal{N}(0, \sigma_1^2),\ t_4 \leq t \leq n. & \end{array}$$

The set E remains equal to Case Study I, so we again have 56 time series. Figure 2 shows an exemplary time series for $\sigma_1 = 1$ and $\sigma_2 = 4$ with four change points that are marked through vertical lines.

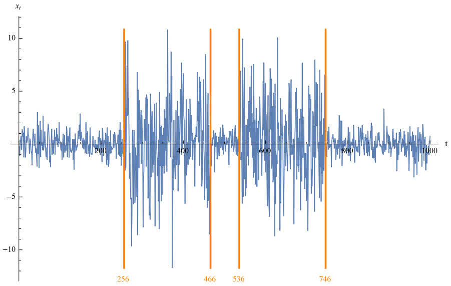


Figure 2: Simulated time series with four change points ($\sigma_1 = 1$ and $\sigma_2 = 4$)

5.2 Simulation Results

We again use the same statistics for comparison of the different methods. As we have time series with more than one change point now, we also want to additionally implement a variance stabilization method for the methods HeurMeth and F-Test, and we will denote these methods with HeurMeth$_{\mathrm{St}}$ and F-Test$_{\mathrm{St}}$. As soon as a change point at time t_0 is found, the data before time t_0 is transformed according to the formula

$$x_t^* = \begin{cases} \frac{1}{t_0-1}\sum_{t=1}^{t_0-1} x_t + \sqrt{F(t_0)}\Big(x_t - \frac{1}{t_0-1}\sum_{t=1}^{t_0-1} x_t\Big) & t < t_0, \\ x_t, & t \geq t_0, \end{cases} \tag{16}$$

where $F(t_0)$ is defined as before as the variance of the later time window divided by the variance of the foregoing time window. With this transformation, the shift in variance at point t_0 is eliminated and the change point detector can proceed further on the transformed time series.

In this section, the definition of a correctly found change point remains equal (interval of length 10 around each change point) and we now have 224 (=56*4) true positives versus 5376 (=56*96) true negatives in the optimal classification case. The parameter selections of all methods stay the same and the computation of all statistics is conducted in an analogous manner. At this point, we consider once again the table with the best change point detection methods concerning our chosen statistics and window lengths in table 6.

Table 6: Change point detection methods that reached the maximum of a special statistic for different window lengths (for time series with several change points and usage of forward and backward computation together)

	Window length 50	Window length 200	Window length 500
Max(F_1 score)	F-Test$_{\mathrm{St}}$ (0.48)	BSOP (0.48)	BSOP (0.55)
Max(Balanced accuracy)	F-Test (0.85)	BSOP (0.85)	BSOP (0.86)
AUC-ROC	F-Test (0.89)	ICSS (0.87)	BSOP (0.86)
AUC-PR	F-Test$_{\mathrm{St}}$ (0.40)	ICSS (0.35)	BSOP (0.39)

For small window lengths, the methods F-Test and F-Test$_{\mathrm{St}}$ showed the best performance. For window length 200, ICSS and BSOP were best. For longer window lengths, BSOP outperformed the other methods. In the mean, all statistics lost on height in comparison to time series with one change point and the usage of forward and backward computation as here (mean decreases for the four statistics sorted as in the first column of table 6: −0.045, −0.056, −0.062 and −0.039).

In figure 13 and 14, the ROC-curves and PR-curves for time series with four change points are shown. Figure 15 and 16 contain the AUC-values for ROC-curve and PR-curve respectively.

Table 7 presents the best methods only for forward computations.

All statistics in table 7 are higher than the statistics in table 6. So, in the case of several change points the sole use of forward computation is more advantageous regarding to the

Table 7: Change point detection methods that reached the maximum of a special statistic for different window lengths (for time series with several change points and usage of forward computation)

	Window length 50	Window length 200	Window length 500
Max(F_1 score)	BSOP (0.70)	BSOP (0.77)	BSOP (0.78)
Max(Balanced accuracy)	ICSS (0.85)	BSOP (0.87)	BSOP (0.87)
AUC-ROC	F-Test (0.89)	ICSS (0.88)	BSOP (0.87)
AUC-PR	BSOP (0.69)	BSOP (0.73)	BSOP (0.73)

optimal parameter settings. If we consider the mean deviation between the F_1 scores and AUC-PR values concerning forward computation against forward and backward computation, the values increased about 0.032 and 0.061 respectively. Nevertheless, the mean of balanced accuracy and AUC-ROC decreased by 0.010 and 0.014. So, regarding to the mean deviations, the statistical values remained equal.

We now have again a look at the performance indicators for the case of forward computation in table 8. Here, the methods ICSS, ICM, SIC, BSOP, BOCPD and F-Test display higher indicators for upward changes, whereas the methods RuLSIF and F-Test$_{\mathrm{St}}$ for downward changes. The methods HeurMeth and HeurMeth$_{\mathrm{St}}$ are equally good for both kinds of change. In the case of time series with four change points, more methods are better for upward changes – as ICM, BSOP and HeurMeth changed their manner. The methods RuLSIF and BOCPD show again high divergences between the statistics of upward and downward changes.

With an additional application of a variance stabilization method, the methods HeurMeth and F-Test could not be further enhanced. One possible reason for that outcome is that some change points were found to early for our defined interval of being a correct change point.

6 Case Study III

6.1 Data Generation

Let $x = \{x_1, \dots, x_n\}$ be again a given time series of length n, where x_t is a realization of the random variable X_t. In this study, 1050 time series of length $n = 600$ and one change point at position $t_0 = 301$ have been generated, where

$$X_t \sim \mathcal{N}(0, \sigma_1^2),\ 1 \leq t \leq t_0 - 1, \quad X_t \sim \mathcal{N}(0, \sigma_2^2),\ t_0 \leq t \leq n.$$

Let $E = \{1, 2, 3, \dots, 13, 14, 15\}$ be a set of values. Then, the standard deviations satisfy $\sigma_1 \in E$ and $\sigma_2 \in E \setminus \{\sigma_1\}$. With all possible combinations, we get $|E|(|E| - 1) = 210$ time series with different changes in variance. We apply this process five times such that in the end we get 1050 time series. Additionally, we generate 1050 time series with no change point. Therefore, $X_t \sim \mathcal{N}(0, \sigma_1^2), 1 \leq t \leq n$ and the set E remains equal. One pass over E generates 15 time series, so we have to apply this process 70 times to get 1050 time series.

Table 8: Performance indicators for forward computation and time series with several change points (note the following abbreviations: I1 = Max(F_1 score), I2 = Max(Balanced accuracy), I3 = AUC-ROC, I4 = AUC-PR, V+/- = all variance changes are considered, V+ = only variance increases are considered, V- = only variance decreases are considered)

		Window length 50				Window length 200				Window length 500			
		I1	I2	I3	I4	I1	I2	I3	I4	I1	I2	I3	I4
ICSS	V+/-	0.68	0.85	0.89	0.68	0.66	0.86	0.88	0.68	0.59	0.84	0.87	0.57
	V+	0.71	0.87	0.91	0.74	0.69	0.88	0.89	0.72	0.62	0.84	0.88	0.62
	V-	0.64	0.84	0.87	0.63	0.63	0.84	0.87	0.63	0.57	0.84	0.85	0.52
ICM	V+/-	0.13	0.60	0.59	0.06	0.12	0.57	0.57	0.06	0.14	0.62	0.62	0.07
	V+	0.19	0.69	0.69	0.09	0.15	0.61	0.60	0.07	0.22	0.70	0.72	0.12
	V-	0.07	0.50	0.48	0.04	0.09	0.54	0.54	0.05	0.09	0.53	0.51	0.04
SIC	V+/-	0.10	0.59	0.59	0.05	0.12	0.61	0.61	0.06	0.14	0.59	0.58	0.07
	V+	0.12	0.66	0.66	0.06	0.13	0.62	0.61	0.06	0.16	0.61	0.60	0.08
	V-	0.08	0.53	0.52	0.04	0.14	0.62	0.60	0.06	0.14	0.58	0.58	0.07
BSOP	V+/-	0.70	0.83	0.87	0.69	0.77	0.87	0.87	0.73	0.78	0.87	0.87	0.73
	V+	0.72	0.84	0.89	0.70	0.82	0.89	0.88	0.78	0.84	0.90	0.89	0.80
	V-	0.68	0.82	0.85	0.67	0.73	0.85	0.86	0.67	0.72	0.85	0.86	0.68
RuLSIF	V+/-	0.32	0.66	0.64	0.24	0.26	0.77	0.84	0.21	0.37	0.67	0.69	0.30
	V+	0.05	0.50	0.39	0.03	0.25	0.76	0.82	0.14	0.33	0.67	0.70	0.27
	V-	0.49	0.82	0.90	0.47	0.33	0.77	0.85	0.28	0.41	0.68	0.68	0.34
BOCPD	V+/-	0.09	0.53	0.49	0.08	0.09	0.53	0.49	0.08	0.09	0.53	0.49	0.08
	V+	0.17	0.56	0.53	0.14	0.17	0.56	0.53	0.14	0.17	0.56	0.53	0.14
	V-	0.00	0.50	0.46	0.04	0.00	0.50	0.46	0.04	0.00	0.50	0.46	0.04
HeurMeth	V+/-	0.18	0.67	0.74	0.12	0.16	0.68	0.74	0.09	0.14	0.71	0.72	0.10
	V+	0.18	0.67	0.74	0.11	0.15	0.68	0.74	0.09	0.15	0.72	0.72	0.09
	V-	0.19	0.67	0.74	0.13	0.16	0.68	0.74	0.10	0.14	0.71	0.71	0.11
HeurM.$_{\mathrm{St}}$	V+/-	0.17	0.67	0.72	0.12	0.12	0.63	0.66	0.07	0.21	0.71	0.74	0.16
	V+	0.18	0.68	0.74	0.12	0.15	0.66	0.70	0.09	0.21	0.71	0.74	0.15
	V-	0.17	0.66	0.70	0.12	0.11	0.62	0.62	0.06	0.23	0.72	0.74	0.19
F-Test	V+/-	0.46	0.85	0.89	0.37	0.18	0.74	0.80	0.12	0.24	0.76	0.80	0.18
	V+	0.47	0.85	0.90	0.38	0.18	0.74	0.80	0.12	0.25	0.76	0.80	0.21
	V-	0.46	0.85	0.88	0.36	0.18	0.74	0.79	0.12	0.23	0.76	0.80	0.16
F-Test$_{\mathrm{St}}$	V+/-	0.51	0.80	0.86	0.41	0.24	0.67	0.73	0.16	0.33	0.76	0.83	0.28
	V+	0.40	0.79	0.86	0.27	0.20	0.63	0.69	0.09	0.34	0.76	0.83	0.35
	V-	0.63	0.82	0.86	0.61	0.32	0.71	0.77	0.29	0.31	0.76	0.82	0.24

6.2 Simulation Results

In this simulation study, we used for each method only the optimal parameters concerning F_1 score and balanced accuracy as found in figures 5 to 8. This allows us to consider a larger amount of time series for simulation study. Our goal is to find the relative number of correctly found change points out of the piecewise stationary time series that have one change point as well as the relative numbers of incorrectly found change points out of the stationary time series with no change point. For finding a correct change point, we again use the interval $\{t_0 - 5, t_0 + 4\}$ around the real change point time t_0. If there are several change points in this interval, we count only one true positive. For the false positives, we also use intervals of length 10, so the stationary time series is split into 60 intervals. Table 9 gives a full overview of our evaluations. Here, we have restricted us to sliding time windows of length 50 and 200.

Table 9: True versus false positives (relative values in percent) based on 1050 time series with one versus no change point (note the following abbreviations: OP(...) = Optimal parameters used as found in Case Study I for forward and backward computation regarding to the statistic ..., TP = true positives, FP = false positives, F = only forward computation, B = only backward computation, F&B = forward and backward computation together)

		Window length 50				Window length 200			
		OP(F_1 score)		OP(Bal. Accu.)		OP(F_1 score)		OP(Bal. Accu.)	
		TP	FP	TP	FP	TP	FP	TP	FP
ICSS	F	29.143	0.024	58.381	0.006	47.619	0.014	57.810	0.005
	B	29.143	0.002	58.667	6.487	47.619	0.006	57.524	2.303
	F&B	29.143	0.003	59.238	12.041	47.619	0.011	58.762	4.003
ICM	F	25.905	13.251	36.952	24.489	22.190	9.197	28.667	13.216
	B	26.571	13.256	36.476	23.992	23.619	9.537	31.619	13.529
	F&B	44.571	24.303	59.048	41.373	40.190	17.984	50.762	25.384
SIC	F	58.381	51.508	58.381	51.508	36.190	24.644	36.190	24.644
	B	55.714	54.702	55.714	54.702	37.905	26.756	37.905	26.756
	F&B	78.381	74.802	78.381	74.802	57.238	42.446	57.238	42.446
BSOP	F	42.190	0.154	45.333	0.567	58.571	0.079	59.619	0.254
	B	42.095	0.154	45.333	0.576	58.667	0.084	59.714	0.263
	F&B	42.190	0.308	45.429	1.133	58.762	0.138	59.810	0.441
RuLSIF	F	18.286	0.229	39.810	10.397	22.952	0.000	55.238	0.038
	B	19.905	0.244	53.619	10.165	29.714	0.000	58.381	0.032
	F&B	33.429	0.425	69.905	24.341	35.714	0.000	78.381	2.106
BOCPD	F	8.952	0.013	10.381	0.001	8.952	0.013	10.381	0.001
	B	9.333	0.011	11.143	0.003	9.333	0.011	11.143	0.003
	F&B	18.286	0.024	21.524	0.005	18.286	0.024	21.524	0.005
HeurMeth	F	16.762	2.678	72.952	35.884	10.667	0.029	69.238	13.843
	B	17.429	2.683	74.476	35.994	10.667	0.029	69.238	13.808
	F&B	18.857	3.546	76.000	39.665	10.667	0.029	69.238	13.843
F-Test	F	48.667	0.110	82.571	24.171	45.238	0.000	74.952	0.125
	B	48.667	0.111	82.571	24.267	45.238	0.000	74.952	0.133
	F&B	48.667	0.110	82.571	24.171	45.238	0.000	74.952	0.125

To get a better overview about the numbers in table 9, we also plot the F_1 scores for the parameter settings obtained from Case Study I through the use of forward and backward computations. So, the values of the columns with heading OP(F_1 score) lead to the figures 3 and 4. These plots show similar results as in Case Study I and II.

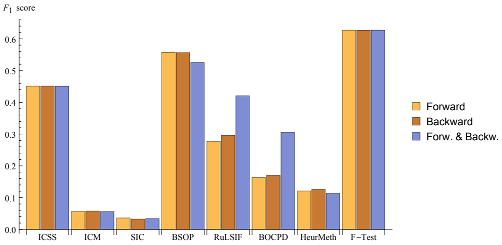


Figure 3: F_1 scores for different change point detection methods and window-length=50

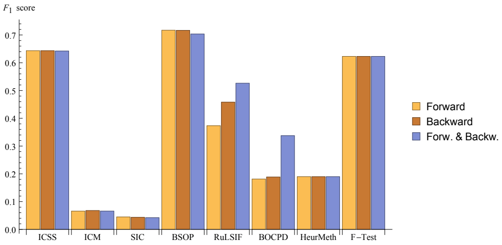


Figure 4: F_1 scores for different change point detection methods and window-length=200

Again the methods ICSS, BSOP and F-Test demonstrate their superiority over the other methods. Furthermore, for the methods RuLSIF and BOCPD the use of forward and backward computation is more advantageous, whereas for the other methods passage in one direction is sufficient.

7 Conclusion

In this study, online as well as typically retrospective methods are compared with the help of sliding time windows in an online manner for data streams.

The method BSOP[1] showed the best performance regarding to the statistics F_1 score and AUC of PR-curve, followed by the methods ICSS and F-Test. The methods ICM, SIC and HeurMeth were not that good in our simulation. Regarding to the statistics balanced accuracy and AUC of ROC-curve, the methods ICSS, BSOP, RuLSIF and F-Test displayed a good performance whereas the methods ICM, SIC and BOCPD were not that convincing. The inferior performance of ICM and SIC method was somehow predetermined, as they are usually used for dependent observations.

The sole application of forward computation (concerning the sliding time windows) is for the most methods as recommendable as the use of forward and backward computation together, as there was no decrease in performance. All methods behaved similar concerning their performance for the different sliding time window lengths. Two methods (RuLSIF and BOCPD) showed differences in performance for upward versus downward

[1] More information about the different change point detection methods can be found in section 3.

changes in variance. The additional application of a variance stabilization method to the methods HeurMeth and F-Test did not show major improvements.

Applications to real datasets can be found in our previous publications (Breitenberger et al., 2015, 2017). There, we used the heuristic method HeurMeth with additional constraints. The reason for using this method lies in the fast execution time and in the fact, that this method finds many change points that can be further restricted. Our further research might compare exhaustively the methods for the correlated data sets.

Acknowledgements

This work has been supported by the Linz Center of Mechatronics (LCM) in the framework of the Austrian COMET-K2 programme and by Smartbow GmbH, by the Austrian Research Promotion Agency (FFG), Project No. 848610 and by the Ministry of Education of the Russian Federation (the Agreement number 02.a03.21.0008 of 24 June 2016).

References

Badagián, A. (2015). *Time series segmentation procedures to detect, locate and estimate change points.* Springer.

Badagián, A., Kaiser, R., and Pena, D. (2009). Time series segmentation by cusum, autoslex and autoparm methods. *Statistics and Econometrics Series*, 25:9–25.

Breitenberger, S., Efrosinin, D., Auer, W., Deininger, A., and Waßmuth, R. (2015). Automatisierte erkennung der trinkmengen und trinkphasen bei klbern anhand mittels ohrmarken erfasster beschleunigungsdaten. In *12. Int. Tagung: Bau, Technik und Umwelt in der landwirtschaftlichen Nutztierhaltung.*

Breitenberger, S., Efrosinin, D., Auer, W., Deininger, A., and Waßmuth, R. (2017). Change point detection in piecewise stationary time series for farm animal behavior analysis. In *Operations Research Proceedings 2015.* Springer.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.

Ibrahim, A. A., Ahmed, M., and Buhamra, S. (2003). *Chapter testing for multiple change point in an autoregressive model using SIC criterion.* Nova Publishers.

Inclán, C. and Tiao, G. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 427:913–923.

Liu, S., Yamada, M., Collier, N., and Sugiyama, M. (2013). Change-point detection in time-series data by relative density-ratio estimation. *arXiv 1203.0453.*

Rohrbeck, C. (2013). Detection of changes in variance using binary segmentation and optimal partitioning. *http://www.lancaster.ac.uk/pg/rohrbeck/ResearchTopicI.pdf.*

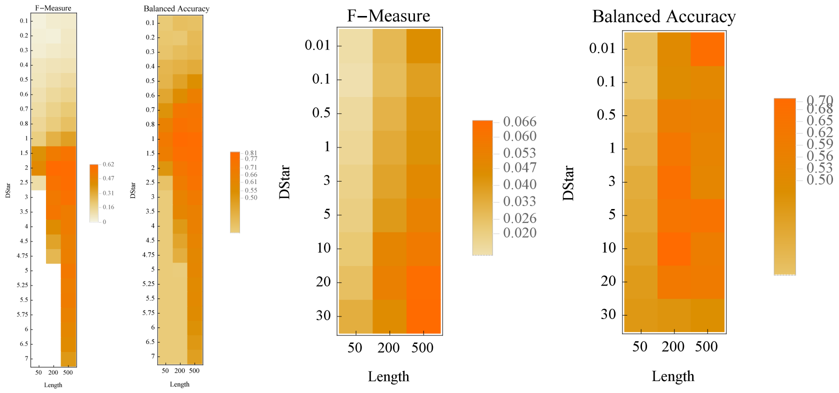
Ryan, P., David, J., and MacKay, C. (2007). Bayesian online changepoint detection. *arXiv 0710.3742.*

Saito, T. and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, 10(3).

Tsay, R. (1988). Outliers, level shifts and variance changes in time series. *Journal of Forcasting*, 7:1–20.

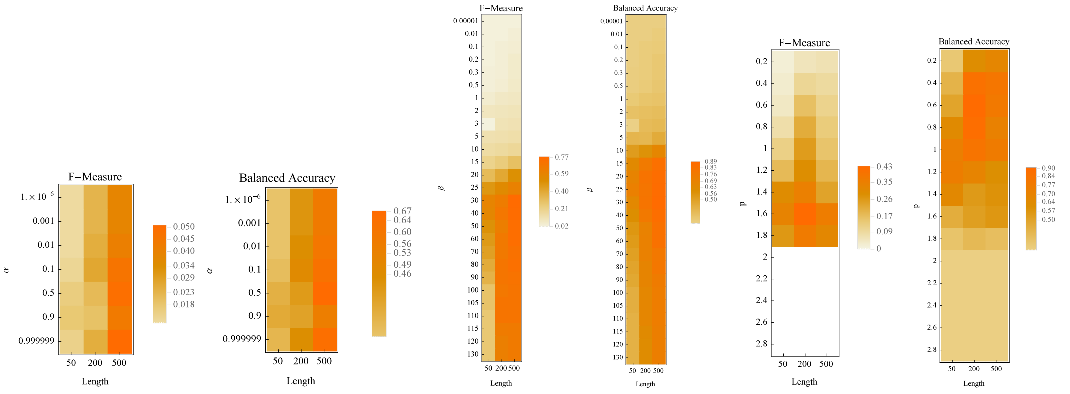
Ureche-Rangau, L. and Speeg, F. (2011). A simple method of variance shift detection at unknown time points. *Economics Bulletin*, 31(3):2204–2218.

Appendix



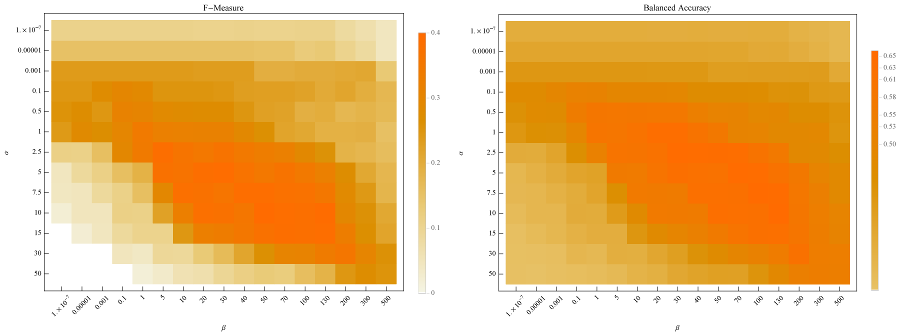
(a) F_1 score of ICSS (b) Bal. accuracy of ICSS (c) F_1 score of ICM (d) Bal. accuracy of ICM

Figure 5: F_1 scores and balanced accuracies for ICSS and ICM method (for time series with one change point and usage of forward and backward computation together)



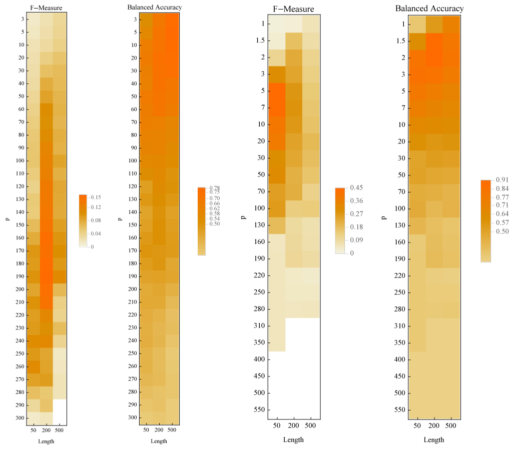
(a) F_1 score of SIC (b) Bal. accuracy of SIC (c) F_1 score of BSOP (d) Bal. accuracy of BSOP (e) F_1 score of RuLSIF (f) Bal. accuracy of RuLSIF

Figure 6: F_1 scores and balanced accuracies for SIC, BSOP and RuLSIF method (method RuLSIF with $k = 10$ and $\alpha = 0.2$ fixed; for time series with one change point and usage of forward and backward computation together)



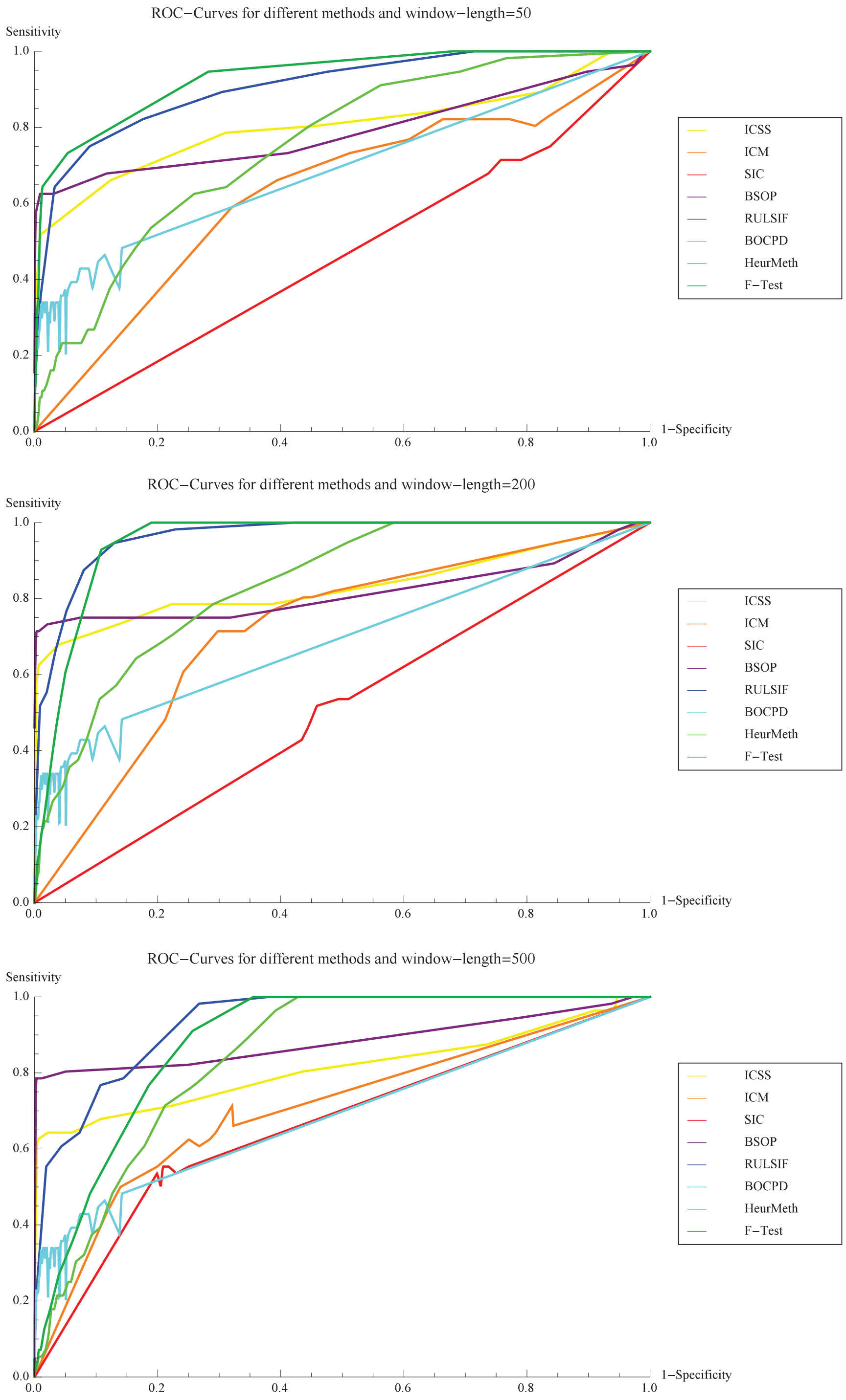
(a) F_1 score of BOCPD (b) Bal. accuracy of BOCPD

Figure 7: F_1 scores and balanced accuracies for BOCPD method with $\lambda = 500$ and $\kappa = 1$; plots for $\kappa = 0.0001$ are not shown here (for time series with one change point and usage of forward and backward computation together)



(a) F_1 score of HeurMeth (b) Bal. accuracy of HeurMeth (c) F_1 score of F-Test (d) Bal. accuracy of F-Test

Figure 8: F_1 scores and balanced accuracies for HeurMeth and F-Test method (for time series with one change point and usage of forward and backward computation together)



ROC−Curves for different methods and window−length=50

ROC−Curves for different methods and window−length=200

ROC−Curves for different methods and window−length=500

Figure 9: ROC-curves for different methods and different window lengths (for time series with one change point and usage of forward and backward computation together)

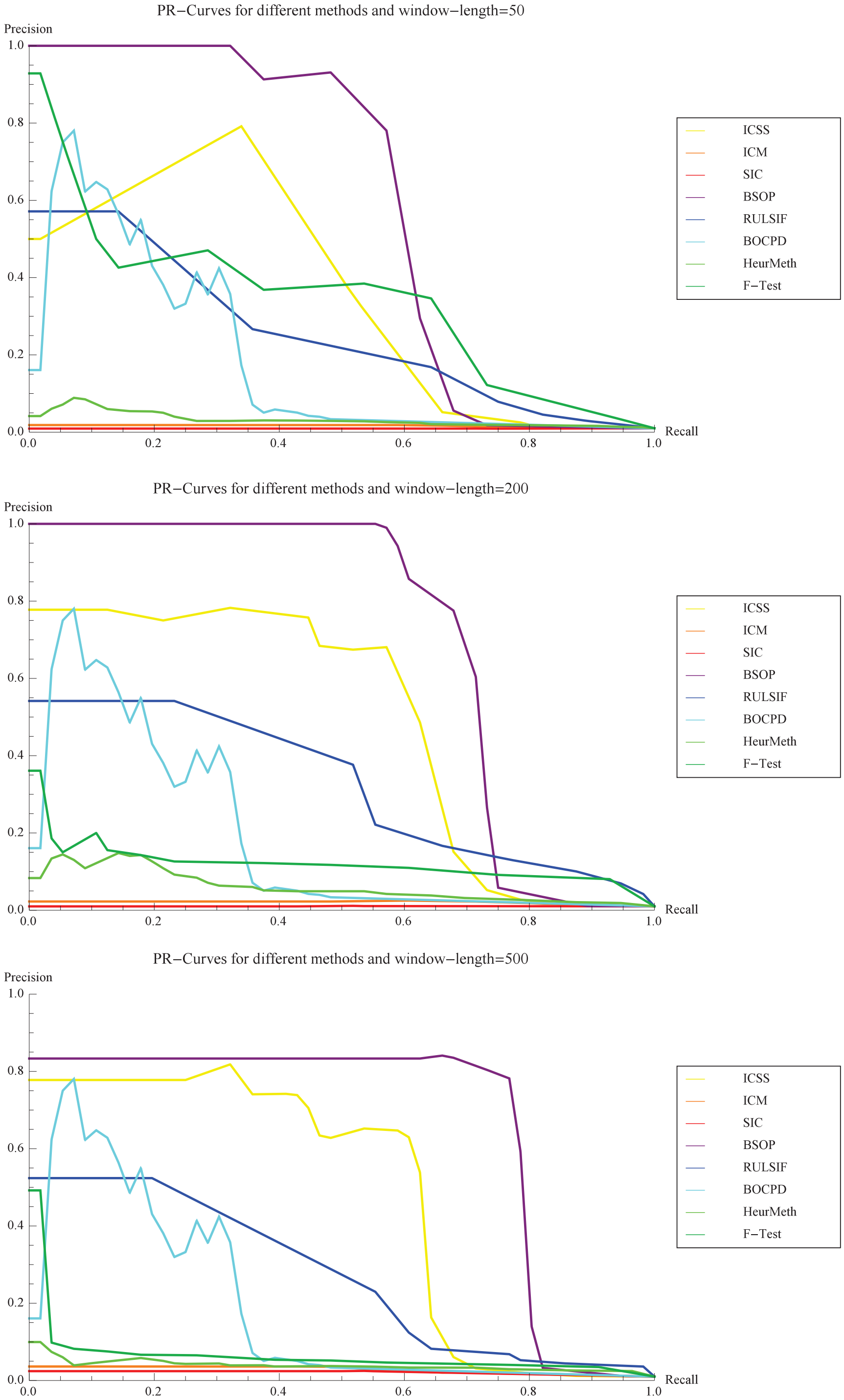


Figure 10: PR-curves for different methods and different window lengths (for time series with one change point and usage of forward and backward computation together)

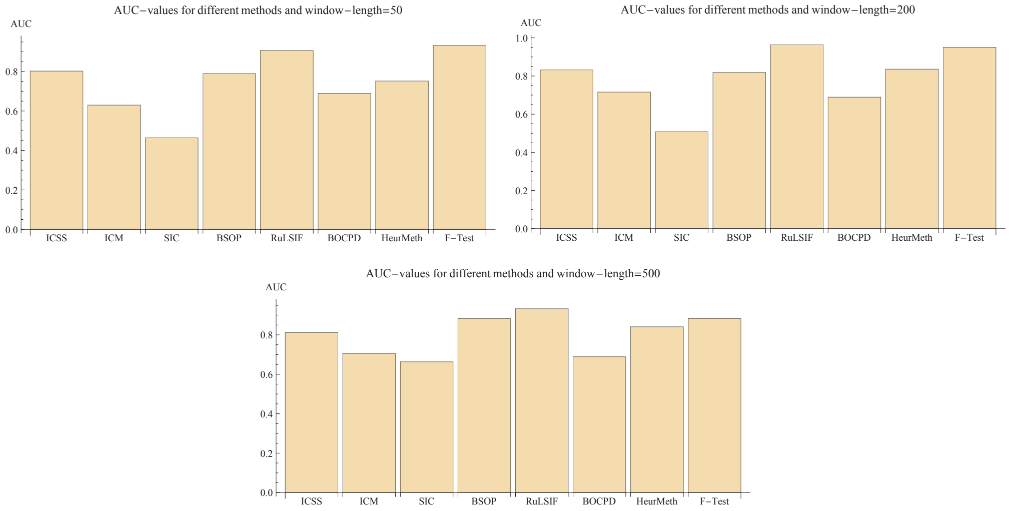


Figure 11: AUC-values of ROC-curve for different methods and different window lengths (for time series with one change point and usage of forward and backward computation together)

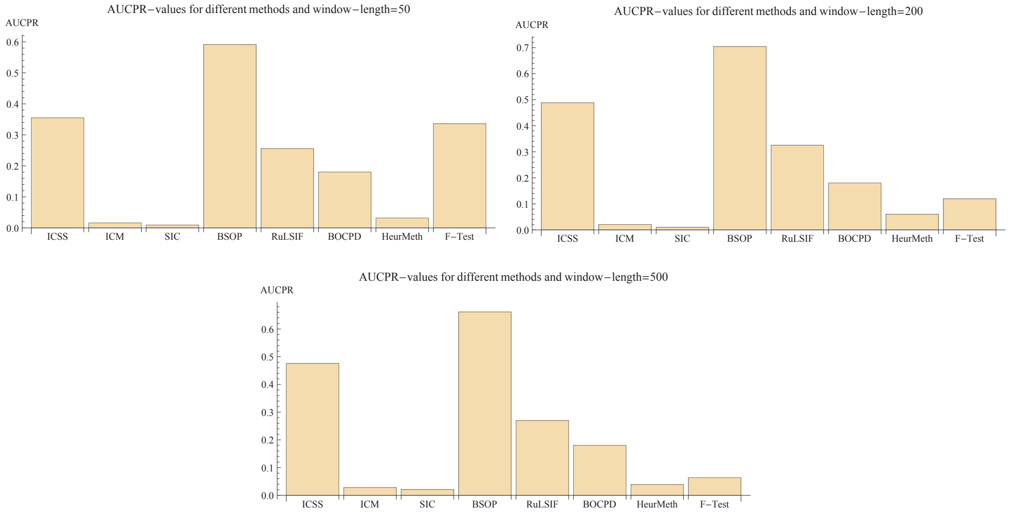


Figure 12: AUC-values of PR-curve for different methods and different window lengths (for time series with one change point and usage of forward and backward computation together)

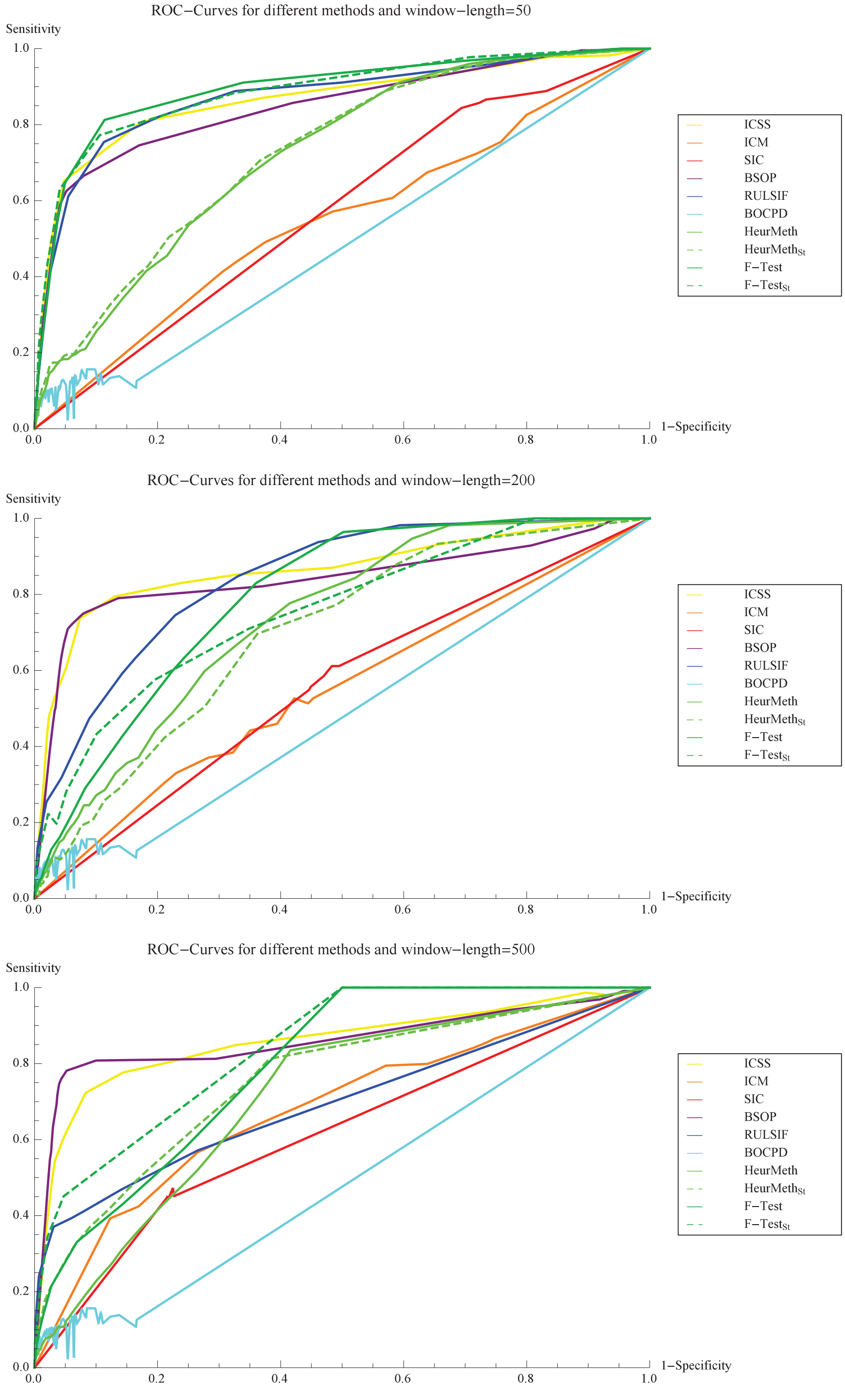


Figure 13: ROC-curves for different methods and different window lengths (for time series with several change points and usage of forward and backward computation together)

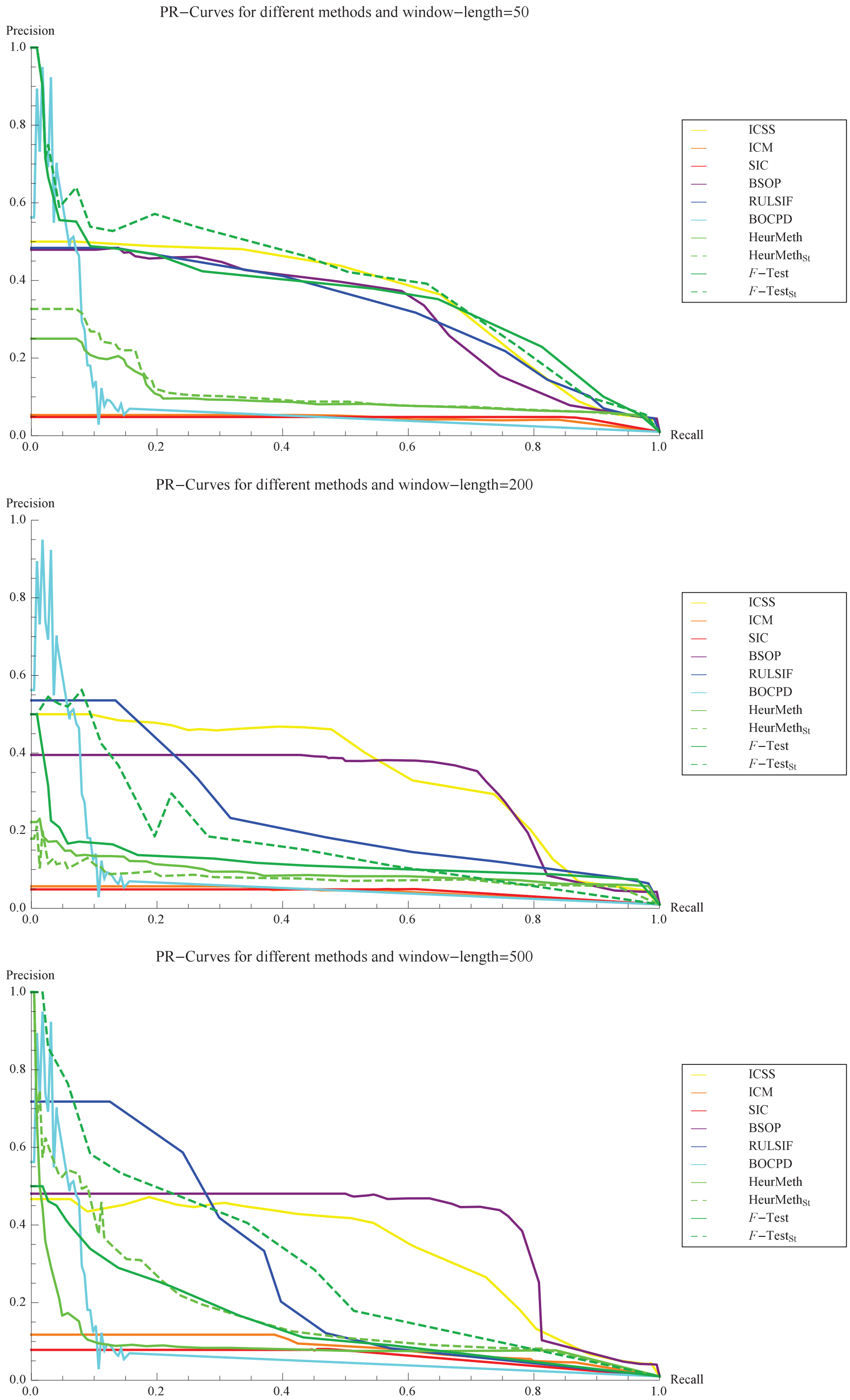


Figure 14: PR-curves for different methods and different window lengths (for time series with several change points and usage of forward and backward computation together)

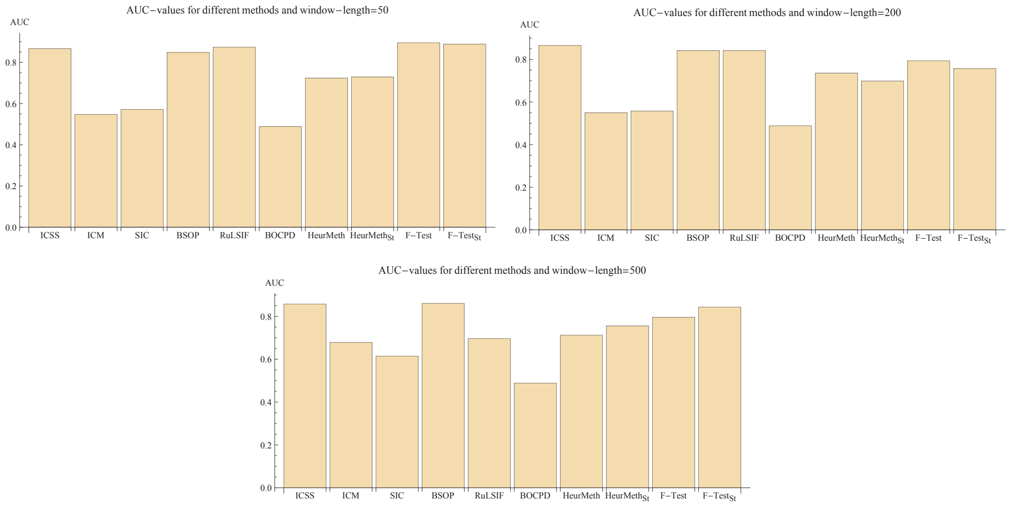


Figure 15: AUC-values of ROC-curve for different methods and different window lengths (for time series with several change points and usage of forward and backward computation together)

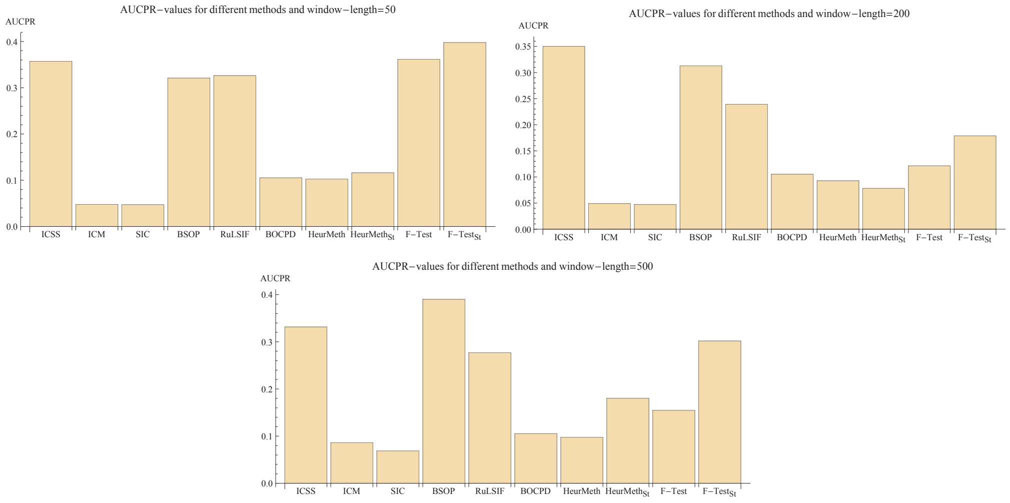


Figure 16: AUC-values of PR-curve for different methods and different window lengths (for time series with several change points and usage of forward and backward computation together)