Data-adaptive structural change-point detection via isolation

Andreas Anastasiou[1] and Sophia Loizidou[2*]

[1]Department of Mathematics and Statistics, University of Cyprus.
[2]Department of Mathematics, University of Luxembourg.

*Corresponding author(s). E-mail(s): sophia.loizidou@uni.lu;
Contributing authors: anastasiou.andreas@ucy.ac.cy;

Abstract

In this supplement we provide further simulation results, an investigation of the impact of the values of the expansion parameter, $\boldsymbol{\lambda_T}$, and the threshold constant in the performance of the algorithm, as well as the step-by-step proof of Theorem 2, which shows the consistency of our method in accurately estimating the true number and the locations of the change-points in the case that the underlying signal $\boldsymbol{f_t}$ is continuous, piecewise-linear.

1 Further simulation results

Some further simulation results are presented in Tables 1 and 2 for piecewise-constant signals and Tables 3 and 4 for continuous, piecewise-linear signals. The signals used are:

(S12) *justnoise*: sequence of length 6000 with no change-points. The standard deviation is $\sigma = 1$.

(S13) *long_signal*: sequence of length 11000 with 1 change-point at 5500 with values before and after the change-point 0 and 1.5. The standard deviation is $\sigma = 1$.

(S14) *small_dist*3: sequence of length 1000 with 6 change-points at 100, 130, 485, 515, 870, 900 with values between change-points 0, 1.5, 0, 1, 0, 1.5, 0. The standard deviation is $\sigma = 1$.

(S15) *teeth*: piecewise-constant signal of length 270 with 13 change-points at 11, 31, 51, 71, 91, 111, 131, 151, 171, 191, 211, 231, 251 with values between change-points 0, 1, 0, 1, …, 0, 1. The standard deviation is $\sigma = 0.4$.

(S16) *justnoise_wave*: piecewise-linear signal without change-points of length 1000. The starting intercept is $f_1 = 0$ and slope $f_2 - f_1 = 1$. The standard deviation is $\sigma = 1$.

(S17) *wave*4: continuous, piecewise-linear signal of length 200 with 9 change-points at 20, 40, ..., 180 with changes in slope 1/6, 1/2, -3/4, -1/3, -2/3, 1, 1/4, 3/4, -5/4. The starting intercept is $f_1 = -1$ and slope $f_2 - f_1 = 1/32$. The standard deviation is $\sigma = 0.3$.

(S18) *wave*5: continuous, piecewise-linear signal of length 350 with 50 change-points at 7, 14, ..., 343 with changes in slope -2.5, 2.5, ..., 2.5, -2.5. The starting intercept is $f_1 = -0$ and slope $f_2 - f_1 = 1$. The standard deviation is $\sigma = 1$.

The general conclusions form the tables are that DAIS performs at least as well as the competitors, often having an advantage in computational time and accuracy compared to the best performing competitors. Signal (S12), in Table 1 has no change-points. Most algorithms have very good results with negligible differences in the accuracy in the number of change-points detected. Similar conclusions hold for signal (S13), in which the large length of the signal forces some algorithms, especially MSCP, to be extremely slow. Moreover, comparing DAIS with ID_th which use the same value for the expansion parameter λ_T, it is evident that DAIS requires a smaller number of expansions and thus is faster. Signal (S14) in Table 2 indicates that in this difficult structure, where pairs of change-points are close to each other and move towards opposite directions such that they 'cancel' each-other out, ID_th has the best performance, followed by DAIS and MOSUM, in terms of accuracy with respect to the number of change-points detected and the MSE. For the signal (S15) all algorithms perform similarly. Regarding continuous, piecewise-linear signals, the signal (S16) in Table 3 has constant slope and no change-points. In this signal, all algorithms perform very well, apart from the MARS, TF, and local_poly methods, which seem to exhibit some overdetection issues. Among the top performing algorithms, DAIS and ID_ic have the lowest computational time. Regarding the performance of the competing methods in Signals (S17) and (S18) as shown in Table 4, DAIS and ID (combined with either the thresholding or the information criterion stopping rule) exhibit very good performance in both signals in terms of accuracy for both the estimated number and the estimated locations of change-points. CPOP has slightly worse (but still quite accurate) performance in both signals. NOT and TS are among the top performing methods in (S17), but not in (S18).

Table 1 Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signals (S12) and (S13). The average MSE and computational times are also given.

Method	Signal	$\hat{N} - N$					MSE	d_H	Time (s)
		-1	0	1	2	≥ 3			
DAIS	(S12)	-	**99**	0	1	0	1.87×10^{-4}	-	0.237
ID_th		-	92	3	5	0	3.59×10^{-4}	-	0.458
ID_ic		-	**100**	0	0	0	$\mathbf{1.39 \times 10^{-4}}$	-	0.193
WBS_th		-	91	8	1	0	2.15×10^{-4}	-	0.050
WBS_ic		-	**100**	0	0	0	$\mathbf{1.39 \times 10^{-4}}$	-	0.073
WBS2		-	91	5	3	1	2.75×10^{-4}	-	3.570
PELT		-	**100**	0	0	0	$\mathbf{1.39 \times 10^{-4}}$	-	0.003
NOT		-	**100**	0	0	0	$\mathbf{1.39 \times 10^{-4}}$	-	0.085
MOSUM		-	94	2	4	0	2.96×10^{-4}	-	0.028
MSCP		-	**100**	0	0	0	$\mathbf{1.39 \times 10^{-4}}$	-	83.25
SeedBS_th		-	88	11	1	0	1.59×10^{-4}	-	0.041
SeedBS_ic		-	**99**	1	0	0	1.68×10^{-4}	-	0.041
DP_univar		-	0	0	0	100	372×10^{-4}	-	30.70
SMUCE		-	78	22	0	0	3.60×10^{-4}	-	0.070
MsFPOP		-	**97**	3	0	0	2.14×10^{-4}	-	0.039
DAIS	(S13)	0	**99**	0	1	0	$\mathbf{4.33 \times 10^{-4}}$	43.10×10^{-4}	0.322
ID_th		0	93	2	5	0	5.61×10^{-4}	135.00×10^{-4}	0.471
ID_ic		0	**100**	0	0	0	$\mathbf{4.37 \times 10^{-4}}$	$\mathbf{1.15 \times 10^{-4}}$	0.391
WBS_th		0	**98**	2	0	0	$\mathbf{4.33 \times 10^{-4}}$	5.74×10^{-4}	0.083
WBS_ic		0	**100**	0	0	0	$\mathbf{4.19 \times 10^{-4}}$	$\mathbf{1.06 \times 10^{-4}}$	0.118
WBS2		0	92	3	4	1	5.46×10^{-4}	176.00×10^{-4}	6.380
PELT		0	**100**	0	0	0	$\mathbf{4.19 \times 10^{-4}}$	$\mathbf{1.06 \times 10^{-4}}$	0.004
NOT		0	**100**	0	0	0	$\mathbf{4.19 \times 10^{-4}}$	$\mathbf{1.06 \times 10^{-4}}$	0.127
MOSUM		0	92	2	6	0	5.49×10^{-4}	203.60×10^{-4}	0.049
MSCP		6	93	1	0	0	341.35×10^{-4}	334.74×10^{-4}	276.294
SeedBS_th		0	90	10	0	0	$\mathbf{4.30 \times 10^{-4}}$	223.00×10^{-4}	0.052
SeedBS_ic		0	**100**	0	0	0	$\mathbf{4.19 \times 10^{-4}}$	$\mathbf{1.06 \times 10^{-4}}$	0.052
DP_univar		0	0	0	0	100	368×10^{-4}	4840.00×10^{-4}	190.80
SMUCE		0	90	7	3	0	5.08×10^{-4}	269.00×10^{-4}	0.117
MsFPOP		0	**99**	1	0	0	$\mathbf{4.32 \times 10^{-4}}$	23.20×10^{-4}	0.068

Table 2 Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signals (S14) and (S15). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$									
Method	Signal	≤ -3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS	(S14)	0	15	3	80	0	2	0	0.0321	0.067	0.004
ID_th		0	2	2	**86**	5	4	1	**0.0303**	**0.031**	0.005
ID_ic		2	35	0	63	0	0	0	0.0358	0.145	0.016
WBS_th		0	4	7	71	14	3	1	**0.0307**	0.049	0.009
WBS_ic		2	33	1	64	0	0	0	0.0337	0.140	0.029
WBS2		0	6	5	70	9	6	4	**0.0306**	0.048	0.504
PELT		26	65	0	9	0	0	0	0.0596	0.437	0.001
NOT		2	33	1	64	0	0	0	0.0332	0.139	0.043
MOSUM		0	13	2	79	3	3	0	**0.0287**	0.063	0.015
MSCP		95	5	0	0	0	0	0	0.1390	0.593	2.610
SeedBS_th		0	4	23	50	15	7	1	0.0338	0.049	0.019
SeedBS_ic		1	29	2	63	3	1	1	0.0331	0.122	0.019
DP_univar		0	0	0	2	6	4	88	0.0602	0.126	0.139
SMUCE		0	8	39	48	5	0	0	0.0405	**0.034**	0.005
MsFPOP		2	34	1	63	0	0	0	0.0332	0.143	0.004
DAIS	(S15)	0	0	1	**94**	4	1	0	0.0272	0.010	0.002
ID_th		0	0	2	88	10	0	0	0.0270	0.012	0.003
ID_ic		0	0	0	86	11	3	0	0.0272	0.011	0.008
WBS_th		0	0	1	84	14	1	0	**0.0231**	0.011	0.004
WBS_ic		0	0	0	**93**	7	0	0	**0.0229**	**0.008**	0.010
WBS2		0	0	1	**96**	3	0	0	**0.0231**	**0.008**	0.096
PELT		1	6	3	90	0	0	0	0.0263	0.016	0.001
NOT		0	0	0	**94**	6	0	0	**0.0242**	**0.008**	0.018
MOSUM		0	0	0	**95**	3	2	0	**0.0226**	**0.008**	0.008
MSCP		100	0	0	0	0	0	0	0.2110	0.261	0.242
SeedBS_th		0	2	9	77	10	2	0	0.0283	0.018	0.026
SeedBS_ic		0	0	0	**93**	5	2	0	**0.0240**	**0.009**	0.026
DP_univar		49	18	2	31	0	0	0	0.0889	0.210	0.003
SMUCE		8	13	25	54	0	0	0	0.0598	0.029	0.002
MsFPOP		100	0	0	0	0	0	0	0.2500	0.930	0.001

Table 3 Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signal (S16). The average MSE and computational times are also given.

		$\hat{N} - N$					
Method	Signal	0	1	2	≥ 3	MSE	Time (s)
DAIS	(S16)	**100**	0	0	0	$\mathbf{2.12 \times 10^{-3}}$	0.032
ID_th		**100**	0	0	0	$\mathbf{2.12 \times 10^{-3}}$	0.049
ID_ic		**100**	0	0	0	$\mathbf{2.12 \times 10^{-3}}$	0.033
CPOP		**100**	0	0	0	2.45×10^{-3}	3.220
NOT		**98**	2	0	0	2.46×10^{-3}	0.427
MARS		0	100	0	0	5.25×10^{-3}	0.006
TF		11	51	30	8	5.29×10^{-3}	0.420
TS		**100**	0	0	0	$\mathbf{2.12 \times 10^{-3}}$	0.618
local_poly		10	5	17	68	6.47×10^{-3}	33.000

Table 4 Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signals (S17) and (S18). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$									
Method	Signal	≤ -15	$(-15, -2]$	-1	0	1	$[2, 15)$	≥ 15	MSE	d_H	Time (s)
DAIS	(S17)	-	0	0	**96**	4	0	0	0.022	0.104	0.003
ID_th		-	0	0	**95**	4	1	0	0.021	0.108	0.002
ID_ic		-	0	0	90	9	1	0	0.021	0.112	0.016
CPOP		-	0	0	87	12	1	0	0.081	0.122	0.017
NOT		-	0	0	**96**	3	1	0	**0.015**	**0.088**	0.065
MARS		-	84	13	3	0	0	0	3.720	2.160	0.003
TF		-	0	0	0	0	21	79	0.024	0.433	0.097
TS		-	0	0	**96**	4	0	0	0.094	0.198	0.110
local_poly		-	100	0	0	0	0	0	0.331	0.956	0.095
DAIS	(S18)	0	1	2	**96**	1	0	0	**0.463**	0.300	0.008
ID_th		0	0	2	**97**	1	0	0	**0.428**	**0.251**	0.008
ID_ic		0	2	0	**97**	1	0	0	0.476	0.287	0.045
CPOP		0	0	0	89	11	0	0	1.650	0.323	0.025
NOT		1	0	3	79	10	7	0	0.611	0.831	0.094
MARS		100	0	0	0	0	0	0	6.590	48.00	0.003
TF		0	0	0	0	0	82	18	1.200	0.323	0.110
TS		100	0	0	0	0	0	0	6.650	49.00	0.197
local_poly		100	0	0	0	0	0	0	6.610	5.220	0.639

2 Investigation of the impact of parameter choice on accuracy

In this section, we investigate the impact of the expansion parameter, λ_T, and the threshold constant, C, in the accuracy of DAIS using a small simulation study. The setup and signals used are the same as in Section 5 of the main paper and Section 1 of the supplementary material. The values used for the expansion parameter are $\lambda_T \in \{1, 3, 5, 10, 15, 20\}$ and for the threshold constant $C \in \{1.5, 1.6, \ldots, 1.9\}$ for the case of piecewise-constant signals and $C \in \{1.9, 2, \ldots, 2.3\}$ for continuous, piecewise-linear signals. The results in Table 5 indicate that our algorithm is robust to small and moderate changes of λ_T. Tables 6 and 7 indicate that any value of the threshold constant in the intervals $[1.5, 1.9]$ for piecewise-constant and $[1.9, 2.3]$ for continuous, piecewise-linear signals has good estimation accuracy.

3 Proof of Theorem 2

For the proof of Theorem 2, we require the following two lemmas.

Lemma 1 *Suppose $\boldsymbol{f} = (f_1, f_2, \ldots, f_T)^T$ is a piecewise-linear vector and $r_1, r_2, \ldots, r_N$ are the locations of the change-points. Suppose $1 \leq s < e \leq T$, such that $r_{j-1} \leq s < r_j < e \leq r_{j+1}$, for some $j = 1, 2, \ldots, N$. Let $\Delta_j^f = \left|2f_{r_j} - f_{r_j+1} - f_{r_j-1}\right|$ and $\eta = \min\{r_j - s, e - r_j\}$. Then,*

$$C_{s,e}^{r_j}(\boldsymbol{f}) = \max_{s<b<e} C_{s,e}^{b}(\boldsymbol{f}) \begin{cases} \geq \frac{1}{\sqrt{24}}\eta^{3/2}\Delta_j^f, \\ \leq \frac{1}{\sqrt{3}}(\eta+1)^{3/2}\Delta_j^f \end{cases}$$

Proof See Lemma 5 from [1]. $\square$

Lemma 2 *Suppose $\boldsymbol{f} = (f_1, f_2, \ldots, f_T)^T$ is a piecewise-linear vector and $r_1, r_2, \ldots, r_N$ are the locations of the change-points. Suppose $1 \leq s < e \leq T$, such that $r_{j-1} \leq s < r_j < e \leq r_{j+1}$, for some $j = 1, 2, \ldots, N$. Let $\rho = |r - b|$, $\Delta_j^f = \left|2f_{r_j} - f_{r_j+1} - f_{r_j-1}\right|$, $\eta_L = r_j - s$ and $\eta_R = e - r_j$. Then,*

$$\|\boldsymbol{\phi_{s,e}^{b}}\langle \boldsymbol{f}, \boldsymbol{\phi_{s,e}^{b}}\rangle - \boldsymbol{\phi_{s,e}^{r}}\langle \boldsymbol{f}, \boldsymbol{\phi_{s,e}^{r}}\rangle\|_2^2 = \left(C_{s,e}^{r_j}(\boldsymbol{f})\right)^2 - \left(C_{s,e}^{b}(\boldsymbol{f})\right)^2.$$

In addition,

1. *for any $r_j \leq b < e$, $\left(C_{s,e}^{r_j}(\boldsymbol{f})\right)^2 - \left(C_{s,e}^{b}(\boldsymbol{f})\right)^2 = \frac{1}{63}\min\{\rho, \eta_L\}^3\left(\Delta_j^f\right)^2$;*
2. *for any $s \leq b < r_j$, $\left(C_{s,e}^{r_j}(\boldsymbol{f})\right)^2 - \left(C_{s,e}^{b}(\boldsymbol{f})\right)^2 = \frac{1}{63}\min\{\rho, \eta_R\}^3\left(\Delta_j^f\right)^2$.*

Proof See Lemma 7 from [1]. $\square$

Table 5 Distribution of $\hat{N} - N$ over 100 simulated data sequences of signals for $\lambda_T \in \{1, 3, 5, 10, 15, 20\}$, for the piecewise-constant signals (S1), (S3), (S4), (S12) and (S15), and the continuous, piecewise-linear signals (S9) and (S10). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$									
λ_T	Signal	≤ 3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
1	(S1)	-	11	4	82	1	1	1	0.013	0.076	0.010
3		-	13	2	85	0	0	0	0.012	0.073	0.004
5		-	16	3	79	2	0	0	0.013	0.095	0.003
10		-	18	2	77	2	1	0	0.013	0.102	0.002
15		-	21	2	76	0	1	0	0.014	0.114	0.001
20		-	22	2	73	2	1	0	0.014	0.122	0.001
1	(S3)	0	0	1	95	4	0	0	0.022	0.009	0.003
3		0	0	0	98	2	0	0	0.022	0.009	0.003
5		0	0	2	97	1	0	0	0.022	0.009	0.002
10		0	0	5	95	0	0	0	0.025	0.011	0.002
15		0	0	3	92	5	0	0	0.026	0.012	0.002
20		0	0	3	91	5	1	0	0.028	0.013	0.003
1	(S4)	0	1	0	89	9	1	0	1.800	0.017	0.003
3		0	1	0	89	9	1	0	1.750	0.018	0.002
5		0	1	0	92	7	0	0	1.790	0.016	0.002
10		0	1	0	95	4	0	0	1.750	0.015	0.002
15		0	4	0	91	5	0	0	1.830	0.018	0.002
20		0	6	0	93	1	0	0	1.890	0.018	0.002
1	(S9)	0	0	0	98	2	0	0	0.031	0.089	0.028
3		0	0	0	98	2	0	0	0.030	0.086	0.011
5		0	0	0	99	1	0	0	0.029	0.085	0.007
10		0	0	0	98	2	0	0	0.028	0.083	0.004
15		0	0	0	99	1	0	0	0.026	0.082	0.004
20		0	0	0	100	0	0	0	0.025	0.072	0.003
1	(S10)	0	0	0	96	4	0	0	0.292	0.325	0.034
3		0	0	0	98	2	0	0	0.265	0.297	0.024
5		0	1	0	97	2	0	0	0.246	0.299	0.023
10		0	0	0	99	1	0	0	0.215	0.251	0.021
15		0	0	0	100	0	0	0	0.228	0.247	0.021
20		1	0	0	81	16	2	0	0.375	0.385	0.021
1	(S12)	-	-	-	99	0	1	0	1.9×10^{-4}	-	0.599
3		-	-	-	100	0	0	0	1.5×10^{-4}	-	0.209
5		-	-	-	100	0	0	0	1.5×10^{-4}	-	0.124
10		-	-	-	100	0	0	0	1.5×10^{-4}	-	0.062
15		-	-	-	100	0	0	0	1.5×10^{-4}	-	0.037
20		-	-	-	100	0	0	0	1.5×10^{-4}	-	0.032
1	(S15)	0	1	0	92	7	0	0	0.028	0.011	0.004
3		0	1	0	94	5	0	0	0.027	0.011	0.003
5		0	1	0	96	3	0	0	0.026	0.010	0.003
10		0	1	0	97	1	1	0	0.025	0.009	0.002
15		0	0	0	100	0	0	0	0.025	0.008	0.003
20		1	0	0	98	1	0	0	0.026	0.011	0.003

The steps of the proof of Theorem 2 are the same as for Theorem 1, which can be found in Appendix B. We work under the notation in (12), (13) and (14) of the main paper.

Table 6 Distribution of $\hat{N} - N$ over 100 simulated data sequences of signals for $C \in \{1.5, 1.6, 1.7, 1.8, 1.9\}$, for the piecewise-constant signals (S1), (S3), (S4) and (S14). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$									
C	Signal	≤ 3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
1.5		-	4	2	68	12	12	2	0.015	0.090	0.004
1.6		-	8	2	85	3	1	1	0.013	0.062	0.004
1.7	(S1)	-	14	2	84	0	0	0	0.013	0.078	0.004
1.8		-	21	2	77	0	0	0	0.014	0.113	0.004
1.9		-	28	4	68	0	0	0	0.016	0.149	0.004
1.5		0	0	0	91	8	1	0	0.021	0.009	0.002
1.6		0	0	0	95	5	0	0	0.021	0.008	0.002
1.7	(S3)	0	0	0	98	2	0	0	0.022	0.009	0.002
1.8		0	0	3	96	1	0	0	0.022	0.010	0.002
1.9		0	1	10	89	0	0	0	0.024	0.013	0.002
1.5		0	0	0	72	21	3	4	1.890	0.022	0.002
1.6		0	0	0	82	15	3	0	1.780	0.018	0.002
1.7	(S4)	0	1	0	89	9	1	0	1.750	0.018	0.002
1.8		0	2	0	96	1	1	0	1.730	0.016	0.002
1.9		0	3	0	96	0	1	0	1.760	0.017	0.002
1.5		-	-	-	75	6	16	3	0.001	-	0.197
1.6		-	-	-	93	2	5	0	0.001	-	0.222
1.7	(S12)	-	-	-	99	0	1	0	0.001	-	0.216
1.8		-	-	-	99	0	1	0	0.001	-	0.223
1.9		-	-	-	100	0	0	0	0.001	-	0.231
1.5		0	0	0	74	16	6	4	0.027	0.013	0.002
1.6		0	0	0	89	7	2	2	0.027	0.011	0.002
1.7	(S15)	0	1	0	94	5	0	0	0.027	0.011	0.002
1.8		1	1	0	96	2	0	0	0.027	0.013	0.002
1.9		1	4	1	93	1	0	0	0.028	0.017	0.002

Proof We will prove the more specific result

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\ldots,N}\left(\left|\hat{r}_j - r_j\right|\left(\Delta_j^f\right)^{2/3}\right) \leq C_3(\log T)^{1/3}\right) \geq 1 - \frac{1}{6\sqrt{\pi T}}, \tag{1}$$

which implies result (18) of the main paper.

Steps 1 & 2: Similar to Theorem 1, denote

$$A_T^* = \left\{\max_{s,b,e:1\leq s\leq b<e\leq T}\left|C_{s,e}^b(\boldsymbol{X}) - C_{s,e}^b(\boldsymbol{f})\right| \leq \sqrt{8\log T}\right\}$$

$$B_T^* = \left\{\max_{1\leq s\leq b<e<T}\max_{\substack{r_{j-1}<s\leq r_j\\ r_j<e\leq r_{j+1}\\ s\leq b<e}}\frac{\left|\left\langle \boldsymbol{\phi_{s,e}^b}\langle\boldsymbol{f},\boldsymbol{\phi_{s,e}^b}\rangle - \boldsymbol{\phi_{s,e}^r}\langle\boldsymbol{f},\boldsymbol{\phi_{s,e}^r}\rangle, \boldsymbol{\epsilon}\right\rangle\right|}{\|\boldsymbol{\phi_{s,e}^b}\langle\boldsymbol{f},\boldsymbol{\phi_{s,e}^b}\rangle - \boldsymbol{\phi_{s,e}^r}\langle\boldsymbol{f},\boldsymbol{\phi_{s,e}^r}\rangle\|_2} \leq \sqrt{8\log T}\right\}. \tag{2}$$

The same reasoning as in the proof of Theorem 1 leads to $\mathbb{P}(A_T^*) \geq 1 - 1/(12\sqrt{\pi}T)$ and $\mathbb{P}(B_T^*) \geq 1 - 1/(12\sqrt{\pi}T)$. Therefore, it holds that

$$\mathbb{P}\left(A_T^* \cap B_T^*\right) \geq 1 - \frac{1}{6\sqrt{\pi T}}$$

Table 7 Distribution of $\hat{N} - N$ over 100 simulated data sequences of signals for $C \in \{1.9, 2, 2.1, 2.2\}$, for the continuous, piecewise-linear signals (S9) and (S10). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$									
C	Signal	≤ 3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
1.9		0	0	0	92	8	0	0	0.032	0.097	0.009
2		0	0	0	97	3	0	0	0.032	0.091	0.010
2.1	(S9)	0	0	0	98	2	0	0	0.030	0.087	0.010
2.2		0	0	0	99	1	0	0	0.029	0.083	0.009
2.3		0	0	0	100	0	0	0	0.028	0.075	0.010
1.9		0	0	0	92	7	1	0	0.269	0.304	0.018
2		0	0	0	96	4	0	0	0.265	0.297	0.017
2.1	(S10)	0	0	0	99	1	0	0	0.265	0.295	0.017
2.2		0	1	0	97	2	0	0	0.267	0.311	0.018
2.3		0	1	0	95	4	0	0	0.270	0.329	0.032

Step 3: From now on, we assume that A_T^* and B_T^* both hold. The constants we use are

$$C_1 = \sqrt{\frac{2}{3}} C_3^{\frac{3}{2}} + \sqrt{8}, \quad C_2 = \frac{1}{8\sqrt{3n^3}} - \frac{2\sqrt{2}}{C^*}, \quad C_3 = 63^{\frac{1}{3}}(2\sqrt{2}+4)^{\frac{2}{3}},$$

where C^* satisfies Assumption (A3), $\delta_T^{3/2}\underline{f}_T \geq C^*\sqrt{\log T}$ and $n \geq 3/2$. As before, for $j \in \{1, 2, \ldots, N\}$ define I_j^L and I_j^R as in (5) of the main paper. For $d_{s,e} = \text{argmax}_{t \in \{s, s+1, \ldots, e-2\}}\{|X_{t+2} - 2X_{t+1} + X_t|\}$ being the location of the largest difference detected in the interval $[s, e]$, $1 \leq s < e \leq T$, define c_m^l and c_k^r as in (B6) of the main paper. Since the length of the intervals in (5) is $\delta_T/2n$, $\lambda_T \leq \delta_T/2n$ ensures that there exists at least one $m \in \{0, 1, \ldots, K^{\max}\}$ and at least one $k \in \{1, 2 \ldots, K^{\max}\}$ such that $c_m^l \in I_j^L$ and $c_k^r \in I_j^R$ for all $j \in \{1, 2, \ldots, N\}$.

At the beginning of DAIS, $s = 1, e = T$ and the first change-point that will get detected depends on the value of $d_{1,T}$. As already explained in Section 2.3 of the main paper, for $j \in \{1, \ldots, N-1\}$, the largest difference $d_{s,e}$ will be at most at a distance $\delta_{1,T}^j$ from the nearest change-point r_j or r_{j+1}, where $\delta_{1,T}^j \leq \frac{\tilde{\delta}_j}{2} - \frac{3\delta_T}{4n}$ for $\tilde{\delta}_j = r_{j+1} - r_j$. The first point to get detected will be the point that is closest to the largest difference $d_{1,T}$. The interval where the detection of this change-point occurs, cannot contain more than one change-points, as was explained in Section 2.3 of the main paper.

Without loss of generality, we suppose that the first change-point to get detected is r_J for some $J \in \{1, 2, \ldots, N\}$. We can show that there exists an interval $[c_m^l, c_k^r]$, for $m, k \in \{0, 1, \ldots, K^{\max}\}$, such that r_J is isolated using exactly the same argument as in (B7) of the main paper. We will now show that for $\tilde{b}_J = \text{argmax}_{c_m^l \leq t < c_k^r} C_{c_m^l, c_k^r}^t(\boldsymbol{X})$, then $C_{c_m^l, c_k^r}^{\tilde{b}_J}(\boldsymbol{X}) > \zeta_T$. Using (2), we have that

$$C_{c_m^l, c_k^r}^{\tilde{b}_J}(\boldsymbol{X}) \geq C_{c_m^l, c_k^r}^{r_J}(\boldsymbol{X}) \geq C_{c_m^l, c_k^r}^{r_J}(\boldsymbol{f}) - \sqrt{8 \log T}. \tag{3}$$

From Lemma 1, we have that

$$C_{c_m^l, c_k^r}^{r_J}(\boldsymbol{f}) \geq \frac{1}{\sqrt{24}}\left(\min\{r_J - c_m^l, c_k^r - r_J\}\right)^{\frac{3}{2}} \Delta_J^f.$$

Showing that

$$\min\{c_k^r - r_J, r_J - c_m^l\} \geq \frac{\delta_T}{2n}. \tag{4}$$

follows the exact same steps as Step 3 in the proof of Theorem 1 and will not be repeated. Now, using Assumption (A3) and the results in (3), (4), we have that

$$\begin{aligned} C_{c_m^l, c_k^r}^{\tilde{b}_J}(\boldsymbol{X}) &\geq \frac{1}{\sqrt{24}}\left(\frac{\delta_T}{2n}\right)^{\frac{3}{2}} \Delta_J^f - \sqrt{8\log T} \geq \frac{1}{\sqrt{24}}\left(\frac{\delta_T}{2n}\right)^{\frac{3}{2}} \underline{f}_T - \sqrt{8\log T} \\ &= \left(\frac{1}{8\sqrt{3n^3}} - \frac{2\sqrt{2\log T}}{\delta_T^{3/2}\underline{f}_T}\right)\delta_T^{3/2}\underline{f}_T \geq \left(\frac{1}{8\sqrt{3n^3}} - \frac{2\sqrt{2}}{C^*}\right)\delta_T^{3/2}\underline{f}_T \\ &= C_2\delta_T^{3/2}\underline{f}_T > \zeta_T \end{aligned} \tag{5}$$

and thus, the change-point will get detected.

Therefore, there will be an interval of the form $[c_{\tilde{m}}^l, c_{\tilde{k}}^r]$, such that the interval contains r_J and no other change-point and $\max_{c_{\tilde{m}}^l \leq t < c_{\tilde{k}}^r} C_{c_{\tilde{m}}^l, c_{\tilde{k}}^r}^t(\boldsymbol{X}) > \zeta_T$. For $k^*, m^* \in \{0, 1, \ldots K^{\max}\}$, denote by $c_{m^*}^l \geq c_{\tilde{m}}^l$ and $c_{k^*}^r \leq c_{\tilde{k}}^r$ the first left- and right-expanding points, respectively, that this happens and let $b_J = \operatorname{argmax}_{c_{m^*}^l \leq t < c_{k^*}^r} C_{c_{m^*}^l, c_{k^*}^r}^t(\boldsymbol{X})$, with $C_{c_{m^*}^l, c_{k^*}^r}^{b_J}(\boldsymbol{X}) > \zeta_T$. Note that b_J cannot be an estimation of $r_j, j \neq J$, as r_J is isolated in the interval where it is detected. Our aim now is to find $\tilde{\gamma}_T > 0$, such that for any $b^* \in \{c_{m^*}^l, c_{m^*}^l + 1, \ldots, c_{k^*}^r - 1\}$ with $|b^* - r_J|\left(\Delta_J^f\right)^2 > \tilde{\gamma}_T$, we have that

$$\left(C_{c_{m^*}^l, c_{k^*}^r}^{r_J}(\boldsymbol{X})\right)^2 > \left(C_{c_{m^*}^l, c_{k^*}^r}^{b^*}(\boldsymbol{X})\right)^2. \tag{6}$$

Proving (6) and using the definition of b_J, we can conclude that $|b_J - r_J|\left(\Delta_J^f\right)^2 \leq \tilde{\gamma}_T$. Now, using (11) of the main paper, it can be shown, for $\boldsymbol{\phi_{s,e}^b}$ as defined in (17) of the main paper, that (6) is equivalent to

$$\begin{aligned} \left(C_{c_{m^*}^l, c_{k^*}^r}^{r_J}(\boldsymbol{f})\right)^2 - \left(C_{c_{m^*}^l, c_{k^*}^r}^{b^*}(\boldsymbol{f})\right)^2 &> \left(C_{c_{m^*}^l, c_{k^*}^r}^{b^*}(\boldsymbol{\epsilon})\right)^2 - \left(C_{c_{m^*}^l, c_{k^*}^r}^{r_J}(\boldsymbol{\epsilon})\right)^2 \\ &+ 2\left\langle \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{b^*}} \langle \boldsymbol{f}, \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{b^*}} \rangle - \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{r_J}} \langle \boldsymbol{f}, \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{r_J}} \rangle, \boldsymbol{\epsilon} \right\rangle. \end{aligned} \tag{7}$$

Without loss of generality, assume that $b^* \in [r_J, c_{k^*}^r)$ and a similar approach holds when $b^* \in [c_{m^*}^l, r_J)$. Denote

$$\Lambda: = \left(C_{c_{m^*}^l, c_{k^*}^r}^{r_J}(\boldsymbol{f})\right)^2 - \left(C_{c_{m^*}^l, c_{k^*}^r}^{b^*}(\boldsymbol{f})\right)^2. \tag{8}$$

For the right-hand side of (7) using (2),

$$\left(C_{c_{m^*}^l, c_{k^*}^r}^{b^*}(\boldsymbol{\epsilon})\right)^2 - \left(C_{c_{m^*}^l, c_{k^*}^r}^{r_J}(\boldsymbol{\epsilon})\right)^2 \leq \max_{s,e,b: s \leq b < e}\left(C_{s,e}^b(\boldsymbol{\epsilon})\right)^2 - \left(C_{c_{m^*}^l, c_{k^*}^r}^{r_J}(\boldsymbol{\epsilon})\right)^2 \leq 8\log T \tag{9}$$

Using Lemma 2, (2) and (8), we have that for the left-hand side,

$$\begin{aligned} &2\left\langle \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{b^*}} \langle \boldsymbol{f}, \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{b^*}} \rangle - \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{r_J}} \langle \boldsymbol{f}, \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{r_J}} \rangle, \boldsymbol{\epsilon} \right\rangle \\ &\quad \leq 2\left\| \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{b^*}} \langle \boldsymbol{f}, \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{b^*}} \rangle - \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{r_J}} \langle \boldsymbol{f}, \boldsymbol{\phi_{c_{m^*}^l, c_{k^*}^r}^{r_J}} \rangle \right\|_2 \sqrt{8\log T} \\ &\quad = 2\sqrt{\Lambda}\sqrt{8\log T}. \end{aligned} \tag{10}$$

Using (8), (9) and (10), we can conclude that (7) is satisfied if $\Lambda > 8 \log T + \sqrt{2}\sqrt{\Lambda}\sqrt{8 \log T}$ is satisfied, which has solution

$$\Lambda > \left(2\sqrt{2}+4\right)^2 \log T.$$

Using Lemma 2, we can conclude that

$$\begin{aligned}
&\Lambda > \left(2\sqrt{2}+4\right)^2 \log T \\
&\Leftrightarrow \frac{1}{63}\left(\min\{|r_J - b^*|, r_J - c^l_{m^*}\}\right)^3 \left(\Delta^f_J\right)^2 > \left(2\sqrt{2}+4\right)^2 \log T \\
&\Leftrightarrow \min\{|r_J - b^*|, r_J - c^l_{m^*}\} > \frac{(63 \log T)^{1/3}\left(2\sqrt{2}+4\right)^{2/3}}{\left(\Delta^f_J\right)^{2/3}} = \frac{C_3 (\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}}
\end{aligned} \tag{11}$$

Now, if for sufficiently large T

$$\min\{r_J - c^l_{m^*}, c^r_{k^*} - r_J\} > 2^{1/3} C_3 \frac{(\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}} - 1, \tag{12}$$

it follows that

$$\min\{r_J - c^l_{m^*}, c^r_{k^*} - r_J\} > C_3 \frac{(\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}},$$

and we can deduce that (11) is restricted to

$$|r_J - b^*| > C_3 \frac{(\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}}$$

which implies (6). So, we conclude that necessarily

$$|r_J - b_J| \left(\Delta^f_J\right)^{2/3} \leq C_3 (\log T)^{1/3}. \tag{13}$$

But (12) must be true since if we assume that

$$\min\{r_J - c^l_{m^*}, c^r_{k^*} - r_J\} \leq 2^{1/3} C_3 \frac{(\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}} - 1,$$

then, using Lemma 1, we have that

$$\begin{aligned}
C^{b_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X}) &\leq C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{f}) + \sqrt{8 \log T} \\
&\leq \frac{1}{\sqrt{3}}\left(\min\{r_J - c^l_{m^*}, c^r_{k^*} - r_J\} + 1\right)^{3/2} \Delta^f_J + \sqrt{8 \log T} \\
&\leq \frac{1}{\sqrt{3}}\left(2^{1/3} C_3 \frac{(\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}}\right)^{3/2} \Delta^f_J + \sqrt{8 \log T} \\
&= \sqrt{\frac{2}{3}} C_3^{3/2} \sqrt{\log T} + \sqrt{8 \log T} \\
&= \left(\sqrt{\frac{2}{3}} C_3^{3/2} + \sqrt{8}\right) \sqrt{\log T} = C_1 \sqrt{\log T} \leq \zeta_T,
\end{aligned}$$

which contradicts $C^{b_J}_{c^l_{m^*},c^r_{k^*}}(\boldsymbol{X}) > \zeta_T$.

Thus, we have proved that for $\lambda_T \leq \delta_T/2n$, working under the assumption that both A^*_T and B^*_T hold, there will be an interval $[c^l_{m^*}, c^r_{k^*}]$ with $C^{b_J}_{c^l_{m^*},c^r_{k^*}}(\boldsymbol{X}) > \zeta_T$, where $b_J = \mathrm{argmax}_{c^l_{m^*} \leq t < c^r_{k^*}} C^t_{c^l_{m^*},c^r_{k^*}}(\boldsymbol{X})$ is the estimated location for the change-point r_J that satisfies (13).

Step 4: After the detection of the change-point r_J at the estimated location b_J in the interval $[c^l_{m^*}, c^r_{k^*}]$, the process is repeated in the disjoint intervals $[1, c^l_{m^*}]$ and $[c^r_{k^*}, T]$. Proving that there is no other change-point in the interval $[c^l_{m^*}, c^r_{k^*}]$ can be done in exactly the same way as Step 4 in the case of piecewise-constant signals and will not be repeated here.

Step 5: After detecting r_J, the algorithm will first check the interval $[1, c^l_{m^*}]$. So, unless $r_J = r_1$ and $[1, c^l_{m^*}]$ contains no other change-points, the next change-point to get detected will be one of $r_1, r_2, \ldots, r_{J-1}$. The location of the largest difference in the interval $[1, c^l_{m^*}]$, $d_{1,c^l_{m^*}}$, will again determine which change-point will be detected next and as in Step 5 for piecewise-constant signals, we only need to consider the case when the next change-point to get detected is r_{J-1}.

Now, concentrating on the case that the next change-point to get detected is r_{J-1}, we mention that since this is the closest change-point to the already detected r_J, we need to make sure that detection is possible. As before, for $k_{J-1} \in \{1, \ldots, K^{\max}\}$ and $m_{J-1} \in \{k_{J-1} - 1, k_{J-1}\}$ we will show that r_{J-1} gets detected in $[c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}]$, where $c^l_{m^*_{J-1}} \geq c^l_{m_{J-1}}$ and $c^r_{k^*_{J-1}} \leq c^r_{k_{J-1}} \leq c^l_{m^*}$ and its detection is at location

$$b_{J-1} = \mathrm{argmax}_{c^l_{m^*_{J-1}} \leq t < c^r_{k^*_{J-1}}} C^t_{c^l_{m^*_{J-1}},c^r_{k^*_{J-1}}}(\boldsymbol{X}),$$

which satisfies $\left|r_{J-1} - b_{J-1}\right| \left(\Delta^f_{J-1}\right)^{2/3} \leq C_3 (\log T)^{1/3}$. Firstly, r_{J-1} is isolated in the interval $[c^l_{m_{J-1}}, c^r_{k_{J-1}}]$ using the same argument as in (B7) of the main paper. Using Lemma 2, we have that for $\tilde{b}_{J-1} = \mathrm{argmax}_{c^l_{m_{J-1}} \leq t < c^r_{k_{J-1}}} C^t_{c^l_{m_{J-1}},c^r_{k_{J-1}}}(\boldsymbol{X})$,

$$\begin{aligned} C^{\tilde{b}_{J-1}}_{c^l_{m_{J-1}},c^r_{k_{J-1}}}(\boldsymbol{X}) &\geq C^{\tilde{r}_{J-1}}_{c^l_{m_{J-1}},c^r_{k_{J-1}}}(\boldsymbol{X}) \geq C^{\tilde{r}_{J-1}}_{c^l_{m_{J-1}},c^r_{k_{J-1}}}(\boldsymbol{f}) - \sqrt{8\log T} \\ &\geq \frac{1}{\sqrt{24}} \left(\min\{r_{J-1} - c^l_{m_{J-1}}, c^r_{k_{J-1}} - r_{J-1}\}\right)^{3/2} \Delta^f_{J-1} - \sqrt{8\log T}. \end{aligned} \tag{14}$$

Before we show that $\min\{c^r_{k_{J-1}} - r_{J-1}, r_{J-1} - c^l_{m_{J-1}}\} \geq \delta_T/2n$, we need show that $c^l_{m^*}$ satisfies $c^l_{m^*} - r_{J-1} \geq \delta_T/2n$ The proof is exactly the same as in Step 5 of the proof of Theorem 1. It can be deduced that the right end-point of the interval will satisfy $c^r_{k_{J-1}} - r_{J-1} > \delta_T/2n$ for some k_{J-1}. For the left end-point of the interval, the same holds as in Step 3. In any case, $\min\{c^r_{k_{J-1}} - r_{J-1}, r_{J-1} - c^l_{m_{J-1}}\} > \delta_T/2n$ holds and so, from (14), using exactly the same calculations as (5), we have that

$$C^{\tilde{b}_{J-1}}_{c^l_{m_{J-1}},c^r_{k_{J-1}}}(\boldsymbol{X}) = C_2(\delta_T)^{3/2}\underline{f}_T > \zeta_T.$$

Therefore, we have shown that there exists an interval of the form $[c^l_{\tilde{m}_{J-1}}, c^r_{\tilde{k}_{J-1}}]$ with $\max_{c^l_{\tilde{m}_{J-1}} \leq b < c^r_{\tilde{k}_{J-1}}} C^t_{c^l_{\tilde{m}_{J-1}},c^r_{\tilde{k}_{J-1}}}(\boldsymbol{X}) > \zeta_T$.

Now, denote $c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}$ the first points where this occurs and b_{J-1} as defined above with $C^{b_J-1}_{c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}}(\boldsymbol{X}) > \zeta_T$. We can show that $|r_{J-1} - b_{J-1}| \left(\Delta^f_{J-1}\right)^{2/3} \leq C_3 (\log T)^{1/3}$, following exactly the same process as in Step 3.

After detecting r_{J-1} in the interval $[c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}]$, DAIS will restart on intervals $[1, c^l_{m^*_{J-1}}]$ and $[c^r_{k^*_{J-1}}, c^l_{m^*}]$. Step 5 can be applied to all intervals, as long as there is a change-point. We can conclude that all change-points will get detected, one by one, and their estimated locations will satisfy $|r_j - b_j| \left(\Delta^f_j\right)^{2/3} \leq C_3 (\log T)^{1/3}$, $\forall j \in \{1, 2, \dots, N\}$. There will not be any double detection issues as each interval contains no previously detected change-points.

Step 6: After detecting all the change-points at locations $b_1, b_2, \dots, b_N$ using the intervals $[c^l_{m^*_j}, c^r_{k^*_j}]$ for $j \in \{1, \dots, N\}$, the algorithm will check all intervals of the form $[c^r_{k^*_{j-1}}, c^l_{m^*_j}]$ and $[c^r_{k^*_j}, c^l_{m^*_{j+1}}]$, with $c^l_{m^*_0} = 1$ and $c^r_{k^*_{N+1}} = T$. At most $N+1$ intervals of this form, containing no change-points will be checked. Denoting by $[s^*, e^*]$ any of those intervals, we can show that DAIS will not detect any change-point as for $b \in \{s^*, s^*+1, \dots, e^*-1\}$,

$$C^b_{s^*,e^*}(\boldsymbol{X}) \leq C^b_{s^*,e^*}(\boldsymbol{f}) + \sqrt{8 \log T} = \sqrt{8 \log T} < C_1 \sqrt{\log T} \leq \zeta_T.$$

The algorithm will terminate after not detecting any change-points in all intervals.

□

References

[1] Baranowski, R., Chen, Y., Fryzlewicz, P.: Narrowest-Over-Threshold Detection of Multiple Change Points and Change-Point-Like Features. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **81**(3), 649–672 (2019) https://doi.org/10.1111/rssb.12322