AIMS Mathematics, 10(12): 29853–29872.
DOI: 10.3934/math.20251311
Received: 28 September 2025
Revised: 06 December 2025
Accepted: 12 December 2025
Published: 18 December 2025

https://www.aimspress.com/journal/Math

Research article

Window-sliding based NSP algorithm for multiple change-points estimation

Xiaoyuan Zhang[1] and Zhanshou Chen[1,2,*]

[1] School of Mathematics and Statistics, Qinghai Normal University, Xining, China
[2] The State Key Laboratory of Tibetan Intelligence, Qinghai Normal University, Xining 810008, China

* **Correspondence:** Email: chenzhanshou@126.com.

Abstract: A window-sliding based narrowest significance pursuit (WNSP) algorithm is proposed for multiple change-points estimation. The algorithm adopts a "post-inference selection" approach: first, it automatically identifies the narrowest significant intervals containing at least one change point using the narrow significance tracking (NSP) method at a global significance level α; then, within each interval, it employs adaptive bandwidth and single-peak detection techniques to achieve precise estimation of change-point locations. Theoretical analysis confirms the method's consistency and finite-sample reliability under general noise conditions. Numerical simulations and real-world data analysis demonstrate the WNSP algorithm's effectiveness and robustness across diverse noise distributions and signal structures.

Keywords: change-points estimation; window-sliding; narrowest significance pursuit; confidence intervals; local statistic
Mathematics Subject Classification: 62F03, 62L10

1. Introduction

Change-points detection and estimation has been a subject of great importance in the field of statistical analysis since the 1950s. Change-point problems have been the focus of academic research in a number of disciplines, including economics [1], finance [2], medicine [3], engineering [4], and environmental studies [5]. A number of studies have provided detailed discussions of change-point models from a theoretical perspective. For surveys, we refer the reader to Chen and Gupta [6] and

Horvath and Rice [7], among many others.

This paper focuses on linear models. Given a design matrix $X = (X_{t,i}), t = 1, \dots, T, i = 1, \dots, p$, the response Y_t is modeled as

$$Y_t = f_t + \varepsilon_t, \qquad t = 1, \dots, T, \tag{1}$$

where the unknown signal function f_t satisfies:

$$f_t = X_{t,i}\beta^{(j)}, \qquad t = \tau_j + 1, \dots, \tau_{j+1}, \qquad j = 0, \dots, N. \tag{2}$$

Here, τ_j denotes a change-point location, i.e., $f_{\tau_j - 1} = f_{\tau_j}$ but $f_{\tau_j} \neq f_{\tau_j + 1}$, and N is the number of change- points. The parameter vector is

$$\beta^{(j)} = \left(\beta_1^{(j)}, \dots, \beta_p^{(j)}\right)'$$

such that $\beta^{(j)} \neq \beta^{(j+1)}$, and ε_t is a zero-mean noise term whose distribution may vary from identically distributed Gaussians to autocorrelated, heavy-tailed, and heteroscedastic forms.

Although the model is linear in β, the design matrix X can be arbitrary. For example, a $T \times 1$ matrix of ones yields a piecewise constant signal with noise. If X is $T \times p$ with its i-th column given by $(t/T)^{i-1}$, the model represents a piecewise polynomial signal. The framework also encompasses regression with coefficient shifts and autoregressive settings where Y_t depends on exogenous covariates and lagged values $Y_{t-1}, Y_{t-2}, \dots$, with abrupt changes in the dependence structure. We now review the uncertainty in multiple change-point problems within the existing literature.

Existing change-point detection methods can be broadly categorized into two types: optimization-based approaches and greedy procedures. The former includes dynamic programming detection methods [8–10] and penalty cost methods [11,12], which often suffer from high computational complexity that increases significantly with sample size n. While the PELT algorithm [13] reduces this to O(n), it requires specific assumptions and lacks statistical consistency guarantees.

The most commonly used greedy process method is the Binary Segmentation (BS) algorithm combined with the CUSUM statistic [14]; these are widely adopted for their efficiency (typically $O(n \log n)$) and programming convenience. However, BS performs poorly when small segments are sandwiched between large ones. Subsequent improvements include Circular Binary Segmentation (CBS) [15], Wild Binary Segmentation (WBS) [16] (whose recent extension, Wild Binary Segmentation 2 (WBS2) [29867], further enhances performance through a steepest-drop model selection criterion), Narrowest-over-threshold method [18], and Seeded Binary Segmentation (SeedBS) method [19], which enhances performance through deterministic interval construction. Recent variants also encompass ensemble binary segmentation for irregular time series [20] and multi-dimensional CUSUM-based approaches [21,22]. Systematic reviews of related methods can be found in Truong et al. [23], Cho and Kirchs [24], and Shi et al. [25].

In contrast to the two global methods above, local approaches leverage local information to achieve lower computational complexity. The Screening and Ranking Algorithm (SaRa) [26], which was originally developed for detecting DNA copy number variations, attains time complexity using a forward algorithm and local CUSUM statistic. However, it requires careful bandwidth selection, as narrow bandwidths increase false positives and wide ones may miss true change-points. Subsequent improvements include the reverse CNV method [27] for high-dimensional short signals, scanning-based confidence intervals [28], the MOSUM method [29] for small samples, rank-based scanning for

heavy-tailed data [30], and the isolation-detection algorithm [31], which improves localization at the cost of speed in long stationary sequences.

Most existing change-point detection methods follow a "model selection before inference" framework, requiring pre-estimation of the number and locations of change-points before constructing confidence intervals. This post-selection inference strategy, however, often suffers from selection bias that distorts conditional coverage. Fryzlewicz [32] proposed the Narrowest Significance Pursuit (NSP) method, which reverses this process via an "post-inference selection" approach. It first identifies the narrowest intervals containing at least one change-point at level α, but only offers unconditional interval inference without point estimates, limiting its use in precision-critical applications. Therefore, extending NSP to provide accurate change-point location estimates while preserving its finite-sample coverage guarantees remains an important open problem.

To overcome this, we propose the WNSP algorithm. It first uses NSP to identify significant intervals, then applies adaptive bandwidth selection and single-peak detection within each interval to accurately locate change-points. By extending the "infer first, then select" principle, WNSP simultaneously provides rigorous interval-level coverage and precise point estimates, ensuring both inferential validity and localization accuracy.

The paper is structured as follows: Section 2 outlines the NSP algorithm. Section 3 details the WNSP algorithm. Section 4 presents its theoretical properties. In Section 5, we provide numerical simulations. Section 6 demonstrates real-world applications. Finally, the work concludes with some final remarks in Section 7.

2. An overview of the NSP algorithm

The NSP algorithm aims to automatically identify the narrowest interval containing at least one change point at a global significance level α. It fits local linear models and examines residuals under general distributional assumptions while maintaining finite-sample coverage. The procedure consists of four layers:

I. Scanning layer: multi-scale local testing

The data is scanned using sliding windows of binary lengths (e.g., $I^d = 1, 3, 7, 15, \ldots, 2^j - 1$) over a sparse grid or random intervals to achieve $O(T \log T)$ complexity. Within each window $[s, e]$, a multi-scale scan statistic is used as the loss function to rapidly fit a model and compute residuals. The same statistic is applied to the residuals to obtain a deviation measure $D(s, e)$.

II. Threshold layer: global significance control

A threshold λ_α is derived from extreme value theory to control the global significance level at α. For Gaussian errors with known variance σ^2, the threshold is set as:

$$\lambda_\alpha = \sigma(a_T + b_T\gamma),$$

where the coefficients a_T and b_T are defined as:

$$a_T = \sqrt{2\log T + \frac{\log\log T}{2} - \frac{\log\pi}{2}},\ b_T = \frac{1}{\sqrt{2\log T}},$$

and γ is the solution to the equation:

$$\alpha = 1 - \exp(-2e^{-\gamma}).$$

For the case of unknown variance or non-Gaussian heavy-tailed errors, robust estimators such as the median absolute deviation (MAD) or self-normalization techniques are employed to estimate the scale parameter. This construction ensures that under the null hypothesis of no change-points,

$$P(\max D(s,e) > \lambda_\alpha) \leq \alpha.$$

The forms of a_T and b_T are derived from the limiting distribution of the multiresolution sup-norm of Gaussian noise processes (see Theorem 2.2 in Fryzlewicz, 2024).

III. Selection layer: shortest significant interval

Among intervals with $D(s,e) \geq \lambda_\alpha$, the shortest interval is chosen, and ties are broken by maximum deviation. Recursive binary search is applied to left and right subintervals until no more significant intervals are found, ensuring local minimality.

IV. Output layer

The algorithm returns a set of non-overlapping intervals S, each guaranteed to contain at least one true change-point with probability no less than $1-\alpha$, a property that holds even in finite samples. The overlap tolerance can be adjusted if overlapping intervals are desired.

In summary, NSP transforms multi-change-point detection into an automated, computationally efficient, and statistically rigorous process through multi-scale scanning, threshold calibration, recursive selection, and guaranteed output.

3. The WNSP algorithm for change-point estimation

The proposed WNSP algorithm is designed to precisely locate the position of a single change-point within a significant interval $[s,e]$ identified by the NSP algorithm. The core idea is to employ a sliding window strategy that computes a local discrepancy statistic at each candidate point within the interval. The point that maximizes this statistic is then identified as the estimated change-point. In the following algorithm description, we use k to index the significant intervals identified by the NSP algorithm (which correspond to the intervals $[s_j, e_j]$ described in Lemma 1). Here, j denotes the true change-point indices, while k denotes the indices of the algorithmically identified intervals. Let $Y_{\{s_k-h_k:e_k+h_k\}} = \{Y_{s_k-h_k}, Y_{s_k-h_k+1}, \ldots, Y_{e_k+h_k}\}$ denote the observed data sequence, which includes the significant interval $[s_k, e_k]$ and an appropriate buffer on both sides. The buffer is introduced to mitigate boundary effects during local estimation.

For each candidate point η within the interval $[s_k, e_k]$, the algorithm defines two adjacent sliding windows of adaptive width h. The left window comprises the data points from index $\eta - h_k$ to $\eta - 1$, and the right window comprises the data points from index η to $\eta + h_k - 1$. These index ranges are denoted as idx_L and idx_R, respectively, in the pseudocode:

$$idx_L = \eta - h_k, \ldots, \eta - 1; idx_R = \eta, \ldots, \eta + h_k - 1;$$

$$W_{L(\eta)} = \{Y_t : t \in idx_L\}; W_{R(\eta)} = \{Y_t : t \in idx_R\},$$

the window width h_k is adaptively determined for each significant interval $[s_k, e_k]$ as

$$h_k = \lfloor 0.1 * (e_k - s_k + 1) \rfloor, \qquad k = 1, \ldots, K, \tag{3}$$

where $\lfloor \cdot \rfloor$ denotes the floor operation, ensuring that h_k is an integer number of data points. This choice ensures that the window size is proportional to the length of the significant interval, providing a balance between estimation stability (larger windows reduce variance) and localization accuracy (smaller windows avoid over-smoothing near change-points). The proportionality constant 0.1 was empirically optimized through simulations (see Section 5), and aligns with the theoretical requirement that h_k should be a small fraction of the interval length to ensure localization accuracy.

When the candidate point η is near the boundaries of the significant interval $[s, e]$ (i.e., $|\eta - s| < h_k$ or $|\eta - e| < h_k$), a dual-buffer truncation strategy is adopted , and a fixed buffer zone of length h_k is established on both sides of the original data range; if either $W_{L(\eta)}$ or $W_{R(\eta)}$ extends beyond this buffer zone, the actual window range is adjusted to $[s, \eta - h_k/2]$ and $[\eta + h_k/2, e]$, with the portions extending beyond the buffer filled with zeros. The local discrepancy statistic $D(\eta)$ is computed by fitting two separate models—one to $W_{L(\eta)}$ and another to $W_{R(\eta)}$—and measuring the absolute difference between the estimated parameters. The specific form of $D(\eta)$ depends on the assumed signal model:

Case 1: Piecewise constant model (mean shift)

$$D(\eta) = \left|\hat{\mu}_{L(\eta)} - \hat{\mu}_{R(\eta)}\right|, \tag{4}$$

where

$$\hat{\mu}_{L(\eta)} = (1/h_k)\sum_{t=\eta-h_k}^{\eta-1} Y_t,\ \hat{\mu}_{R(\eta)} = (1/h_k)\sum_{t=\eta}^{\eta+h_k-1} Y_t.$$

Case 2: Piecewise linear model (slope change)

$$D(\eta) = \left|\hat{\xi}_{L(\eta)} - \hat{\xi}_{R(\eta)}\right|, \tag{5}$$

where $\hat{\xi}_{L(\eta)}$ and $\hat{\xi}_{R(\eta)}$ are the estimated slope coefficients obtained by performing linear regression on the data in the left and right windows, respectively.

The local discrepancy statistic $D(\eta)$, defined as the L_1 distance $|| \hat{\xi}_{L(\eta)} - \hat{\xi}_{R(\eta)} ||_1$ between parameter estimates from adjacent windows, serves as a sensitive measure for change-points. Its rationale is that if η lies within a stationary segment, the value of $D(\eta)$ will be close to zero; if η is located at a change-point, $D(\eta)$ will attain a local maximum. This behavior is rigorously justified in Supplementary A.2, where we show that under mild assumptions, $D(\eta)$ concentrates around its expected value, and the maximum occurs near the true change-point with high probability.

The precise location of the change- point is then estimated by maximizing $D(\eta)$ over the candidate points:

$$\hat{\tau}_k = argmax_{\{\eta \in [s_k, e_k]\}} D(\eta). \tag{6}$$

The algorithm requires only a single pass over the significant interval $[s_k, e_k]$. The computational complexity for processing one such interval of length $n_k = e_k - s_k + 1$ is $O(n_k)$, making the WNSP algorithm both robust and computationally efficient. A schematic illustration of the algorithm's operation is provided in Figures 1 and 2 (Section 5).

The pseudocode for the WNSP algorithm is summarized as follows:

Input: Significant intervals I= $[s_1, e_1], \ldots, [s_k, e_k]$ from NSP.
Data sequence $\{Y_t\}_{t=1}^{T}$.
Model type: "constant" or "linear"
Output: Estimated change-points $\hat{\mathcal{T}} = \hat{\tau}_1, \ldots, \hat{\tau}_k$.

(1): Initialize $\hat{\mathcal{T}} = \emptyset$
(2): **for** $k = 1$ *to* k **do**
(3): $L_k = e_k - s_k + 1$
(4): $h_k = \lfloor 0.1 * L_k \rfloor$ ▷ *Adaptive window width*
(5): Initialize $D[\,]$ as an empty list
(6): **for** $\eta = s_k + h_k$ *to* $e_k - h_k + 1$ **do** ▷ *Avoid boundaries*
(7): $idx_L = \{\eta - h_k, \ldots, \eta - 1\}$ ▷ *Left window indices*
(8): $idx_R = \{\eta, \ldots, \eta + h_k - 1\}$ ▷ *Right window indices*
(9): **if** $mode_type ==$ constant then
(10): $\hat{\mu}_L = mean((Y[idx_L])$
(11): $\hat{\mu}_R = mean((Y[idx_R])$
(12): $D(\eta) = |\hat{\mu}_L - \hat{\mu}_R|$
(13): **else if** $mode_type ==$ linear then
(14): $\hat{\beta}_L = slope(Y[idx_L] \sim t[idx_L])$ ▷Linear regression on left window
(15): $\hat{\beta}_R = slope(Y[idx_R] \sim t[idx_R])$ ▷Linear regression on right window
(16): $D(\eta) = \left|\hat{\beta}_L - \hat{\beta}_R\right|$
(17): **end if**
(18): Append $D(\eta)$ to $D[\,]$
(19): **end for**
(20): $\hat{\tau}_k = \arg\max D[\,]$
(21): **end for**
(22): **return** $\hat{\mathcal{T}}$

4. Consistency results

This section establishes the theoretical foundation for the WNSP algorithm, demonstrating that it satisfies the deterministic coverage property and achieves consistent estimation of change-point locations. We begin by recalling key assumptions and results for the NSP algorithm, upon which our method builds.

Assumption 1. (Minimum spacing and signal strength) For each change-point $\tau_j, j = 1, \ldots, N$, assume that

$$\tau_{j+1} - \tau_j \geq 2\bar{d}_{j+1} + 2\bar{d}_j - 2, j = 1, \ldots, N-1, \tau_1 - \tau_0 \geq 2\bar{d}_1 - 1, \tau_{N+1} - \tau_N \geq 2\bar{d}_N - 1,$$

where $\tau_0 = 0, \tau_{N+1} = T,$ and

$$\bar{d}_j = \left\lceil \frac{16\lambda_\alpha^2}{\left|f_{\tau_{j+1}} - f_{\tau_j}\right|^2} \right\rceil + 1. \tag{7}$$

This assumption ensures that the change-points are well-separated and their magnitudes are sufficiently large to be identifiable.

Assumption 2. There exists a positive sequence ψ_T such that $\psi_T \to \infty$ as $T \to \infty$, and the length of each significant interval $[s_j, e_j]$ identified by the NSP algorithm satisfies: $(e_j - s_j + 1) \geq \psi_T$, for $j = 1, \ldots, N$.

This assumption ensures that every significant interval is sufficiently long, which guarantees that the adaptive window width h_k also tends to infinity, thereby enabling consistent parameter estimation within the left and right windows of the WNSP algorithm.

Lemma 1. (NSP coverage guarantee) Under Assumption 1 and on the set $\| \varepsilon \|_I \leqslant \lambda_\alpha$, if the NSP algorithm considers all non-overlapping subintervals with threshold λ_α, it returns exactly N significant intervals $[s_1, e_1] <\cdots< [s_N, e_N]$ such that

$$\tau_j \in [s_j, e_j - 1],\ e_j - s_j + 1 \leq 2\bar{d}_j, j = 1, \ldots, N.$$

This lemma guarantees that with high probability (controlled by the confidence level α), the coarse detection stage successfully identifies N intervals, each containing one and only one true change-point.

Proof. See Supplementary A.1.

Theorem 1. (WNSP estimation consistency) If assumptions in Lemma 1 hold, then there is a constant C>0, such that

$$P(max_{1 \leq j \leq N} \mid \hat{\tau}_j - \tau_j \mid \leq C\log T) \to 1, T \to \infty.$$

This theorem shows that the WNSP algorithm achieves the sure coverage property, and the estimation error is bounded by $O(\log T)$ with high probability.

Proof. See Supplementary A.2.

5. Simulation study

5.1. Simulation data and experimental design

This section presents numerical simulations to verify that, under the same global coverage probability $(\geq 1 - \alpha)$ as NSP, the WNSP method utilizes adaptive window widths and single-peak detection techniques to accurately locate change- points within significant intervals. We first introduce the competing methods. WNSP is a local optimization approach, while SaRa (another competitive local optimization method) uses a window width set as $h = C\log n$, where $C \in 1, 2, 3$ as recommended by Niu and Zhang [26]. The Wild Binary Segmentation (WBS) method [16], a variational search technique based on random subintervals, is widely adopted due to its simplicity and computational efficiency. As a significant advancement of WBS, WBS2 [29867] addresses certain limitations of the original WBS in model selection, providing stronger theoretical guarantees for detecting frequent change-points. The WNSP framework proposed in this paper shares the goal of enhancing the robustness and accuracy of change-point detection with WBS2 but follows a distinct technical path: WNSP focuses on local refined search within statistically significant candidate intervals, rather than global optimization. The Narrowest-Over-Threshold (NOT) method [18] shares conceptual

similarities with WBS, as both improve binary segmentation via random subinterval sampling.

WNSP employs parameters identical to those of NSP, using a deterministic grid with $M = 1000$ intervals. The threshold is set as $\lambda_\alpha = \sigma(a_T + b_T\gamma)$, where σ is estimated via the median absolute deviation (MAD). Simulations were conducted on both piecewise constant and piecewise linear signals under three error settings: independent and identically distributed $(i.i.d.)$ Gaussian, $t(5)$, and heteroscedastic ε_t. The performance of each algorithm was systematically evaluated. A summary of the models used is provided in Table 1.

Table 1. Models for comparative simulation studies in section 5; "NO. of cpts" denotes the number of control points.

Model name	NO. of cpts	Sample path execution in R
M1	4	$\mathrm{c}\big(\mathrm{rep}(0,100),\mathrm{rep}(2,100),\mathrm{rep}(5,100),\mathrm{rep}(1,100),\mathrm{rep}(3{,}100)\big) + \mathrm{rnorm}(500)$
M2	4	$\mathrm{c}\big(\mathrm{rep}(0,100),\mathrm{rep}(2,100),\mathrm{rep}(5,100),\mathrm{rep}(1,100),\mathrm{rep}(3{,}100)\big) + \mathrm{rt}(500,\mathrm{df}=5)$
M3	4	$\mathrm{c}\big(\mathrm{rep}(0,100),\mathrm{rep}(2,100),\mathrm{rep}(5,100),\mathrm{rep}(1,100),\mathrm{rep}(3{,}100)\big) + \mathrm{rnorm}(500) * \mathrm{seq}(0.2{,}1,\mathrm{length.out}=500)$
M4	2	$\mathrm{ifelse}\left(\mathrm{t}\le 100,\frac{\mathrm{t}}{100},\mathrm{ifelse}\left(\mathrm{t}\le 200,1-\frac{\mathrm{t}-100}{100},\frac{\mathrm{t}-200}{100}\right)\right) + \mathrm{rnorm}(300)$
M5	2	$\mathrm{ifelse}\left(\mathrm{t}\le 100,\frac{\mathrm{t}}{100},\mathrm{ifelse}\left(\mathrm{t}\le 200,1-\frac{\mathrm{t}-100}{100},\frac{\mathrm{t}-200}{100}\right)\right) + \mathrm{rt}(300,\mathrm{df}=5)$
M6	2	$\mathrm{ifelse}\left(\mathrm{t}\le 100,\frac{\mathrm{t}}{100},\mathrm{ifelse}\left(\mathrm{t}\le 200,1-\frac{\mathrm{t}-100}{100},\frac{\mathrm{t}-200}{100}\right)\right) + \mathrm{rnorm}(300) * \mathrm{seq}(0.2{,}1,\mathrm{length.out}=300)$

Table 1 provides a detailed summary of the parameter settings for the six data-generating processes $(M1 - M6)$. $M1 - M3$ are piecewise constant models, each containing four change-points, differing only in their noise distributions: $M1$ employs standard Gaussian noise, $M2$ uses heavy-tailed $t - distributed$ noise $(df = 5)$, and M3 utilizes heteroscedastic Gaussian noise. $M4 - M6$ are continuous piecewise linear models, each containing two change-points and exhibiting a triangular wave pattern, with noise distributions corresponding to Gaussian, t-distribution, and heteroscedastic Gaussian, respectively. In all models, t represents the time index, ranging from 1 to $\mathrm{T}(M1 - M3{:}\,\mathrm{T} = 500; M4 - M6{:}\,\mathrm{T} = 300)$. The R code column displays the specific data generation commands.

To better demonstrate the performance of the WNSP algorithm on analog signals, we present the interval detection results of NSP and the precise change-point localization effects of WNSP on the segmented constant model (M1) and the segmented linear model (M4), respectively.

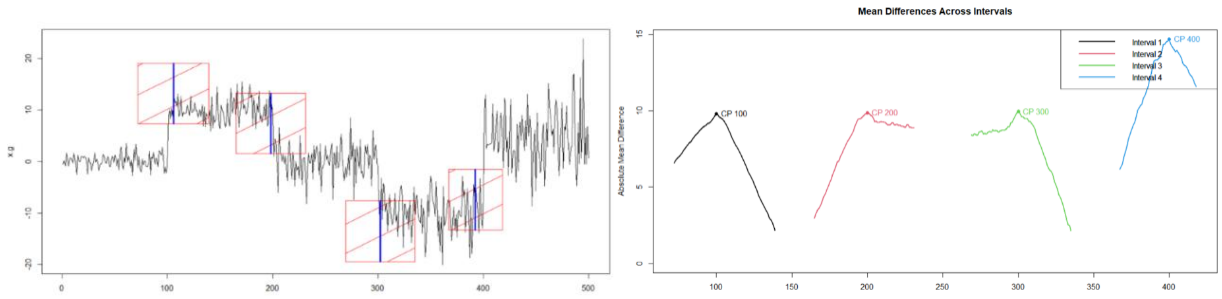

Figure 1. (model M1) Significant intervals (red shaded boxes) and their midpoints (blue) returned by the NSP (Left). Precise change-point estimation by WNSP via local statistic peaks (Right).

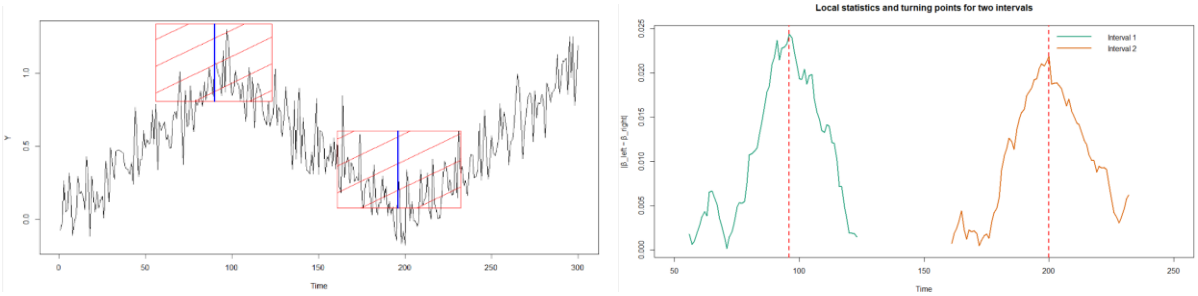

Figure 2. (model M4) Significant intervals (red shaded boxes) and their midpoints (blue) returned by the NSP (Left). Precise change-point estimation by WNSP via local statistic peaks (Right).

5.2. Results and discussion

For each method, we show a frequency table for the distribution of $\widehat{N} - N$, where $\widehat{N}$ is the number of the estimated change-points, and N denotes the true number of change-points. We also report Monte Carlo estimates of the mean-squared error of the estimated signal, given by

$$MSE = E\left\{\frac{1}{T}\sum_{t=1}^{T}\left(f_t - \hat{f}_t\right)^2\right\},$$

where f_t denotes the true signal value, $\hat{f}_t$ denotes the segmented constant or linear signal value reconstructed using estimated change points, T denotes the total length of the observed data, and $E[\cdot]$ denotes averaging over multiple simulation experiments to reduce the impact of random error.

To assess the performance of each method in terms of the accuracy of the estimated locations of the change-points, we report estimates of the (scaled) Hausdorff distance

$$d_H = T^{-1}E\left[max\left\{\max_{j=0,\ldots,N+1}\min_{l=0,\ldots,\widehat{N}+1}\left|\tau_j - \hat{\tau}_l\right|, \max_{l=0,\ldots,\widehat{N}+1}\min_{j=0,\ldots,N+1}\left|\hat{\tau}_l - \tau_j\right|\right\}\right],$$

where

$$0 = \tau_0 < \tau_1 < \cdots < \tau_N < \tau_{N+1} = T \text{ and } 0 = \hat{\tau}_0 < \hat{\tau}_1 < \cdots < \hat{\tau}_N < \hat{\tau}_{N+1} = T$$

denote true and estimated locations of the change-points, respectively. From the definition above, it follows that $0 \leq d_H \leq 1$. An estimator is regarded as performing well when its d_H is close to 0. However, d_H would be large when the number of change points is underestimated or some of the estimated change points are far from the real change-points.

Empirical results in Table 2 demonstrate that WNSP achieves competitive localization performance.

Table 2. Performance comparison of competing methods based on 100 simulations, showing the error distribution in change-point number ($\widehat{N} - N$), location accuracy ($d_H \times 10^2$), and MSE.

Model	Method	Results for the following values of $\widehat{N} - N$:							*MSE*	$d_H \times 10^2$
		-3≤	-2	-1	0	1	2	≥3		
M1	*WNSP*	0	0	3	94	3	0	0	0.026	0.46
	NOT	0	0	2	95	3	0	0	0.023	0.41
	WBS	0	1	3	90	4	2	0	0.032	0.49
	WBS2	0	1	2	93	3	1	0	0.021	0.42
	SaRa(C=1)	0	4	7	87	2	0	0	0.061	0.53
	SaRa(C=2)	1	5	9	84	1	0	0	0.054	0.50
	SaRa(C=3)	0	2	3	90	4	1	0	0.031	0.48
M2	*WNSP*	0	0	4	91	4	1	0	0.042	0.56
	NOT	0	0	5	90	4	1	0	0.040	0.51
	WBS	0	1	5	90	4	0	0	0.042	0.59
	WBS2	0	0	5	90	5	0	0	0.042	0.58
	SaRa(C=1)	0	0	17	73	9	1	0	0.086	0.93
	SaRa(C=2)	1	3	10	77	7	2	0	0.097	0.89
	SaRa(C=3)	0	1	6	81	9	3	0	0.078	0.62
M3	*WNSP*	0	1	5	93	2	0	0	0.057	0.50
	NOT	0	0	0	92	5	2	0	0.059	0.49
	WBS	0	0	2	91	4	3	0	0.059	0.51
	WBS2	0	0	1	92	3	3	1	0.059	0.50
	SaRa(C=1)	0	5	12	63	13	5	2	0.178	2.95
	SaRa(C=2)	0	3	9	74	11	3	0	0.135	1.49
	SaRa(C=3)	0	2	8	76	11	3	0	0.113	1.30
M4	*WNSP*	0	0	0	97	3	0	0	0.038	0.49
	NOT	0	0	1	98	1	0	0	0.039	0.44
M5	*WNSP*	0	0	1	94	5	0	0	0.052	0.52
	NOT	0	0	2	93	5	0	0	0.050	0.58
M6	*WNSP*	0	0	2	90	8	0	0	0.041	0.52
	NOT	0	0	2	92	6	0	0	0.040	0.50

Specifically, its Hausdorff distance is comparable to that of the NOT method on most models, though NOT shows a slight advantage on $M1, M2, M4,$ and $M6$. More importantly, a comprehensive evaluation must consider detection completeness. As detailed in Appendix A.3, WNSP achieves the highest recall rate across all models, meaning it is the most reliable method in avoiding missed detections. Consequently, its overall detection performance, as measured by the F1-score, is superior or comparable to NOT (see Table A1). In summary, WNSP consistently attains among the lowest MSE and Hausdorff distance values, outperforms WBS, WBS2 and SaRa, and matches NOT in localization while excelling in recall. The method shows no systematic bias and maintains robustness across diverse noise distributions, including Gaussian, heavy-tailed, and heteroscedastic errors. Its adaptive window strategy avoids sensitivity to preset bandwidths, eliminating the need for manual parameter tuning. Notably, compared with the latest WBS2 method, WNSP demonstrates comparable or superior Hausdorff distance on models $M1 - M3$, while maintaining its inherent advantage in achieving a higher recall rate, which underscores the effectiveness of its local refinement strategy within statistically significant intervals.

The simulation design focuses on comparing methods that provide point estimates. Since the NSP algorithm outputs significant intervals rather than point estimates, it is not evaluated on localization metrics (e.g., d_H), though WNSP's first stage is functionally equivalent to NSP. Additionally, the SaRa method is applied only to piecewise-constant models $(M1 - M3)$ because its core statistic, the local CUSUM, is ineffective for detecting the slope changes present in the piecewise-linear models $(M4 - M6)$.

5.3. Integrated performance and efficiency analysis

To preliminarily investigate the algorithm's sensitivity to errors in the public matrix $\hat{\Sigma}_B$, we conducted an additional test. Using Model M1, we compared the performance of WNSP when employing the precise matrix ($\varepsilon = 0$) versus a significantly perturbed matrix ($\varepsilon = 0.3$, defined in Appendix A.4). The results show that under significant perturbation, WNSP's Hausdorff distance ($d_H \times 10^2$) increased from 0.46 to 0.85, while the NOT method, used as a control, remained stable (0.41). This finding confirms that WNSP's performance can degrade when there is a substantial discrepancy between the public matrix and the true data-generating mechanism. Therefore, in practical applications, it is advisable to estimate $\hat{\Sigma}_B$ using macro data that is homogeneous with or from the same distribution as the target data to ensure optimal performance. A systematic sensitivity analysis covering various perturbation levels and models is a valuable direction for future work.

6. Real data analysis

6.1. Application to CPI data

In this section, we use the WNSP algorithm to analyze the time series of US ex-post real interest rates (the 3-month Treasury bill rate net of CPI inflation) in the United States considered by Garcia and Perron [33] and Bai and Perron [34]. The dataset is available at http://qed.econ.queensu.ca/jae/datasets/bai001/. The dataset contains quarterly observations from Q1 1961 to Q3 1986, totaling $T = 103$ points.

The goal is to identify structural change-points reflecting shifts in economic mechanisms, thereby

offering empirical support for monetary policy and macroeconomic analysis. Such changes may signal responses to policy adjustments, external shocks, or business cycle transitions, exemplified by the 1973 oil crisis and the 1979 Federal Reserve policy shift.

Using WNSP with parameters $M = 1000$ and $\alpha = 0.1$, and variance estimated via MAD, two significant intervals were detected: $Y_{(25:56)}$ and $Y_{(78:84)}$. Adaptive window widths were applied within these intervals, leading to change-point estimates at $\hat{\tau}^1 = 47$ and $\hat{\tau}^2 = 81$. These align with existing results: the first corresponds to the early-1973 oil shock (mean decrease), and the second to mid-1981 fiscal deficit surge (mean increase), as shown in Figure 3.

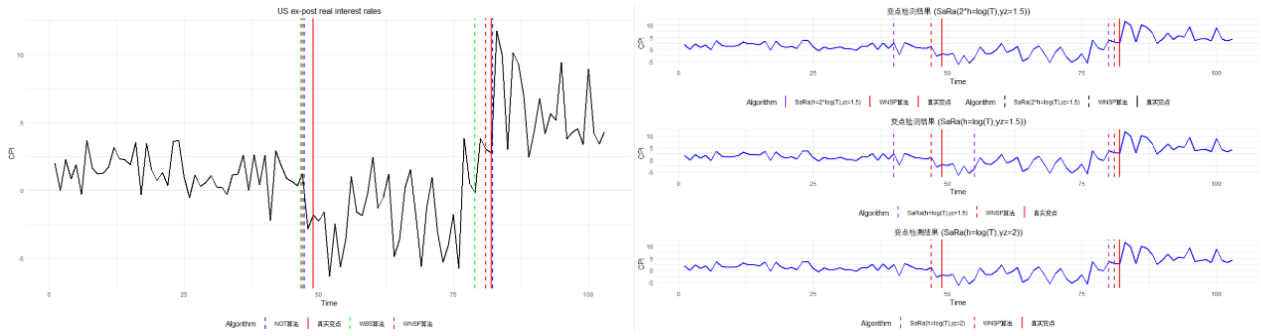

Figure 3. Comparison of experimental results between the WNSP algorithm and NOT/WBS algorithms on CPI data (left). Comparison of experimental results between the WNSP algorithm and SaRa algorithm (right).

Compared with other methods, the results reveal the following: (1) The SaRa algorithm is highly dependent on bandwidth selection, whereas WNSP can adaptively determine thresholds and bandwidth based on the data; (2) WNSP exhibits estimation performance comparable to the NOT algorithm and superior precision to other methods.

6.2. Application to UK house price index

This study analyzes the monthly percentage change in the House Price Index (HPI) for Tower Hamlets (Figure 4) and Hackney (Figure 5), UK, from January 1995 to October 2022. The data for this study were obtained from the UK Land Registry (http://landregistry.data.gov.uk/app/ukhpi). A comparison of change-point detection methods—WNSP, NOT, WBS, and SaRa—reveals that WNSP and NOT yield similar results, while SaRa is highly sensitive to bandwidth choice (e.g., $\log(T)$, $2\log(T)$, $3\log(T)$). All methods detected a significant change-point around March 2008 and September 2009, aligning with the peak of the global financial crisis. These breaks likely reflect the crisis' impact on the London property market, including falling prices and tighter credit conditions. The results illustrate how statistical approaches can capture market shock timing and urban housing dynamics during economic turmoil.

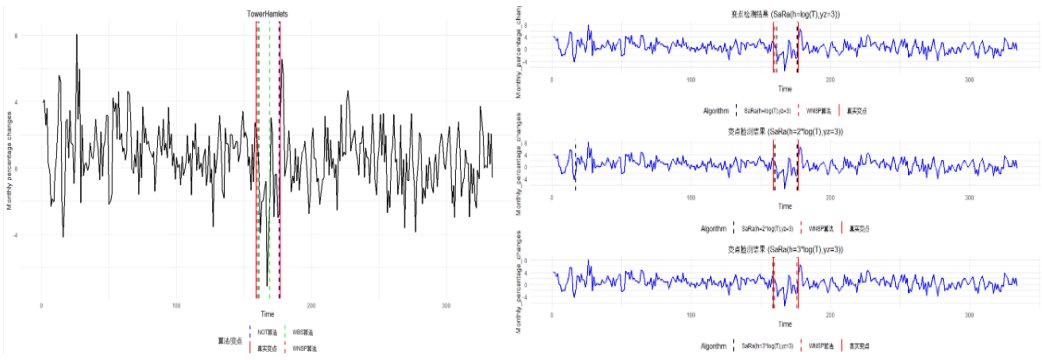

Figure 4. Comparison of experimental results of the WNSP algorithm with NOT and WBS for house price indicators in Tower Hamlets (left). Comparison of experimental results of the WNSP algorithm with the SaRa algorithm (right).

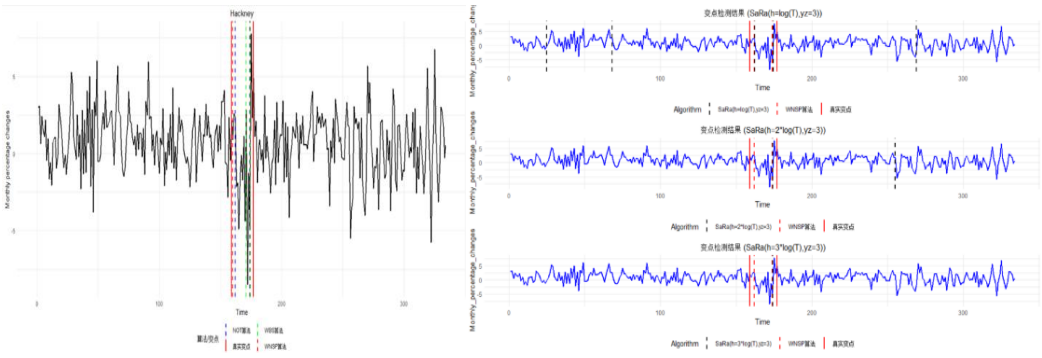

Figure 5. Comparison of experimental results of the WNSP algorithm with NOT and WBS for house price indicators in Hackney (left). Comparison of experimental results of the WNSP algorithm with the SaRa algorithm (right).

7. Conclusions

This paper proposed a WNSP algorithm, which effectively addresses the challenge of accurately estimating change-point locations while preserving the statistical reliability of the NSP framework. By extending the "post-inference selection" paradigm into a two-stage process of interval detection and localized point estimation, WNSP achieves both rigorous coverage guarantees and high localization accuracy. Theoretical analysis confirms that the algorithm possesses consistency and finite-sample coverage properties. Extensive simulations and real-data applications, including economic indicators and housing market series, demonstrate that the WNSP performs robustly under various noise distributions and structural complexities. It outperforms competing methods in terms of accuracy and adaptability without requiring manual bandwidth selection.

The WNSP algorithm performs well when change-points are widely spaced but struggles in dense scenarios where the spacing is smaller than required by Assumption 1. In such cases, the NSP step may return fewer intervals than the actual number of change-points, with some intervals containing multiple change-points. This occurs because NSP guarantees only that each significant interval covers at least one change-point but does not ensure their complete separation.

Future work will focus on several directions: (a) extending the WNSP framework to multivariate, online, and nonlinear settings, (b) conducting a more comprehensive, cross-scenario systematic performance comparison with the latest change-point detection methods, such as WBS2, and (c) exploring strategies to relax Assumption 1 for better performance in dense change-point scenarios.

Author contributions

Xiaoyuan Zhang: Formal analysis, Writing–original draft; Zhanshou Chen: Methodology, Writing–review & editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.12161072) and the Natural Science Foundation of Qinghai Province (No.2024-ZJ-933).

Conflict of interest

The authors declare no conflict of interest.

References

1. G. B. Pezzatti, T. Zumbrunnen, M. Bürgi, P. Ambrosetti, M. Conedera, Fire regime shifts as a consequence of fire policy and socio-economic development: An analysis based on the change point approach, *Forest Policy Econom.*, **29** (2013), 7–18. https://doi.org/10.1016/j.forpol.2011.07.002
2. K. J. Oh, I. Han, Using change-point detection to support artificial neural networks for interest rates forecasting, *Expert Syst. Appl.*, **19** (2000), 105–115. https://doi.org/10.1016/s0957-4174(00)00025-7
3. C. Contal, J. O'Quigley, An application of changepoint methods in studying the effect of age on survival in breast cancer, *Comput. Stat. Data Anal.*, **30** (1999), 253–270. https://doi.org/10.1016/s0167-9473(98)00096-6
4. W. Gu, J. Choi, M. Gu, H. Simon, K. Wu, Fast change point detection for electricity market analysis, In: *2013 IEEE International Conference on Big Data*, **2** (2013), 50–57. https://doi.org/10.1109/bigdata.2013.6691733
5. D. Jarušková, Some problems with application of change-point detection methods to environmental data, *Environm. Official J. Int. Environm. Soc.*, **8** (1997), 469–483. https://doi.org/10.1002/(SICI)1099-095X(199709/10)8:5%3C469::AID-ENV265%3E3.0.CO;2-J

6. J. Chen, A. K. Gupta, A. Gupta, *Parametric Statistical Change Point Analysis*, Boston: Birkhäuser, 2000. https://doi.org/10.1007/978-1-4757-3131-6
7. L. Horváth, G. Rice, *Change Point Analysis for Time Series*, New York: Springer, 2024. https://doi.org/10.1007/978-3-031-51609-2_2
8. D. M. Hawkins, Fitting multiple change-point models to data, *Comput. Stat. Data Anal*., **37** (2001), 323–341. https://doi.org/10.1016/s0167-9473(00)00068-2https://doi.org/10.1016/s0167-9473(00)00068-2
9. J. Pan, J. Chen, Application of modified information criterion to multiple change point problems, *J. Multivar. Anal*., **97** (2006), 2221–2241. https://doi.org/10.1016/j.jmva.2006.05.009
10. Y. C. Yao, S. T. Au, Least-squares estimation of a step function, *Sankhyā: Indian J. Stat.*, **51** (1989), 370–381.
11. Z. Harchaoui, C. Lévy-Leduc, Multiple change-point estimation with a total variation penalty, *J. Amer. Stat. Assoc.*, **105** (2010), 1480–1493. https://doi.org/10.1198/jasa.2010.tm09181
12. Q. Li, L. Wang, Robust change point detection method via adaptive LAD-LASSO, *Stat. Papers*., **61** (2020), 109–121. https://doi.org/10.1007/s00362-017-0927-3
13. R. Killick, P. Fearnhead, I. A. Eckley, Optimal detection of changepoints with a linear computational cost, *J. Amer. Stat. Assoc.*, **107** (2012), 1590–1598. https://doi.org/10.1080/01621459.2012.737745
14. L. Y. Vostrikova, Detecting "disorder" in multidimensional random processes, *Russian Acad. Sci.*, **259** (1981), 270–274. https://doi.org/10.1007/978-1-4614-2386-7_3
15. A. B. Olshen, E. S. Venkatraman, R. Lucito, M. Wigler, Circular binary segmentation for the analysis of array-based DNA copy number data, *Biostatistics*., **5** (2004), 557–572. https://doi.org/10.1093/biostatistics/kxh008
16. P. Fryzlewicz, Wild binary segmentation for multiple change-point detection, *Ann. Statist.*, **42** (2014), 2243–2281. https://doi.org/10.1214/14-aos1245
17. P. Fryzlewicz, Detecting possibly frequent change-points: Wild Binary Segmentation 2 and steepest-drop model selection, *J. Korean Stat. Soc.*, **49** (2020), 1027–1070. https://doi.org/10.1007/s42952-020-00060-x
18. R. Baranowski, Y. Chen, P. Fryzlewicz, Narrowest-over-threshold detection of multiple change points and change-point-like features, *J. Royal Stat. Soc. Ser. B: Stat. Methodol*., **81** (2019), 649–672. https://doi.org/10.1111/rssb.12322
19. S. Kovács, P. Bühlmann, H. Li, A. Munk, Seeded binary segmentation: A general methodology for fast and optimal changepoint detection, *Biometrika*., **110** (2023), 249–256. https://doi.org/10.1093/biomet/asac052
20. K. K. Korkas, Ensemble binary segmentation for irregularly spaced data with change-points, *J. Korean Stat. Soc.*, **51** (2022), 65–86. https://doi.org/10.1007/s42952-021-00120-w
21. Z. Chen, Y. Hu, Cumulative sum estimator for change-point in panel data, *Stat. Papers*, **58** (2017), 707–728. https://doi.org/10.1007/s00362-015-0722-y
22. H. Cho, P. Fryzlewicz, Multiple-change-point detection for high dimensional time series via sparsified binary segmentation, *J. Royal Stat. Soc. Ser. B: Stat. Methodol.*, **77** (2015), 475–507. https://doi.org/10.1111/rssb.12079
23. C. Truong, L. Oudre, N. Vayatis, Selective review of offline change point detection methods, *Signal Proc*., **167** (2020), 107–299. https://doi.org/10.1016/j.sigpro.2019.107299

24. H. Cho, C. Kirch, Data segmentation algorithms: Univariate mean change and beyond, *Econom. Stat.*, **30** (2024), 76–95. https://doi.org/10.1016/j.ecosta.2021.10.008
25. X. Shi, C. Gallagher, R. Lund, R. Killick, A comparison of single and multiple changepoint techniques for time series data, *Comput. Stat. Data Anal.*, **170** (2022), 107–433. https://doi.org/10.1016/j.csda.2022.107433
26. Y. S. Niu, H. Zhang, The screening and ranking algorithm to detect DNA copy number variations, *Ann. App. Stat*., **6** (2012), 1306–1326. https://doi.org/10.1214/12-aoas539
27. S. Jun Shin, Y. Wu, N. Hao, A backward procedure for change-point detection with applications to copy number variation detection, *Canad. J. Stat*., **48** (2020), 366–385. https://doi.org/10.1002/cjs.11535
28. C. Y. Yau, Z. Zhao, Inference for multiple change points in time series via likelihood ratio scan statistics, *J. Royal Stat. Soc. Ser. B: Stat. Methodol*., **78** (2016), 895–916. https://doi.org/10.1016/j.econlet.2019.03.017
29. B. Eichinger, C. Kirch, A MOSUM procedure for the estimation of multiple random change points, *Bernoulli*, **24** (2018), 526–524. https://doi.org/10.3150/16-BEJ887
30. Z. Chen, Q. Xu, H. Li, Inference for multiple change points in heavy-tailed time series via rank likelihood ratio scan statistics, *Econom. Lett*., **179** (2019), 53–56. https://doi.org/10.3150/16-bej887
31. A. Anastasiou, P. Fryzlewicz, Detecting multiple generalized change-points by isolating single ones, *Metrika*, **85** (2022), 141–174. https://doi.org/10.1007/s00184-021-00821-6
32. P. Fryzlewicz, Narrowest significance pursuit: inference for multiple change-points in linear models, *J. Amer. Stat. Assoc.*, **119** (2024), 1633–1646. https://doi.org/10.1080/01621459.2023.2211733
33. R. Garcia, P. Perron, An analysis of the real interest rate under regime shifts, *Review Econom. Stat.*, **78** (1996), 111–125. https://doi.org/10.2307/2109851
34. J. Bai, P. Perron, Computation and analysis of multiple structural change models, *J. Appl. Economet.*, **18** (2003), 1–22. https://doi.org/10.1002/jae.659

Appendix

A.1. Proof of Lemma 1

Proof. Assume initially that f_t has a single change-point τ_1. As NSP considers all intervals by the assumption of the theorem, it will certainly consider intervals symmetric to the true change-point, i.e., $[\tau_1 - d + 1, \tau_1 + d]$, for all appropriate d. In a piecewise-constant signal, there is an explicit formula for the deviation measure $D[s,e]$ on any interval $[s,e]$, given by

$$D[s,e] = \max_{\eta\in\{1,\dots,e-s+1\}} \frac{1}{2\sqrt{\eta}} \left(\max_{s_1\in\{s,\dots,e+1-\eta\}} \sum_{t=s_1}^{s_1+\eta-1} Y_t - \min_{s_1\in\{s,\dots,e+1-\eta\}} \sum_{t=s_1}^{s_1+\eta-1} Y_t \right). \quad (A1)$$

Without loss of generality, assume

$$f_{\tau_1} > f_{\tau_1+1}.$$

Representation (A1) implies

$$D_{[\tau_1-d+1,\tau_1+d]} \geq \frac{1}{2\sqrt{d}}\left(\max_{s_1\in\{[\tau_1-d+1,\ldots,\tau_1+d\}}\sum_{t=s_1}^{s_1+d-1} Y_t - \min_{s_1\in\{[\tau_1-d+1,\ldots,\tau_1+d\}}\sum_{t=s_1}^{s_1+d-1} Y_t\right)$$

$$\geq \frac{1}{2\sqrt{d}}\left(\sum_{t=\tau_1-d+1}^{\tau_1} Y_t - \sum_{t=\tau_1+1}^{\tau_1+d} Y_t\right) \geq \frac{1}{2}\left|f_{\tau_1+1}-f_{\tau_1}\right|\sqrt{d} - \| \varepsilon \|_I$$

$$\geq \frac{1}{2}\left|f_{\tau_1+1}-f_{\tau_1}\right|\sqrt{d} - \| \varepsilon \|_I. \quad (A2)$$

On the set $\| \varepsilon \|_I \leqslant \lambda_\alpha$, (2) is further bounded from below by $\frac{1}{2}\left|f_{\tau_1+1}-f_{\tau_1}\right|\sqrt{d}-\lambda_\alpha$. From the definition of the NSP algorithm, detection on $[s,e]$ is triggered by the event $D[s,e]>\lambda_\alpha$, so detection on $[\tau_1-d+1,\tau_1+d]$ is triggered if (note: not "only if", as we are using lower bounds here) $\frac{1}{2}\left|f_{\tau_1+1}-f_{\tau_1}\right|\sqrt{d}-\lambda_\alpha>\lambda_\alpha$, or

$$\left|f_{\tau_1+1}-f_{\tau_1}\right|\sqrt{d} > 4\lambda_\alpha. \quad (3)$$

As NSP looks for the shortest intervals of detection, the NSP interval of significance around τ_1 will definitely be no longer than $2d=|[\tau_1-d+1,\tau_1+d]|$. However, from (3), it is sufficient for detection to be triggered if $d>\frac{16\lambda_\alpha^2}{\left|f_{\tau_{j+1}}-f_{\tau_j}\right|^2}$. This shows that the maximum length of an NSP interval of significance will not exceed $2\bar{d}$, where $\bar{d}=\left\lceil\frac{16\lambda_\alpha^2}{\left|f_{\tau_{j+1}}-f_{\tau_j}\right|^2}\right\rceil+1$. We now turn our attention to the multiple change-point case. For each change-point τ_j, define its corresponding $\bar{d}_j$ as in formula (7) of the main paper. Recall that we are on the set $\| \varepsilon \|_I \leqslant \lambda_\alpha$. Note first that even though the NSP interval of significance around τ_j is guaranteed to be of length at most $2\bar{d}_j$, it will not necessarily be a subinterval of $[\tau_j-\bar{d}_j+1,\tau_j+\bar{d}_j]$ (as NSP simply looks for the shortest intervals of significance, and interval symmetry around the true change-point is not explicitly promoted). Therefore, to ensure that an interval detection around τ_j does not interfere with detections around τ_{j-1} and τ_{j+1}, the distances τ_j-τ_{j-1} and τ_{j+1}-τ_{j-1} must be suitably long, but this is guaranteed by Assumption 1 from the main paper. This completes the proof.

As an aside, note in addition that in the Gaussian case $\varepsilon\sim N(0,\ 1)$, Theorem 1.3 in Kabluchko (2007) implies $\lambda_\alpha=O\left(\log^{1/2}T\right)$; in fact, for $\alpha=0.05$, we have $\lambda_\alpha\leq 1.33\sqrt{2\log T}$ for $T\geq 100$, and for $\alpha=0.1$, we have $\lambda_\alpha\leq 1.25\sqrt{2\log T}$ over the same range of T.

A.2. Proof of Theorem 1

Proof. The proof consists of two main steps. First, we leverage the result of Lemma 1, which guarantees that the coarse NSP algorithm returns intervals $[s_j,e_j]$ such that τ_j in $[s_j,e_j]$ and $L_j\equiv e_j-s_j+1=O(1)$ (or more precisely, $L_j\leq 2\bar{d}_j$). Second, we focus on a single generic significant interval $[s,e]$ containing one true change-point τ and show that the WNSP estimator $\hat{\tau}$ applied to this interval satisfies $|\hat{\tau}-\tau|=O_p(\log T)$.

Consider a piecewise constant model for simplicity (extension to piecewise linear is similar):

$$Y_t = f_t + \varepsilon_t, \qquad f_t = \begin{cases} \mu_1 \, for \, t < \tau, \\ \mu_2 \, for \, t \geq \tau, \end{cases}$$

With $\Delta = \mid \mu_2 - \mu_1 \mid > 0$. The error terms ε_t satisfy the general conditions specified in Model (1) of the main text. Crucially, for our theoretical analysis, we rely on Assumption 2, which states that the length of the significant interval $L = e - s + 1 \geq \psi_T \to \infty$.

Within $[s, e]$, the WNSP algorithm computes the local discrepancy statistic for each candidate point $\eta \in [s + h, e - h + 1]$ (this range ensures that the sliding windows remain within the buffered data). The adaptive window width is set to $h = \lfloor 0.1 * (e_k - s_k + 1) \rfloor$. The statistic is defined as:

$$D(\eta) = \mid \hat{\mu}_L(\eta) - \hat{\mu}_R(\eta) \mid,$$

where

$$\hat{\mu}_L(\eta) = \frac{1}{h} \sum_{t=\eta-h}^{\eta-1} Y_t \,, \qquad \hat{\mu}_R(\eta) = \frac{1}{h} \sum_{t=\eta}^{\eta+h-1} Y_t.$$

The estimator is $\hat{\tau} = argmax_{\{\eta \in [s,e]\}} D(\eta)$.

Assume $s + h \leq \tau \leq e - h + 1$ (this holds with high probability due to buffer and Lemma 1). Then, $E[D(\tau)] = \mid E[\hat{\mu}_L(\tau)] - E[\hat{\mu}_R(\tau)] \mid = \mid \mu 1 - \mu 2 \mid = \Delta$.

To control the deviation of $D(\tau)$ from its mean, we employ a moment-based approach. Under the general error structure, there exists a constant $C_1 > 0$, dependent on the moments of ε_t, such that $\mathrm{Var}\big(D(\tau)\big) \leq C_1 / w$. Applying Chebyshev's inequality, $P(\mid D(\tau) - \Delta \mid \geq u) \leq \frac{C_1}{hu^2}$.

Choosing $u = \Delta/2$, we obtain:

$$P\left(D(\tau) \leq \frac{\Delta}{2}\right) \leq \frac{4C_1}{h\Delta^2}. \tag{A4}$$

Since $h \geq \lfloor \psi_T / 2 \rfloor \to \infty$ by Assumption 2, this probability decays to zero. Thus, $D(\tau) \geq \Delta/2$ with high probability.

Next, we examine points η far from τ, i.e., $\mid \eta - \tau \mid \geq \delta$, and bound $P(D(\eta) > \Delta/2)$.

Case 1: $\mid \eta - \tau \mid \geq h$. In this case, both sliding windows $W_L(\eta)$ and $W_R(\eta)$ lie entirely within a single constant segment. Consequently, $\mathbb{E}[D(\eta)] = 0$. There exists a constant $C_2 > 0$ such that:

$$P\left(D(\eta) > \frac{\Delta}{2}\right) \leq \frac{4\mathrm{Var}\big(D(\eta)\big)}{\Delta^2} \leq \frac{C_2}{h\Delta^2}. \tag{A5}$$

Case 2: $0 < \mid \eta - \tau \mid < h$. This is the critical case, where one window contains data from both segments. Consider $\eta = \tau + d$ with $0 < d < h$. The expectations are:

$$\mathbb{E}[\hat{\mu}_L(\eta)] = \frac{h-d}{h}\mu_1 + \frac{d}{h}\mu_2, \mathbb{E}[\hat{\mu}_R(\eta)] = \mu_2,$$

which yields:

$$\mathbb{E}[D(\eta)] = \mid \frac{h-d}{h}\mu_1 + \frac{d}{h}\mu_2 - \mu_2 \mid = \left(1 - \frac{d}{h}\right)\Delta < \Delta.$$

Again, using moment bounds, there exists a constant $C_3 > 0$ such that:

$$P\left(D(\eta) > \frac{\Delta}{2}\right) \leq \frac{C_3}{h\Delta^2}. \tag{A6}$$

Let A be the event in which the maximum of $D(\eta)$ occurs at a point distant from τ: $A = \exists\, \eta$ with $|\eta - \tau| > M$ such that $D(\eta) \geq D(\tau)$. Then, $P(|\hat{\tau} - \tau| > M) \leq P(A)$.

From (4), $D(\tau) \geq \Delta/2$ with high probability. From (5) and (6), for any η with $|\eta - \tau| > M$, we have: $P\left(D(\eta) \geq \frac{\Delta}{2}\right) \leq \frac{K}{h\Delta^2}$, where $K = \max C_2, C_3$. The number of such candidate points η in the interval is at most $L = O(1)$. Applying a union bound, $P(A) \leq L \cdot \frac{K}{h\Delta^2}$.

By Assumption 2, $h \asymp \psi_T \to \infty$, where the notation $\asymp$ indicates that h and ψ_T are of the same asymptotic order. We can therefore choose the sequence ψ_T such that $\psi_T \geq \frac{2LK}{\Delta^2}\log T$. This implies: $P(A) \leq \frac{1}{2\log T} \to 0$ as $T \to \infty$.

Thus, $|\hat{\tau} - \tau| \leq M$ with high probability, where $M \asymp h \asymp \log T$.

Finally, applying this result uniformly across all N change-points (where N is fixed), and using a union bound,

$$P\left(\max_{1\leq j\leq N} \left|\hat{\tau}_j - \tau_j\right| > C\log T\right) \leq \sum_{j=1}^{N} P(|\hat{\tau}_l - \tau_j| > C\log T) \leq \frac{N}{2\log T} \to 0.$$

This completes the proof, showing that the WNSP estimator consistently locates change-points with an error rate of $O(\log T)$.

A.3. Detection completeness and F1-score analysis

To comprehensively evaluate the change-point detection performance of the WNSP algorithm beyond localization accuracy ($d_H \times 10^2$), we introduce matching-based detection completeness metrics. These metrics require matching each estimated change-point $\hat{\tau}_l$ to a true change-point τ_j. The matching follows the commonly used tolerance window approach: if $|\hat{\tau}_l - \tau_j| \leq \delta$, the estimated change-point is considered to have correctly detected the true change-point τ_j, where the tolerance δ is set to $\lfloor 0.02T \rfloor$ (i.e., 2% of the sample size).

Based on this matching, we define:

True positive (TP): The number of correctly matched estimated change-points.

False positive (FP): The number of estimated change-points does not match any true change-point.

False negative (FN): The number of true change-points does not match any estimated change-point.

The following metrics are then calculated:

Precision: $P = \frac{TP}{TP+FP}$;

Recall: $R = \frac{TP}{TP+FN}$;

F1-score: $F1 = 2 \cdot (P \cdot R)/(P + R)$;

Precision measures the reliability of the detections (what fraction of alarms are real), recall measures the completeness of detection (what fraction of true change-points are found), and the F1-score is their harmonic mean, serving as a comprehensive metric for overall detection performance.

Table A1. Comparison of detection completeness metrics for different methods (mean ± standard deviation, based on 100 simulations).

Model	Method	Precision	Recall	F1 − score
M1	*WNSP*	0.96	0.98	0.97
	NOT	0.97	0.96	0.96
	WBS	0.94	0.95	0.94
	WBS2	0.95	0.96	0.95
M2	*WNSP*	0.93	0.97	0.95
	NOT	0.95	0.93	0.94
	WBS	0.92	0.94	0.93
	WBS2	0.93	0.95	0.94
M3	*WNSP*	0.92	0.96	0.94
	NOT	0.94	0.92	0.93
	WBS	0.90	0.93	0.91
	WBS2	0.94	0.95	0.94
M4	*WNSP*	0.97	0.99	0.98
	NOT	0.98	0.97	0.97
M5	*WNSP*	0.94	0.98	0.96
	NOT	0.96	0.95	0.95
M6	*WNSP*	0.95	0.97	0.96
	NOT	0.96	0.96	0.96

Note: For brevity, the SaRa method is not included in this table as it is not applicable to piecewise linear models (M4–M6) and generally underperforms WNSP, NOT, and WBS on piecewise constant models (see Table 2 in the main text).

A.4. Method for generating perturbations to the public matrix

The "significantly perturbed matrix" mentioned in this appendix and Section 5.3 of the main text is generated as follows: Let Σ be the true covariance matrix. Generate a random symmetric matrix E of the same dimension as Σ, with its independent and identically distributed (i.i.d.) elements following the standard normal distribution $N(0,1)$. The perturbed public matrix is then defined as:

$$\hat{\Sigma}_B = \Sigma + \epsilon \cdot \frac{E}{\|E\|_F} \| \Sigma \|_F,$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and ϵ is the perturbation intensity coefficient. The value $\epsilon = 0.3$ used in the main text represents a relatively strong level of perturbation.