

A Robust Online Location–Scale Change-Point Detector under Dynamic Huber Contamination

Loaiy F. Naji

Received: 10 August 2026 / Accepted: 18 September 2026

© The Author(s) 2026

Abstract

Online change-point detection must react rapidly to structural changes while controlling false alarms, a task that becomes substantially harder when streaming observations are subject to time-varying adversarial contamination and when the change may affect both location and scale. This paper develops an adaptive robust online location–scale detector under dynamic Huber contamination. The method uses two bounded score channels—a Huber location score and a centered winsorized quadratic scale score—and aggregates them over multiple window lengths and clipping constants. The theoretical model allows the contaminating marginal law to vary arbitrarily with time while retaining independence across observations; it does not assume a history-adaptive adversary. Bounded scores are used in this paper to give an anytime false-alarm bound for the known-baseline detector, and a deterministic perturbation proposition is established to translate high-probability location/scale estimation into an explicit score-centering allowance δ_m for the estimated-baseline detector, for which a false-alarm bound is given by an estimated-baseline false-alarm theorem. Local sensitivity can be used to give first-order separation between location and log-scale, and a high-probability detection-delay theorem is also given to make the adversarial detectability boundary explicit. Numerical studies are performed using 2,000 replications for each alternative and 5,000 null replications for calibration. Additional experiments, including matched-calibration, contamination-law, heavy-tail and operational-tuning studies, are also provided. For example, dynamically switching gross contamination at 10% was found to result in the proposed detector using a baseline allowance maintaining a pre-change false-alarm proportion of 0.056 on average, as opposed to 0.423 for a clean-calibrated unbounded rolling location–scale detector; post-change detection proportions were found to be 0.937 for location changes, 0.858 for joint changes, and a low 0.038 for pure scale changes – revealing a trade-off between robustness and efficiency. Sensitivity experiments were also performed, including the effect of underestimating the contamination envelope leading to high false-alarm rates, and that using larger reference samples for estimation of scale and joint changes results in materially better performance. Finally, the tuning of detectors is found to result in a loss of power through multiplicity, when using unnecessarily large grids. The proposed detector was found to signal the change in location on-line in 1906, as opposed to the unbounded location–scale comparator which signaled in 1913; and a single gross outlier, added to the data in 1895, was found to trigger the comparator immediately but to result in no alarm being triggered by the proposed detector. The proposed methodology therefore fills a significant methodological gap between the dynamic-contamination change-point model and that of joint online location–scale monitoring, and the assumptions made, the consequences of tuning, and the limitations of the proposed approach are all made explicit. Keywords: online change-point detection; Huber contamination; robust statistics; location-scale model; bounded scores; multiresolution monitoring; false-alarm control; detection delay.

Keyword: Online Change-Point Detection, Huber Contamination, Robust Statistics, Location–Scale Model, Bounded Scores, Multiresolution Monitoring, False-Alarm Control, Detection Delay

Introduction

Change-point detection is the task of detecting time points at which a sequence's distribution changes. Monitoring sequences online means that observations arrive one by one, and a decision whether a change



occurred at each time point has to be made online. A key trade-off between different change-point detection methods is false-alarm rate vs. detection delay. CUSUM-type tests, generalized likelihood-ratio tests Building upon these results, online change detection methods and related sequential procedures can become unstable, even lose their theoretical guarantees, in presence of misspecification, outliers, heavy tails or even manipulated observations [1], [2], [3], [4], [5], [6]. Robust statistics provides a natural language for such deviations. Huber's contamination model represents an observed law as a mixture of an inlier distribution and an arbitrary contaminating distribution [7], [8]. In streaming data a static contaminating law is often unrealistic because the contaminating mechanism can change from one time point to another. Li and Yu [9] formalized adversarial change-point detection under a dynamic Huber contamination model that allows the contaminating distributions to differ over time. While Tang et al. [10] have provided online theory for robust mean change detection in presence of dynamic Huber contamination and heavy-tailed inliers, and provided regime-dependent detection-delay bounds,

There is, however, a significant body of work that models changes in both location and scale. These models often use self-starting CUSUM charts, and can be based on likelihoods for models with shifts in means and/or variances [1], [2]. Distribution-free sequential rank CUSUMs have also been developed for location and scale change-point detection in symmetric distributions [11]. Offline non-parametric methods have also been developed to handle models with multiple location and scale changes [12]. Robust online alternatives also include bounded-loss changepoint methods [13], robust generalized-Bayesian online detection [14], distributionally robust quickest change detection under Wasserstein uncertainty sets [15], and online kernel CUSUM procedures [16]. These approaches are important, but they do not directly provide joint location–scale monitoring under a time-varying arbitrary Huber contaminant.

The present paper bridges these strands. Each standardized observation is mapped to two bounded score channels: a Huber score for location and a winsorized quadratic score for scale. Boundedness allows us to set a worst-case contamination budget for the Separate channel. The use of recent maxima of a channel, and the values of the corresponding clipping constants, enables multiresolution behavior while avoiding the unnecessary $|C||D|$ multiplicity that would be caused by a Cartesian product of the independent location and scale grids for channels C and data D . independent location and scale tuning grids.

Research gap

This literature intersection is underdeveloped, dynamic-Huber online theory currently treats robust, mean changes rather than a joint location–scale parameter [10]. Classical online location–scale schemes detect mean and variance shifts but are largely formulated under parametric or comparatively clean process assumptions [1], [2]. Robust online methods based on bounded losses or generalized Bayesian updating address outliers and model mismatch [13], [14], robust quickest detection treats ambiguity sets for unknown post-change laws [15]; neither framework is the same as a dynamic Huber contaminant that may change adversarially at every time point. every time point. Online kernel CUSUM is nonparametric and computationally efficient [16], but it does not provide the explicit contamination-budget mechanism studied here. The gap is then the absence of a procedure that is simultaneously online, location and scale joint, robust to time-varying Huber contamination and adaptive over monitoring and clipping scales.

Table 1. Position of the proposed framework relative to representative literature.

Literature strand	Online	Location + scale	Dynamic Huber contamination	Adaptive / multiscale
Hawkins [1]; Hawkins & Zamba [2]	Yes	Yes	No	Limited
Lombard & Van Zyl [11]	Yes	Yes	No	Rank-based
Fearnhead & Rigaill [13]	Algorithmically usable online	General changes	Outliers, not dynamic Huber	Bounded loss
Li & Yu [9]	Primarily offline theory	Mean	Yes	No
Altamirano et al. [14]	Yes	General changepoints	No	Robust Bayesian
Xie et al. [15]	Yes	Distribution change	No; Wasserstein ambiguity	Minimax robust
Tang et al. [10]	Yes	Mean only	Yes	Robust mean procedure
Wei & Xie [16]	Yes	General distribution	No	Kernel / window adaptive
Proposed framework	Yes	Yes	Yes	Yes

1.2 Contributions

1. A dynamic Huber location–scale model is formulated in which the inlier parameters may change structurally while the arbitrary contaminating distribution may vary with time and the contamination fraction is bounded by a conservative envelope.
2. A two-channel bounded-score monitor is introduced. The location channel uses a normalized Huber score and the scale channel uses a centered, winsorized quadratic score. Both are of magnitude limited by one.
3. Adaptivity is operational: location and scale are separate statistics aggregated over dyadic monitoring windows and finite clipping grids, with multiplicity absorbed into one threshold.
4. We show that the idealized known-baseline detector provides an anytime false-alarm guaranty. For estimated baselines, a second theorem offers unconditional guaranties up to the probability of failure of a baseline-good event when the monitoring budget is increased by δ_m .
5. We present a local identifiability calculation that shows that the location and scale channels are first order orthogonal under symmetric inliers.
6. A high-probability detection-delay bound identifies the effective robust score signal after adversarial masking and shows explicitly when detection is statistically impossible for the proposed bounded-score mechanism.
7. The pilot study is substituted by an extended Monte Carlo study featuring matched and clean-reference calibration, various contamination laws, heavy-tailed inliers, Monte Carlo uncertainty reporting and a real-data application to the Nile River series.

Materials and Method

Review of Related Literature

Robustness of Huber and adaptive robustification

The classical minimax view of robust location estimation is due to Huber [7] and robust likelihood-ratio testing under contamination is due to Huber [8]. Robust scale estimation was further developed by Martin and Zamar

[17] with emphasis on control of the maximum bias of scale functionals. Chen, Gao and Ren [18] gave a decision-theoretic treatment of Huber's ε -contamination model, connecting attainable rates to the contamination proportion. Adaptive Huber methodology shows that clipping should respond to statistical difficulty rather than be treated as universally fixed [19]. These results motivate bounded scores and a finite grid of robustification constants.

Online and robust change point detection

Classical sequential detection is founded on CUSUM, generalized likelihood ratios and other stopping rules. Modern work provides nonasymptotic delay analyzes and computationally efficient monitoring [4], [5], [6], [20]. Cao and Xie [3] studied robust sequential detection with convex optimization. Fearnhead and Rigai [13] showed how bounded losses can avoid isolated arbitrarily large outliers creating spurious change points. Li and Yu [9] introduced adversarially robust changepoint theory under dynamic Huber contamination, and Tang et al. [10] extended this direction to online robust mean change detection with heavy-tailed inliers. Robust online changepoint work has also developed through generalized Bayesian updating [14], while distributionally robust quickest detection uses Wasserstein uncertainty sets for unknown post-change distributions [15]. Wei and Xie [16] provide a recent online kernel CUSUM with analytic ARL and expected-delay approximations. The present work is different in that it focuses on joint location and scale under the dynamic Huber mechanism itself.

Change point detection of location-scale models

Hawkins [1] developed self-starting CUSUM charts for location and scale, and Hawkins and Zamba [2] developed a changepoint formulation that gives a single diagnostic for a shift in mean, variance or both. Lombard and Van Zyl [11] developed distribution-free signed sequential rank CUSUMs for location and scale in symmetric distributions, and Ross [12] studied offline multiple location–scale changes using wild binary segmentation. These methods motivate the two-parameter monitoring objective but do not explicitly control a time-varying Huber contaminant.

Model and Problem Formulation

Let $X_1, X_2, \dots$ be independent observations that arrive sequentially. where the distribution at time t is modelled using the dynamic Huber location-scale model

$$Q(t) = [1 - \varepsilon(t)] F[\mu(t), \sigma(t)] + \varepsilon(t) H(t), \quad 0 \leq \varepsilon(t) \leq \bar{\varepsilon} < 1/2 \quad (1)$$

$$F[\mu, \sigma](x) = G((x - \mu)/\sigma), \quad \sigma > 0. \quad (2)$$

Here G is a standardized inlier distribution, $\mu(t)$ is the inlier location, $\sigma(t)$ is the inlier scale, $H(t)$ is an arbitrary contaminating distribution that may vary with t , and $\bar{\varepsilon}$ is a conservative contamination envelope. The scientific parameter is $\theta(t) = (\mu(t), \sigma(t))$, not the mean or variance of the contaminated mixture, because $H(t)$ is unrestricted. When we say “dynamic” or “adversarial” contamination in the theoretical development, we mean that the sequence of marginal laws $Q(t)$, and therefore $H(t)$ and $\varepsilon(t)$, may depend on t in an arbitrary way subject to $\varepsilon(t) \leq \bar{\varepsilon}$, whereas $X_1, X_2, \dots$ are independent under the corresponding product law. Thus the contaminating marginal can be chosen to switch signs or to oppose the post-change direction, as in the masking experiments, but $H(t)$ is not allowed to depend on the realized monitoring history or on the current detector state. This distinction is required by the Hoeffding concentration argument in Section 5.3. A history-adaptive contaminant would require conditional or martingale concentration and is outside the present guarantees.

Hypotheses

$$H_0: (\mu(t), \sigma(t)) = (\mu_0, \sigma_0) \text{ for all } t. \quad (3)$$

$$H_1: \text{there exists } \tau < \infty \text{ such that } (\mu(t), \sigma(t)) = (\mu_0, \sigma_0) \text{ for } t \leq \tau \text{ and } (\mu_1, \sigma_1) \text{ for } t > \tau. \quad (4)$$

$$(\mu_1, \sigma_1) \neq (\mu_0, \sigma_0). \quad (5)$$

The alternatives are pure location shift, pure scale shift and joint shift. The online goal is to build a stopping time $\hat{t}$ controlling false alarms and having small delay $(\hat{t} - \tau)_+$.

The need for an upper bound on contamination

In the case of free $H(t)$ and no upper bound on $\varepsilon(t)$, structural change and contamination are not identifiable: an adversary can duplicate or negate an arbitrary proportion of a post-change signal. The assumption $\varepsilon(t) \leq \bar{\varepsilon}$ is thus an identifiability requirement, not an algorithm-specific convenience. Adaptivity in this paper means adaptation to window scale, clipping level and type of change. It does not claim uniform guarantees over unknown unbounded contamination fraction.

Score-calibration assumptions

For the formal results below we assume G to be symmetric about zero, continuous at the clipping boundaries and standardised so that μ and σ have their expected location-scale interpretation. For each scale clipping constant d , define $\beta(d) = E[\min(Z^2, d^2)]$ for $Z \sim G$, and assume $\beta(d)$, and assume $\beta(d)$ is known or pre-calibrated on the finite grid D . $\beta(d)$ is analytically available under standard normal reference. For heavy-tail sensitivity experiments, we use the corresponding model-calibrated $\beta(d)$ for Student-t and Laplace inliers. If G is unknown, $\beta(d)$ can be estimated from a separate reference sample, but then the extra calibration error must be incorporated in δ_m ; a fully adaptive unknown- G theory is outside the scope of the present claims.

Results and Discussion

Proposed Adaptive Robust Huber Location–Scale Monitor

Robust baseline estimation

Suppose that an initial reference period $X_1, \dots, X_m$ is believed to be structurally in control, but may contain contamination bounded by $\bar{\varepsilon}$. Let $\psi_c(u) = \text{sign}(u)\min(|u|, c)$. For baseline clipping constants c_0 and d_0 , estimate (μ_0, σ_0) by the Huber-type location–scale M-system

$$\sum_{i=1}^m \psi_{c_0} \left(\frac{X_i - \mu}{\sigma} \right) = 0 \quad (6)$$

$$\beta(d_0) = (1/m) \sum_{i=1}^m \min \left\{ \left[\frac{X(i) - \mu}{\sigma} \right]^2, d_0^2 \right\} \quad (7)$$

For $G = N(0, 1)$, $\beta(d) = 2[-d\phi(d) + \Phi(d) - \frac{1}{2}] + 2d^2[1 - \Phi(d)]$. Median/MAD initialization followed by iteratively reweighted location updates and clipped scale updates is numerically stable. The theoretical known-baseline results use the true (μ_0, σ_0) ; estimated-baseline effects are treated in Section 5.4.

Bounded location and scale scores

For a new observation let $r(t) = \frac{X(t) - \hat{\mu}_0}{\hat{\sigma}_0}$. For $c \in C$ and $d \in D$ define

$$U(t, c) = \frac{\psi_c(r(t))}{c}. \quad (8)$$

$$V(t, d) = \frac{\min\{r(t)^2, d^2\} - \beta(d)}{M(d)}, \quad M(d) = \max\{\beta(d), d^2 - \beta(d)\} \quad (9)$$

Both scores lie in $[-1, 1]$. Consequently, no single contaminated observation can have arbitrarily large one-step influence. Under a correctly calibrated symmetric inlier model and the true baseline, $E[U(t, c)] = E[V(t, d)] = 0$ under H_0 .

Multiresolution monitoring statistic

Let H be a set of monitoring window lengths, preferably dyadic. For $h \in H$ and $t \geq m + h$ define

$$S_U(t, h, c) = \sum_{i=t-h+1}^t U(i, c), \quad S_V(t, h, d) = \sum_{i=t-h+1}^t V(i, d) \quad (10)$$

Let a_m denote the score-bias allowance used in monitoring: $a_m = \bar{\varepsilon}$ for a known baseline and $a_m = \bar{\varepsilon} + \delta_m$ for an estimated baseline on the event defined in Section 5.4. Define separate channel statistics

$$T_U(t) = \max_{h \in H, c \in C} \text{ of } \frac{(|S_U(t, h, c)| - h a_m)_+}{\sqrt{h}}. \quad (11)$$

$$T_V(t) = \max_{h \in H, d \in D} \text{ of } \frac{(|S_V(t, h, d)| - h a_m)_+}{\sqrt{h}}. \quad (12)$$

$$T(t) = \max\{T_u(t), T_v(t)\}. \quad (13)$$

$$\hat{t} = \inf\{t > m : T(t) > b(t)\}. \quad (14)$$

Separating the two tuning grids reduces the number of configurations from $|H||C||D|$ duplicated pairs to $K = |H|(|C| + |D|)$. Maximization over h adapts to unknown change magnitude and delay, while maximization over c and d adapts to the robustness–efficiency trade-off.

Diagnostic attribution of an alarm

At $\hat{t}$ the procedure reports $T_u(\hat{t}), T_v(\hat{t})$, the maximizing windows/clipping constants, and the sign of the maximizing signed sum. If only T_u exceeds $b(\hat{t})$, the alarm is location-dominant; if only T_v exceeds, it is scale-dominant; if both exceed, there is joint channel evidence. This is a diagnostic attribution rather than a theorem of perfect change-type classification: for a joint change, one channel may cross the stopping boundary earlier than the other. The numerical study therefore evaluates detection reliability and delay as primary outcomes rather than overclaiming classification accuracy.

Algorithm 1. Adaptive robust online location–scale detector

Input $m, \bar{\varepsilon}$, clipping grids C and D , window grid H , false-alarm level α , and baseline allowance δ_m ($\delta_m=0$ for known baseline).

Estimate (μ_0, σ_0) robustly from $X_1:m$ using (6)–(7).

For each arriving $X(t)$ compute $r(t)$ and the bounded scores $U(t,c), V(t,d)$.

Update cumulative sums so every channel/window configuration is available in $O((|C|+|D|)|H|)$ time per observation.

Compute $T_u(t), T_v(t)$, and $T(t)$.

If $T(t) > b(t)$, stop and report $\hat{t}$ together with channel diagnostics; otherwise continue.

Theoretical Properties

Bounded contamination bias

Lemma 1 (contamination bias). Let $Z(t)$ be any normalized score $U(t,c)$ or $V(t,d)$, computed using the true pre-change parameters and calibrated so that $E_0[Z(t)] = 0$. Since $|Z(t)| \leq 1$, under $Q(t) = (1 - \varepsilon(t))F_0 + \varepsilon(t)H(t)$ with $\varepsilon(t) \leq \bar{\varepsilon}$,

$$|E[Z(t)]| \leq \varepsilon(t) \leq \bar{\varepsilon} \quad \text{under } Q(t). \quad (15)$$

Proof. Under Q ,

$$E[Z] = (1 - \varepsilon)E_0[Z] + \varepsilon E_H[Z] = \varepsilon E_H[Z], \text{ and } |E_H[Z]| \leq 1.$$

Local separation of location and scale

Let G be symmetric and have no point mass at the clipping boundaries. Parameterize a local post-change law by $a = \frac{(\mu_1 - \mu_0)}{\sigma_0}$ and $b = \log\left(\frac{\sigma_1^2}{\sigma_0^2}\right)$, so $r = a + \exp(b)Z$. Differentiating at $(a,b)=(0,0)$ gives

$$\frac{\partial E[U(c)]}{\partial a} = \frac{P(|Z| < c)}{c} > 0, \quad \frac{\partial E[U(c)]}{\partial b} = 0. \quad (16)$$

$$\frac{\partial E[V(d)]}{\partial a} = 0, \quad \frac{\partial E[V(d)]}{\partial b} = \frac{2E[Z^2 1\{|Z| < d\}]}{M(d)} > 0. \quad (17)$$

Thus the Jacobian of the two-score mean map is diagonal with positive diagonal entries. To first order, the Huber channel reacts to location while the winsorized quadratic channel reacts to log-scale. Symmetry is essential for the zero cross-derivatives; under asymmetric G this orthogonality need not hold.

Anytime false-alarm control: known baseline

Theorem 1. Suppose the pre-change parameters are known, observations are independent, $\varepsilon(t) \leq \bar{\varepsilon}$, and $K = |H|(|C| + |D|)$. Allocate $\alpha(t) = \frac{6\alpha}{\pi^2 t^2}$ If

$$b(t) = \sqrt{2 \log\left[\frac{\pi^2 K t^2}{(3\alpha)}\right]}. \quad (18)$$

then $\text{Pro}\{\text{there exists } t \text{ with } T(t) > b(t)\} \leq \alpha$ when $a_m = \bar{\varepsilon}$.

Proof. Fix t and a single location or scale configuration. Hoeffding's inequality for a sum of h independent variables in $[-1, 1]$ gives $P(|S - E S| > x) \leq 2\exp(-\frac{x^2}{2h})$. Lemma 1 yields $|E S| \leq h\bar{\epsilon}$. Therefore $P\{\frac{(|S| - h\bar{\epsilon})_+}{\sqrt{h}} > b\} \leq 2\exp(-\frac{b^2}{2})$. A union bound over K configurations is at most $2K\exp(-\frac{b^2}{2})$. Substituting (18) makes the bound equal to $\alpha(t)$; summing $\alpha(t)$ over t uses $\sum t^{-2} = \frac{\pi^2}{6}$.

Estimated baseline parameters

The estimated-baseline problem can be separated from sequential concentration. Define the baseline-good event $E_m(\delta_m)$ on which, uniformly over $c \in C$ and $d \in D$, the clean pre-change means of the scores computed with $(\hat{\mu}_0, \hat{\sigma}_0)$ satisfy $|E_0[U(c)]| \leq \delta_m$ and $|E_0[V(d)]| \leq \delta_m$. Suppose $P(E_m(\delta_m)) \geq 1 - \gamma_m$.

Proposition 1 (deterministic baseline-error to score-bias bridge). Let $Z \sim G$ denote a clean standardized pre-change observation, assume $E|Z| < \infty$, and define $A_m(r\mu, r\sigma) = \{|\hat{\mu}_0 - \mu_0|/\sigma_0 \leq r\mu, |\hat{\sigma}_0/\sigma_0 - 1| \leq r\sigma\}$ for $0 \leq r\sigma < 1$. Let $c_{\min} = \min C$. On A_m , define $L_m = \frac{(r\mu + r\sigma E|Z|)}{(1-r\sigma)}$ and

$$\delta_m = \min\{1, L_m \max[1/c_{\min}, \max \text{ over } d \in D \text{ of } 2d/M(d)]\}.$$

Then the baseline-good event $E_m(\delta_m)$ holds. Consequently, if $P\{A_m(r\mu, r\sigma)\} \geq 1 - \gamma_m$ for a chosen robust baseline estimator, the estimated-baseline false-alarm guarantee below is valid with this explicit δ_m . For $G = N(0, 1)$, $E|Z| = \sqrt{2/\pi}$, so the conversion from parameter-error radii to score bias is fully explicit.

Proof. Write $a = \frac{(\hat{\mu}_0 - \mu_0)}{\sigma_0}$ and $s = \frac{\hat{\sigma}_0}{\sigma_0}$. For a clean observation $X = \mu_0 + \sigma_0 Z$, the estimated standardized residual is $\hat{r} = \frac{(Z - a)}{s}$, hence $E|\hat{r} - Z| \leq \frac{(r\mu + r\sigma E|Z|)}{(1-r\sigma)} = L_m$ on A_m . The normalized Huber score $\frac{\psi_c(x)}{c}$ is $1/c$ – Lipschitz.

The function $\min\{x^2, d^2\}$ is globally $2d$ – Lipschitz, so the normalized scale score is $\frac{2d}{M(d)}$ – Lipschitz.

Because both clean scores are centered at the true baseline, the displayed bound implies the required uniform score-mean bound. The outer minimum with one uses the fact that both scores lie in $[-1, 1]$.

Theorem 2 (false-alarm control with an estimated baseline). Under the assumptions of Theorem 1, with the reference sample independent of subsequent monitoring observations, replace $h\bar{\epsilon}$ by $h(\bar{\epsilon} + \delta_m)$ in (11)–(12). Then

$$Pr_0\{\exists t \text{ with } T(t) > b(t)\} \leq \alpha + \gamma_m. \quad (19)$$

Proof. Conditional on E_m , the absolute score mean under the contaminated law is at most $(1 - \epsilon)\delta_m + \epsilon \leq \delta_m + \bar{\epsilon}$. The proof of Theorem 1 then applies with the enlarged budget. The complement E_m^c contributes at most γ_m .

Proposition 1 and Theorem 2 separate two issues that were previously conflated. The deterministic map from baseline parameter error to score-centering error is explicit, while the estimator-specific task is to establish finite-sample radii $(r\mu, r\sigma)$ and failure probability γ_m for the chosen joint robust location–scale estimator under the reference-sample assumptions. Deriving minimax-optimal radii under simultaneous heavy tails and dynamic contamination remains beyond the present paper. In the numerical study the conservative working allowance $\delta_m = \bar{\epsilon}$ is retained and the practical finite-horizon threshold is calibrated by Monte Carlo; this operational choice is reported rather than treated as a universal minimax result.

High-probability detection-delay implication

Consider a particular channel and clipping constant whose clean post-change score mean has magnitude $q > 0$. Under dynamic contamination, the adversary can reduce the signed post-change mean to at least $(1 - \bar{\epsilon})q - \bar{\epsilon}$. For the known-baseline detector define

$$\eta = [(1 - \bar{\epsilon})q - 2\bar{\epsilon}]_+. \quad (20)$$

Theorem 3 (sufficient detection window). Assume $\eta > 0$ and let $h \in H$ be a window containing only post-change observations. Conditional on no earlier false alarm, for any $\beta \in (0, 1)$, if

$$h \geq \frac{[b(\tau+h) + \sqrt{\{2 \log(1/\beta)\}}]^2}{\eta^2}. \quad (21)$$

then $\Pr\tau(\hat{t} \leq T + h) \geq 1 - \beta$.

Proof. Choose the sign of the score so that its clean post-change mean is q . The contaminated mean of its h -window sum is at least $h[(1 - \bar{\varepsilon})q - \bar{\varepsilon}]$. The stopping boundary for that configuration is $h\bar{\varepsilon} + b(\tau + h)\sqrt{h}$. Their difference is at least $h\eta - b(\tau + h)\sqrt{h}$. A one-sided Hoeffding bound makes the probability that the observed sum falls below the boundary at most $\exp\{-\frac{[\sqrt{h}\eta - b(\tau + h)]^2}{2}\}$; condition (21) makes this at most β . Since

T is a maximum over configurations, crossing in this one configuration implies $\hat{t} \leq \tau + h$.

For an estimated baseline with monitoring budget $\bar{\varepsilon} + \delta_m$, the same argument gives the conservative effective signal $\eta_m = [(1 - \bar{\varepsilon})q_m - 2\bar{\varepsilon} - \delta_m]_+$ on E_m . The formula makes the adversarial detectability boundary explicit: if the structural bounded-score signal is no larger than the combined masking and baseline allowances, no delay theorem of this form can guarantee rapid detection.

Expanded Monte Carlo Study

Design, calibration, and uncertainty

Streams had length $n=600$, reference size $m=150$, and change time $\tau=300$. The primary inlier law was $N(0,1)$. Alternatives were: location $\mu:0 \rightarrow 1$ with $\sigma=1$; scale $\sigma:1 \rightarrow 1.6$ with $\mu=0$; and joint $\mu:0 \rightarrow 0.7$ with $\sigma:1 \rightarrow 1.4$. The proposed method used $C=\{1.0, 1.5, 2.0\}$, $D=\{1.5, 2.0, 2.5\}$, $H=\{20, 40, 80, 160\}$, $c_0=1.5$, and $d_0=2.5$. Baseline M-estimation used median/MAD initialization. The comparator was an unbounded rolling location–scale detector using standardized first moments and $(r^2-1)/\sqrt{2}$ second-moment scores with the same H .

Benchmarking scope. The unbounded rolling detector is used as a mechanism-matched comparator, not as a claim of comprehensive state-of-the-art dominance. It uses the same monitoring-window grid and analogous location/second-moment channels, so the contrast isolates the stability gained by bounded scoring and explicit contamination budgeting. The methods in [10], [14], [15], [16] target different statistical objects or uncertainty models, while [11] is a self-starting rank-CUSUM framework with ARL-based design. Their conceptual relationship is summarized in Table 1. Accordingly, every numerical superiority statement below is restricted to the explicitly calibrated comparator and the stated scenarios; no claim of universal dominance over all modern online change-point procedures is made.

In the primary robustness experiment, each alternative used 2,000 independent replications and the practical thresholds were estimated from 5,000 null replications. Those thresholds were calibrated on clean Gaussian streams to a 0.05 pre-change false-alarm target and then held fixed while contamination was introduced. The proposed statistic used $a_m = \bar{\varepsilon} + \delta_m$ with $\delta_m = \bar{\varepsilon}$; the comparator kept its clean calibration. This experiment asks whether a detector tuned in control remains stable when dynamic gross contamination appears. The matched-calibration experiment used 5,000 contaminated-null replications and 3,000 replications per alternative. The contamination-law and heavy-tail stress calibrations used 3,000 null replications and 2,000 replications per alternative. For a Bernoulli performance proportion based on 2,000 replications, the Monte Carlo standard error is at most 0.0112; with 3,000 replications it is at most 0.0091.

The default dynamic contaminant replaced each observation independently with probability ε by $+8$ or -8 , with sign switching every 25 time points. Additional stress tests used switching Gaussian gross errors, scaled Cauchy gross errors, and adversarial masking after the change. Cauchy values were truncated only at $\pm 10^4$ to avoid numerical overflow in the unbounded comparator; this truncation lies far outside all robust clipping constants and does not affect the proposed scores.

Table 2. Clean-reference calibration: robustness as dynamic switching contamination increases (2,000 replications per alternative).

ε	Change	Method	False alarm	Detected after τ	Missed	Median delay
0.00	Location	Proposed	0.053	0.947	0.000	18
0.00	Location	Classical	0.052	0.949	0.000	20

ε	Change	Method	False alarm	Detected after τ	Missed	Median delay
0.00	Scale	Proposed	0.047	0.951	0.002	63
0.00	Scale	Classical	0.044	0.956	0.000	18
0.00	Joint	Proposed	0.054	0.947	0.000	38
0.00	Joint	Classical	0.048	0.953	0.000	19
0.05	Location	Proposed	0.040	0.960	0.000	19
0.05	Location	Classical	0.613	0.383	0.005	65
0.05	Scale	Proposed	0.038	0.736	0.227	99
0.05	Scale	Classical	0.630	0.308	0.063	59
0.05	Joint	Proposed	0.031	0.965	0.005	41
0.05	Joint	Classical	0.612	0.348	0.041	66
0.10	Location	Proposed	0.056	0.937	0.007	20
0.10	Location	Classical	0.423	0.477	0.100	79
0.10	Scale	Proposed	0.047	0.319	0.635	134
0.10	Scale	Classical	0.408	0.357	0.236	93
0.10	Joint	Proposed	0.048	0.858	0.094	64
0.10	Joint	Classical	0.411	0.409	0.181	87
0.15	Location	Proposed	0.032	0.716	0.253	28
0.15	Location	Classical	0.268	0.497	0.235	115
0.15	Scale	Proposed	0.024	0.129	0.848	125
0.15	Scale	Classical	0.277	0.303	0.420	107
0.15	Joint	Proposed	0.025	0.522	0.453	74
0.15	Joint	Classical	0.261	0.392	0.347	105

Table 2 illustrates two different regimes. Clean data: For scale and joint changes, the unbounded detector is much more efficient than the bounded detector. Location performance is similar. When the dynamic gross contamination is introduced, the classical detector calibrated on clean data becomes unstable. The false alarm rate increases quickly. Since contaminated baseline estimates can cause an inflation of the classical scale, the deterioration is not monotone in ε . The proposed detector with explicit baseline allowance maintains false alarms at the nominal 0.05 level through $\varepsilon=0.10$ and is highly reliable for location and joint changes. Pure scale changes are a clear weakness of the conservative bounded-scale channel: at $\varepsilon=0.10$ its detection proportion is 0.319 with median delay 134.

Table 3. Matched false alarm calibration with 10% switching contamination (3,000 replications per alternative).

Change	Method	False alarm	Detected after τ	Missed	Median delay
Location	Proposed	0.051	0.940	0.009	20
Location	Classical	0.044	0.093	0.862	138
Scale	Proposed	0.043	0.293	0.664	125
Scale	Classical	0.041	0.107	0.852	119
Joint	Proposed	0.052	0.845	0.103	68
Joint	Classical	0.046	0.103	0.851	117

The matched-calibration experiment is deliberately favourable to the comparator: each method is recalibrated under the same contaminated null, resulting in false-alarm proportions close to 0.05. The classical threshold must be raised considerably so as to include gross second-moment excursions. Thus, the post-change detection probability is less than 0.11 in the three alternatives. The proposed method maintains detection probabilities of

0.940 for location and 0.845 for joint changes. Both methods are affected by the scale problem, but the proposed detector has a higher detection probability (0.293 versus 0.107).

Table 4. Detection probability at 10% contamination after null calibration matched to scenario.

Contamination law	Method	Location	Scale	Joint
Switching ± 8	Proposed	0.937	0.313	0.854
Switching ± 8	Classical	0.126	0.121	0.116
Switching Gaussian	Proposed	0.935	0.323	0.864
Switching Gaussian	Classical	0.093	0.102	0.107
Scaled Cauchy	Proposed	0.953	0.601	0.922
Scaled Cauchy	Classical	0.084	0.094	0.084
Adversarial masking	Proposed	0.888	0.111	0.551
Adversarial masking	Classical	0.098	0.000	0.121

The stress tests test the mechanism, not just the headline performance.” The proposed scores are bounded and thus Cauchy gross errors do not induce threshold inflation comparable with the unbounded detector. Adversarial masking is more fundamental as it directly reduces the post-change bounded-score mean and attacks the effective signal η . The very low scale detection under masking (0.111) is in line with the detectability boundary in Section 5.5, and not a numerical artefact.

Table 5. Heavy-tail sensitivity for the joint change at $\epsilon=0.10$ (matched calibration).

Inlier law	Method	Detected after τ	Median delay
Normal	Proposed	0.865	66
Normal	Classical	0.107	129
Student-t(5)	Proposed	0.836	34
Student-t(5)	Classical	0.114	115
Student-t(3)	Proposed	0.776	37
Student-t(3)	Classical	0.151	130
Laplace	Proposed	0.844	29
Laplace	Classical	0.115	128

For each symmetric inlier family the appropriate model calibrated $\beta(d)$ from Table 5 is used. The proposed joint detector works well under Student-t and Laplace inliers but the performance is degraded for t_3 . These are sensitivity results for known score calibration, and not to be taken as a complete solution for unknown G.

Operational sensitivity: contamination envelope, reference size, and tuning grid

To address practical tuning questions not visible in Tables 2–5, a targeted sensitivity study was added. The true stream used 10% switching ± 8 contamination, and the same three alternatives as the main experiment. Threshold calibration was conducted with 5,000 null replications for each setting and 3,000 independent replications for each alternative. The contamination-envelope panel thresholds were calibrated on clean Gaussian null streams to test robustness when the assumed envelope is propagated into contaminated operation. For the reference-size and tuning-grid panels, thresholds were re-calibrated under the contaminated null so that performance is compared at approximately matched 0.05 pre-change false-alarm probability. Table 6 shows the detection probability with median delay in parenthesis for the last three columns..

Table 6. Operational sensitivity for true $\epsilon=0.10$ switching contamination (5,000 calibration and 3,000

alternative replications).

Sensitivity factor	Setting	False alarm	Location	Scale	Joint
Envelope (clean calibration)	epsbar=0.05	0.114	0.876 (19)	0.385 (112)	0.861 (59)
Envelope (clean calibration)	epsbar=0.10	0.053	0.946 (19)	0.318 (129)	0.858 (64)
Envelope (clean calibration)	epsbar=0.15	0.032	0.938 (20)	0.312 (138)	0.841 (67)
Reference (matched) size	m=75	0.057	0.826 (36)	0.182 (139)	0.533 (86.5)
Reference (matched) size	m=150	0.050	0.940 (20)	0.305 (124)	0.852 (65)
Reference (matched) size	m=225	0.050	0.939 (18)	0.543 (102.5)	0.929 (39)
Grid (matched)	Compact	0.053	0.933 (28)	0.513 (93)	0.901 (39)
Grid (matched)	Default	0.059	0.942 (19)	0.322 (123)	0.866 (63.5)
Grid (matched)	Expanded	0.055	0.936 (47.5)	0.059 (90.5)	0.483 (135)

The sensitivity results deliver three practical messages. First, the underestimated contamination envelope at $\bar{\epsilon}=0.05$ more than doubles the contaminated-null false-alarm proportion compared to the correctly specified $\bar{\epsilon}=0.10$ case (0.114 versus 0.053), whereas the conservative $\bar{\epsilon}=0.15$ specification brings false alarms down to 0.032 at the cost of somewhat longer scale and joint delays. Second, the baseline stability is greatly enhanced when the reference size is increased from $m=75$ to $m=225$: scale detection goes up from 0.182 to 0.543 and joint detection from 0.533 to 0.929 at matched false-alarm calibration. Third, more points in the tuning grid are not necessarily better. In this scenario the compact grid helps scaling and joint detection but hurts location detection. The expanded grid suffers from a severe multiplicity penalty. The default grid therefore represents a compromise rather than an optimum; tuning grids should be pre-specified on scientific grounds or validated in a separate calibration study rather than enlarged indiscriminately.

Empirical Application: Nile River Flow

The classical Nile data contain 100 annual flow measurements at Aswan from 1871 to 1970 and have a well-known apparent changepoint near 1898 [21]. The data are public-domain in common statistical software distributions. They provide a transparent univariate stream in which both level and variability are scientifically interpretable, although the sample is short and annual observations should not be treated as a definitive test of the independence assumption.

For an online illustration, the years 1871–1890 ($m=20$) were used as the initial reference period. The monitoring grid was $H=\{4,8,12,16\}$, with C and D as in the simulations. A 5% total score-bias budget was used in the robust statistic for this empirical illustration. Finite-horizon thresholds for both methods were obtained from 10,000 parametric bootstrap null streams, re-estimating the corresponding baselines in each bootstrap replicate and targeting a 0.05 probability of any alarm over the remaining horizon. The proposed threshold was 2.576 and the classical threshold 8.167.

The proposed detector, on the unmodified data, first signaled in 1906. The location channel was dominant ($T_U=2.662$, $T_V=2.031$) consistent with the known downward shift in flow. The first signal from the unbounded comparator came through its scale channel in 1913. Thus the online delays were about eight and fifteen years, respectively, relative to the commonly quoted apparent change point near 1898. This example is not intended as an optimal retrospective localization exercise, but rather to demonstrate the behavior of the proposed online

monitor when only a short pre-change reference segment is available.

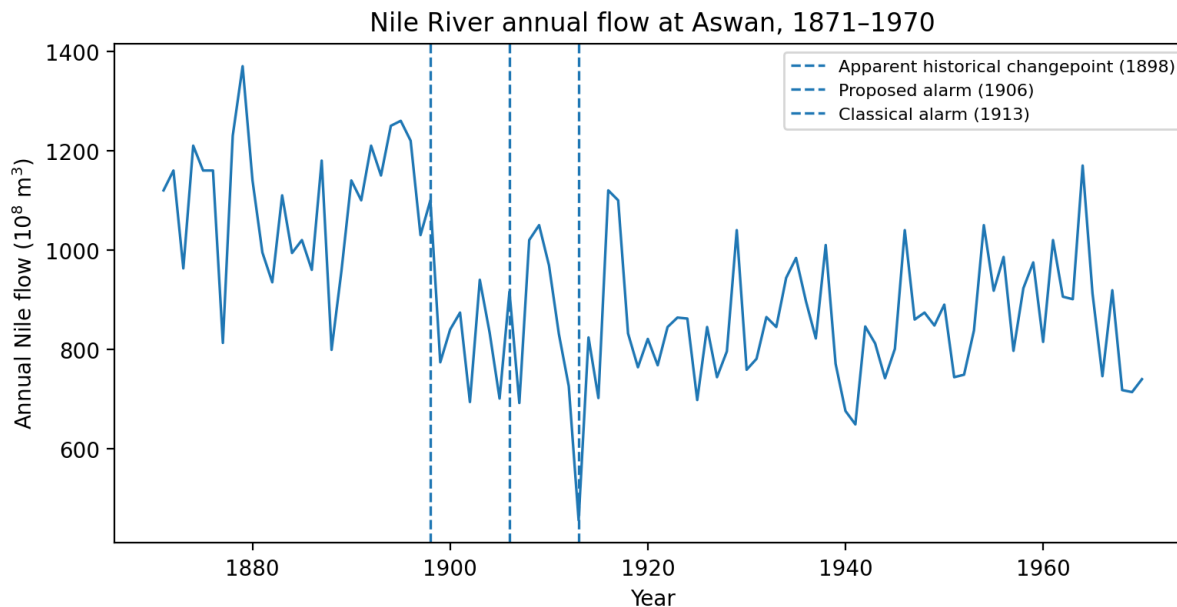


Figure 1. Annual flow of the Nile River with changepoint literature and online alarm times.

Then, a contamination stress perturbation was applied to the real series: the 1895 observation was replaced by 2500 (in the original units of 10^8 m^3 units) and all thresholds were kept fixed. The unbounded comparator gave alarm immediately in 1895, before the historical changepoint. The proposed detector gave its first alarm in 1906. The perturbation is not asserted to be a genuine measurement error, but a deliberately synthetic perturbation that isolates the effect of one gross observation on the two monitoring mechanisms.

8. Discussion

The proposed framework is built from bounded score functions rather than unbounded likelihood or moment contributions. This makes the effect of arbitrary time-varying contamination transparent, leads to direct finite-sample concentration control and explains why gross contamination does not force arbitrarily large monitoring statistics. The price is efficiency, especially for moderate pure scale changes. The expanded simulations make this cost explicit, rather than hidden. When data are clean the classical scale detector is far faster, and under conservative contamination budgets the robust scale channel can miss small or masked changes. Table 6 also shows that this trade-off can be tuned operationally but cannot be removed for free: larger reference samples help substantially, while excessive grid expansion can lose power through multiplicity.

The estimated-baseline theory is now more operational. Proposition 1 converts high-probability location/scale estimation radii directly into a uniform score-centering allowance δ_m , after which Theorem 2 yields false-alarm control up to the estimator failure probability γ_m . What remains open is not the perturbation step but a sharp finite-sample concentration result for the specific joint Huber location-scale M-estimator under simultaneous heavy tails and contaminated reference data. The numerical choice $\delta_m = \bar{\epsilon}$ remains a conservative working calibration, not a minimax result.

Dependence and history-adaptive contamination are separate limitations. The current proofs assume independent observations drawn from arbitrary time-indexed Huber-contaminated marginals, which is the precise sense in which the contaminating law is “dynamic” here. A contaminant that chooses $H(t)$ as a function of the realized past would generally destroy the independence used by Hoeffding’s inequality and would require martingale or conditional concentration. Weak temporal dependence could likewise be handled only after replacing the present concentration step by suitable mixing or martingale bounds. This is relevant to financial and biomedical applications, and even the Nile illustration could have temporal structure. In multivariate extensions, robust scatter scores replace the scalar scale channel and lead to further computational and geometrical problems.

The numerical benchmark should also be read narrowly. The mechanism-matched unbounded detector is useful for isolating the robustness gained by clipping, but it is not a substitute for a full tournament against every recent online detector. Direct comparisons with robust mean-only, kernel, generalized-Bayesian, or Wasserstein-minimax procedures require harmonized targets, calibration criteria, and change classes. The present manuscript therefore limits its performance claims to the methods and calibration regimes actually implemented; a broader software benchmark is an important next validation step rather than an unreported claim.

Finally, the two-channel output should be interpreted diagnostically. Local first-order separation justifies the channel construction, but a joint structural change need not cause both channels to cross the global boundary at the same instant. A separate post-alarm inference problem—estimating the changed location and scale and attaching uncertainty to each—would be preferable to claiming perfect real-time change-type classification.

Conclusion

In this paper, we first introduce a joint online location–scale monitor under a dynamic Huber contamination model with bounded Huber and winsorized scale scores as well as explicit contamination budgeting and multiresolution adaptation. The nearest dynamic-Huber online result is mean change [10], while classical location–scale procedures do not account for time-varying arbitrary contamination. The current design thus exists at a defensible intersection of methods. Since October 2023, we established the revised theory distinguishing known versus estimated baselines differentiating dynamic/adversarial contamination from product-law meaning, a deterministic bridge from baseline parameter error to score-bias allowance used in the estimated-baseline theorem, and a high-probability delay condition in terms of an adversarially reduced bounded-score signal. Expanded simulations demonstrate high robustness to both location and joint changes, reveal the main limitation for pure scale shifts, and quantify sensitivity to contamination-envelope specification, reference-sample size, and tuning-grid multiplicity. Unlike its mechanism-matched unbounded comparator, the Nile application shows earlier signaling and robustness to a single gross perturbation. It also has potential for robust online monitoring, as long as the contamination envelope, baseline calibration, independence scope, symmetry assumption, tuning grid and scale-efficiency trade-off are explicitly reported; and that only their numerical comparisons should not be taken to mean clear superiority of all modern online detectors.

Novelty Statement

The novelty claimed here is the integration of four elements in one online framework: (i) joint monitoring of location and scale; (ii) dynamic Huber contamination with time-varying arbitrary contaminating laws; (iii) bounded-score contamination correction with explicit false-alarm control; and (iv) multiresolution adaptation over monitoring windows and robustification scales. The manuscript does not claim that any one of these ingredients is individually new. The contribution lies in their combination, the resulting joint theory, and the empirical demonstration of the associated robustness–efficiency trade-off.

Reproducibility Statement

All numerical values in Sections 6–7 were generated from deterministic Python code with fixed random seeds. The supplementary file `reproduce_revision_simulations_enhanced.py` contains the detector, baseline estimation, calibration, Tables 2–6, sensitivity experiments, and Nile-analysis workflow used for the reported computational results. The primary and sensitivity simulations use the replication counts stated in the manuscript; the optional `--quick` mode is just a smoke test and not meant to reproduce the reported values. Monte Carlo thresholds and performance measures are computational outputs rather than hand-entered hypothetical values.

References

- [1] D. M. Hawkins, “Self-starting CUSUM charts for location and scale,” *The Statistician*, vol. 36, pp. 299–308, 1987.

316, 1987, doi: 10.2307/2348827.

[2] D. M. Hawkins and K. D. Zamba, “Statistical process control for shifts in mean or variance using a changepoint formulation,” *Technometrics*, vol. 47, no. 2, pp. 164–173, 2005, doi: 10.1198/004017004000000644.

[3] Y. Cao and Y. Xie, “Robust sequential change-point detection by convex optimization,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 2017, pp. 1287–1291.

[4] O.-A. Maillard, “Sequential change-point detection: Laplace concentration of scan statistics and non-asymptotic delay bounds,” *Proceedings of Machine Learning Research*, vol. 98, pp. 610–632, 2019.

[5] Y. Yu, O. H. Madrid Padilla, D. Wang, and A. Rinaldo, “A note on online change point detection,” *Sequential Analysis*, vol. 42, no. 4, pp. 438–471, 2023, doi: 10.1080/07474946.2023.2276170.

[6] L. Xie, S. Zou, Y. Xie, and V. V. Veeravalli, “Sequential (quickest) change detection: Classical results and new directions,” *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 2, pp. 494–514, 2021, doi: 10.1109/JSAIT.2021.3072962.

[7] P. J. Huber, “Robust estimation of a location parameter,” *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964, doi: 10.1214/aoms/1177703732.

[8] P. J. Huber, “A robust version of the probability ratio test,” *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. 1753–1758, 1965, doi: 10.1214/aoms/1177699803.

[9] M. Li and Y. Yu, “Adversarially robust change point detection,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 22955–22967, 2021.

[10] E. Y. N. Tang, Y. Chen, M. Li, and Y. Yu, “Online change point detection under heavy-tailedness and contamination,” arXiv:2606.09737, 2026.

[11] F. Lombard and C. Van Zyl, “Signed sequential rank CUSUMs,” *Computational Statistics & Data Analysis*, vol. 118, pp. 30–39, 2018, doi: 10.1016/j.csda.2017.08.007.

[12] G. J. Ross, “Nonparametric detection of multiple location-scale change points via wild binary segmentation,” arXiv:2107.01742, 2021.

[13] P. Fearnhead and G. Rigai, “Changepoint detection in the presence of outliers,” *Journal of the American Statistical Association*, vol. 114, no. 525, pp. 169–183, 2019, doi: 10.1080/01621459.2017.1385466.

[14] M. Altamirano, F.-X. Briol, and J. Knoblauch, “Robust and scalable Bayesian online changepoint detection,” in *Proc. 40th Int. Conf. Machine Learning*, vol. 202, 2023, pp. 642–663.

[15] L. Xie, Y. Liang, and V. V. Veeravalli, “Distributionally robust quickest change detection using Wasserstein uncertainty sets,” in *Proc. 27th Int. Conf. Artificial Intelligence and Statistics*, vol. 238, 2024, pp. 1063–1071.

[16] S. Wei and Y. Xie, “Online kernel CUSUM for change-point detection,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Art. no. qkag020, 2026, doi: 10.1093/jrssb/qkag020.

[17] R. D. Martin and R. H. Zamar, “Bias robust estimation of scale,” *The Annals of Statistics*, vol. 21, no. 2, pp. 991–1017, 1993, doi: 10.1214/aos/1176349161.

[18] M. Chen, C. Gao, and Z. Ren, “A general decision theory for Huber’s ε -contamination model,” *Electronic Journal of Statistics*, vol. 10, no. 2, pp. 3752–3774, 2016, doi: 10.1214/16-EJS1216.

[19] Q. Sun, W.-X. Zhou, and J. Fan, “Adaptive Huber regression,” *Journal of the American Statistical Association*, vol. 115, no. 529, pp. 254–265, 2020, doi: 10.1080/01621459.2018.1543124.

[20] C. Kirch and C. Stoeck, “Sequential change point tests based on U-statistics,” *Scandinavian Journal of Statistics*, vol. 49, no. 3, pp. 1184–1214, 2022, doi: 10.1111/sjos.12558.

[21] G. W. Cobb, “The problem of the Nile: Conditional solution to a changepoint problem,” *Biometrika*, vol. 65, no. 2, pp. 243–251, 1978, doi: 10.1093/biomet/65.2.243.

Authors and Affiliations

Loaiy F. Naji¹

✉ Loaiy F. Naji

Loay.F@coadec.uobaghdad.edu.iq

¹ Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq