Journal of Neutrosophic and Fuzzy Systems

ISSN: 2771-6449 Vol: 11, No. 02, 2026 PP. 09–16

Article DOI: https://doi.org/10.54216/JNFS.110202

FULL LENGTH ARTICLE

Persistent Choquet Fuzzy Evidence for Detecting and Attributing Distribution Drift in Data Streams

Rama Asad Nadweh[1,*]

[1] Department of Science and Information Technology, Islamic Online University, Doha, Qatar

Email: ramaanadwehh@gmail.com

Received: November 01, 2025 Revised: January 05, 2026 Accepted: March 03, 2026 ⋆ Corresponding author

ABSTRACT

Data-stream drift is rarely one-dimensional: a changed stream may move in location, inflate in scale, alter its tail geometry, or differ globally even when no single moment changes decisively. This paper develops a fuzzy monitoring layer for a scalar stream $z_t \in \mathbb{R}$ by comparing adjacent windows A_t and B_t through four robust evidences $d_{j,t}$: median displacement, robust log-scale change, interquantile tail-shape change, and normalized one-dimensional transport. Stationary calibration maps each $d_{j,t}$ to a fuzzy grade $u_{j,t} \in [0,1]$. A normalized 2-additive capacity then aggregates the evidence by

$$q_t = \sum_{j=1}^{4} m_j u_{j,t} + \sum_{j<k} m_{jk} \min(u_{j,t}, u_{k,t}) \in [0,1],$$

so pairwise reinforcement is modeled explicitly rather than hidden inside an arithmetic score. Persistence is separated from instantaneous evidence through $A_t = [\lambda A_{t-1} + q_t - \delta]_+$, and an alarm occurs when $A_t \geq h$. The resulting Persistent Choquet Fuzzy Drift Monitor (PCFDM) also admits an exact component decomposition $q_t = \sum_j \phi_{j,t}$ for drift attribution. A reproducible Monte Carlo study uses 120 independent stationary calibration streams and 220 test replications for each of seven scenarios. At matched stream-wise calibration, PCFDM detects mean, scale, mixed, and gradual drifts in 94.5%, 89.5%, 95.0%, and 92.3% of runs, with median delays 72, 88, 72, and 192 samples. Its transient-shock alarm rate is 40.5%, compared with 49.1% for fuzzy-mean evidence, 76.8% for maximum fuzzy evidence, and 65.9% for transport alone. Heavy-tail drift remains more difficult (48.2% detection), revealing a genuine trade-off between persistent multi-evidence confirmation and sensitivity to isolated shape changes. The contribution is therefore a mathematically decomposable fuzzy evidence mechanism for monitoring and explaining drift, not a claim of universal dominance over specialized change detectors.

Keywords: Fuzzy measure ▪ Choquet integral ▪ Concept drift ▪ Distribution drift ▪ Data streams ▪ Change detection ▪ Evolving fuzzy systems ▪ Uncertainty aggregation

1. INTRODUCTION

Streaming environments violate the static-distribution assumption whenever the law generating observations changes with time. If P_t denotes the distribution at time t, drift is expressed generically as $P_t \neq P_{t+s}$ for some $s > 0$; in supervised learning this may involve $P_t(X)$, $P_t(Y)$, or the conditional mechanism $P_t(Y \mid X)$ [1–3]. The distinction matters. A detector operating directly on an unlabeled signal z_t identifies *distribution drift* in that monitored quantity; if z_t is a residual, loss, score, or model diagnostic, such drift can become evidence of model degradation without being logically identical to concept drift itself [4, 5]. We preserve that distinction throughout.

Classical sequential detectors are deliberately economical. CUSUM accumulates persistent signed deviation [6], change-point analysis studies structural breaks in ordered observations [7], and later stream-learning methods use error-rate or adaptive-window statistics [8–10]. These methods are powerful when their monitored statistic is well matched to the change. Yet a practical stream may shift through several mechanisms at once. If location evidence is denoted d_1, scale evidence d_2, tail evidence d_3, and distributional transport d_4, then a single scalar d_j inevitably discards the remaining coordinates of the evidence vector $d = (d_1, d_2, d_3, d_4)$.

Fuzzy systems offer a natural language for this multi-evidence setting because a statement such as "the scale has changed" need not be binary. Following the basic fuzzy-set principle $\mu_A(x) \in [0,1]$ [11], we transform each standardized discrepancy into a grade $u_j \in [0,1]$. The central issue then becomes aggregation. A simple mean $\bar{u} = \frac{1}{4}\sum_j u_j$ assumes additivity; a maximum $\max_j u_j$ assumes that one strong component is sufficient. Neither represents the interaction "moderate location evidence becomes more convincing when accompanied by strong transport evidence." Fuzzy measures and Choquet integration provide an established mechanism for such non-additive evidence interaction [12–14].

The proposed Persistent Choquet Fuzzy Drift Monitor (PCFDM) is built around that observation. Two adjacent windows A_t and B_t generate four robust discrepancies $d_{j,t}$. Stationary quantiles calibrate these values into fuzzy grades $u_{j,t}$. A fixed nonnegative 2-additive Möbius representation assigns singleton masses m_j and pairwise masses m_{jk}, producing $q_t \in [0,1]$. A separate persistence state S_t suppresses evidence that does not last, while an exact allocation $\phi_{j,t}$ explains how each component contributed to q_t. In symbols, the information flow is

$$\begin{aligned}(A_t^{\mathrm{win}}, B_t^{\mathrm{win}}) &\longrightarrow d_t \longrightarrow u_t \\ &\longrightarrow q_t \longrightarrow S_t \longrightarrow \{\mathrm{alarm}, \phi_t\}.\end{aligned}$$

Here window notation and accumulator notation are distinguished in the derivation below to avoid ambiguity.

The paper makes four contributions. First, it constructs a robust four-coordinate drift representation whose components have different invariance properties: d_1 is location-sensitive, d_2 is multiplicative-scale sensitive, d_3 isolates quantile geometry, and d_4 compares complete empirical distributions. Second, it introduces a normalized 2-additive Choquet fusion in which interactions are explicit and q_t remains bounded. Third, it separates fuzzy evidence from temporal persistence and proves finite-step accumulation/decay bounds for the leaky state. Fourth, it derives an exact attribution $q_t = \sum_j \phi_{j,t}$ and evaluates both detection and attribution numerically, including negative-control transient shocks and parameter sensitivity. The structure is intentionally different from a predictive evolving fuzzy system: PCFDM monitors a stream and emits interpretable evidence; it does not update a classifier or regression rule.

The remainder is organized as follows. Section 2 positions the method relative to drift detection, evolving fuzzy systems, and fuzzy integrals. Section 3 develops the mathematical evidence representation. Section 4 gives the proposed algorithm and its structural properties. Section 5 specifies the experimental design. Section 6 reports numerical results, Section 7 provides robustness and sensitivity analysis, Section 8 discusses scope and limitations, and Section 9 concludes.

2. RELATED WORK

2.1 Sequential and stream-oriented drift detection

Sequential monitoring predates modern data-stream learning. Page's cumulative-sum construction is designed to accumulate small persistent deviations rather than react only to isolated observations [6], while Hinkley's change-point formulation studies inference on the location of a distributional break [7]. In online machine learning, DDM monitors changes in predictive error [8]; ADWIN maintains an adaptive window and tests whether subwindow means differ sufficiently [9]; and EWMA monitoring provides a computationally light layer for tracking classifier error [10]. Unlabeled-drift detection extends the problem beyond immediate label availability [4].

The broader stream literature emphasizes adaptation as well as detection. SEA combines classifiers trained on sequential chunks [15], while MOA provides a standard environment for online algorithms and evaluation [16]. Surveys distinguish drift types, detector assumptions, adaptation mechanisms, and recurring-stream challenges [1, 3, 17]. A key lesson is that there is no scalar statistic that is uniformly optimal for every drift family. If the affected functional changes from $\mathbb{E}[Z]$ to $\mathrm{Var}(Z)$ or to tail shape, a detector tied tightly to one functional can lose power even when $P_t \neq P_{t+s}$ is substantial.

2.2 Evolving fuzzy systems under nonstationarity

Evolving fuzzy systems address nonstationarity by adapting rule bases, prototypes, antecedent regions, or parameters online. Early work explicitly treated drifts and shifts through evolving fuzzy models [18], and the survey of Baruah and Angelov [19] organized the emerging data-stream perspective. Later reviews cover evolving fuzzy and neuro-fuzzy approaches across clustering, regression, identification, and classification [20]. Self-learning thresholds and model adaptation appear in the evolving fuzzy stream framework of Ge and Zeng [21], while Gu and Shen [22] develop a self-adaptive fuzzy learning system for streaming prediction.

Recent systems continue this adaptive direction. Dynamic similarity weighting has been used to respond to concept drift [23], while X-Fuzz combines evolving neuro-fuzzy learning with local explanations for interpretable adaptation to nonstationary data streams [24]. PCFDM has a narrower objective. It does not attempt to learn $\widehat{y}_t = f_t(x_t)$ or update a fuzzy rule base. Instead, for a monitored scalar z_t, it estimates an evidence vector $u_t \in [0,1]^4$ and an alarm state. This separation allows the same monitor to be attached to a raw sensor, a residual sequence, or a model-loss stream.

2.3 Why a Choquet fuzzy aggregation?

A fuzzy measure $\nu : 2^{\{1,\ldots,p\}} \to [0,1]$ is normalized when $\nu(\varnothing) = 0$ and $\nu(N) = 1$, and monotone when $S \subseteq T$ implies $\nu(S) \leq \nu(T)$. Choquet integration aggregates grades with respect to such a non-additive measure [12]. In multicriteria analysis, this machinery is useful because evidence sources may reinforce or substitute one another rather than contributing independently [13]. The k-additive representation of Grabisch [14] controls complexity by setting higher-order Möbius terms to zero. With $p = 4$ and a 2-additive nonnegative representation, the proposed monitor can express pairwise reinforcement while retaining a transparent formula and exact component allocation.

3. MATHEMATICAL FRAMEWORK

3.1 Adjacent-window representation

Let $z_1, z_2, \ldots$ be a scalar data stream. At evaluation time $t \geq 2w$, define adjacent windows of equal width w,

$$A_t = (z_{t-2w}, \ldots, z_{t-w-1}), \qquad B_t = (z_{t-w}, \ldots, z_{t-1}). \quad (1)$$

The reference A_t and recent window B_t are compared without assuming Gaussianity. For a finite sample W, let

$$\widetilde{\mu}(W) = \mathrm{med}(W), \qquad \widetilde{\sigma}(W) = 1.4826\,\mathrm{MAD}(W), \quad (2)$$

where $\mathrm{MAD}(W) = \mathrm{med}_i\,|W_i - \mathrm{med}(W)|$ and a small positive numerical floor prevents division by zero. Under normal sampling, the factor 1.4826 makes the MAD scale comparable to the standard deviation, while its median construction limits the effect of extreme observations.

The first discrepancy measures standardized robust location change,

$$d_{1,t} = \frac{|\widetilde{\mu}(B_t) - \widetilde{\mu}(A_t)|}{\widetilde{\sigma}(A_t)}. \quad (3)$$

The second measures a multiplicative scale change on a symmetric log scale,

$$d_{2,t} = \left| \log \frac{\widetilde{\sigma}(B_t)}{\widetilde{\sigma}(A_t)} \right|. \quad (4)$$

Thus a scale ratio r and its reciprocal $1/r$ yield equal evidence $|\log r|$.

To separate broad tail geometry from the robust central scale, define

$$\tau(W) = \frac{Q_{0.95}(W) - Q_{0.05}(W)}{Q_{0.75}(W) - Q_{0.25}(W)}, \qquad d_{3,t} = \left| \log \frac{\tau(B_t)}{\tau(A_t)} \right|. \quad (5)$$

Because τ is a ratio of interquantile ranges, it is invariant to affine transformations $W \mapsto a + bW$ with $b > 0$. Hence d_3 responds to a change in relative tail spread even when global units differ.

Finally, let $A_{t,(i)}$ and $B_{t,(i)}$ denote order statistics. Equal sample sizes make the empirical one-dimensional Wasserstein-1 distance especially simple. We normalize it by the reference robust scale,

$$d_{4,t} = \frac{1}{w\,\widetilde{\sigma}(A_t)} \sum_{i=1}^{w} |A_{t,(i)} - B_{t,(i)}|. \quad (6)$$

The vector $d_t = (d_{1,t}, d_{2,t}, d_{3,t}, d_{4,t})$ therefore contains related but non-identical information. A pure location shift tends to raise d_1 and d_4; a scale change raises d_2 and often d_4; and a tail-shape change can raise d_3 without requiring a large median displacement.

3.2 Stationary fuzzy calibration

Raw discrepancies are not directly commensurate. Let $\mathscr{C}_0$ be an independent stationary calibration collection, and let D_j denote all calibration values of component j. Define

$$a_j = Q_{0.75}(D_j), \qquad b_j = Q_{0.98}(D_j) - a_j > 0. \quad (7)$$

The fuzzy drift grade is

$$u_j(d) = 1 - \exp\left[- \left(\frac{[d - a_j]_+}{b_j} \right)^2 \right]. \quad (8)$$

Hence $u_j = 0$ for $d \leq a_j$, while $u_j \uparrow 1$ as $d \to \infty$. For $d > a_j$ and $z = (d - a_j)/b_j$,

$$\frac{\partial u_j}{\partial d} = \frac{2z}{b_j} e^{-z^2} > 0, \quad (9)$$

so stronger standardized discrepancy never lowers its fuzzy evidence. The anchors in (7) are estimated only from stationary calibration streams; no drifted test stream is used to tune a_j or b_j.

3.3 A 2-additive fuzzy capacity

Let $N = \{1, 2, 3, 4\}$. We use singleton Möbius masses

$$(m_1, m_2, m_3, m_4) = (0.10, 0.10, 0.10, 0.20)$$

and nonzero pair masses

$$m_{14} = m_{24} = m_{34} = 0.15, \qquad m_{12} = 0.05,$$

with all other $m_{jk} = 0$. Every mass is nonnegative and

$$\sum_{j=1}^{4} m_j + \sum_{1 \leq j < k \leq 4} m_{jk} = 1. \quad (10)$$

For any $S \subseteq N$, $\nu(S) = \sum_{j \in S} m_j + \sum_{\{j,k\} \subseteq S} m_{jk}$ is therefore a normalized monotone 2-additive capacity. The larger interactions involving component 4 encode the intended interpretation that global transport reinforces a more specific location, scale, or tail discrepancy.

For $u_t = (u_{1,t}, \ldots, u_{4,t})$, the corresponding nonnegative 2-additive Choquet form is

$$q_t = \sum_{j=1}^{4} m_j u_{j,t} + \sum_{j<k} m_{jk} \min(u_{j,t}, u_{k,t}). \quad (11)$$

Unlike $\frac{1}{4} \sum_j u_j$, Equation (11) does not treat the four sources as independent. Unlike $\max_j u_j$, it does not make one isolated grade sufficient to represent the entire evidence state.

Proposition 1 (Bounded monotone fuzzy evidence) *If $u \in [0,1]^4$, all Möbius masses are nonnegative, and (10) holds, then $0 \leq q(u) \leq 1$. Moreover, q is nondecreasing in every coordinate u_j.*

Proof. Every term in (11) is nonnegative. Because $u_j \leq 1$ and $\min(u_j, u_k) \leq 1$, $q(u)$ is no larger than the total mass in (10), namely one. Increasing a coordinate cannot decrease either its singleton term or any pairwise minimum containing it;

with nonnegative coefficients, coordinatewise monotonicity follows. □

4. PROPOSED ALGORITHM

4.1 Persistent Choquet Fuzzy Drift Monitor

Instantaneous evidence q_t should not be confused with an alarm. A short shock can make several $u_{j,t}$ large for one or two evaluations without indicating a persistent regime change. PCFDM therefore uses a leaky accumulator

$$S_t = [\lambda S_{t-1} + q_t - \delta]_+, \qquad 0 < \lambda < 1, \quad \delta \geq 0, \tag{12}$$

initialized at $S_0 = 0$. An alarm is issued at the first time

$$T_h = \inf\{t : S_t \geq h\}. \tag{13}$$

The parameters have distinct roles: q_t represents fuzzy cross-sectional evidence, δ is the evidence offset required for accumulation, λ determines memory, and h is an externally calibrated alarm boundary.

Proposition 2 (Persistence decay and accumulation) *Suppose (12) holds. If $q_{t+r} \leq \delta$ for $r = 1,\ldots,R$, then $S_{t+R} \leq \lambda^R S_t$. Conversely, if $q_{t+r} \geq \delta + \eta$ for $r = 1,\ldots,R$ with $\eta > 0$, then*

$$S_{t+R} \geq \lambda^R S_t + \eta \frac{1-\lambda^R}{1-\lambda}. \tag{14}$$

Starting from $S_t = 0$, any threshold $h < \eta/(1-\lambda)$ is crossed after finitely many persistent evaluations.

Proof. Under $q_{t+r} - \delta \leq 0$, the positive-part operator gives $S_{t+r} \leq \lambda S_{t+r-1}$; iteration yields the decay bound. Under $q_{t+r} - \delta \geq \eta > 0$, truncation is inactive after adding a positive increment, so $S_{t+r} \geq \lambda S_{t+r-1} + \eta$. Iterating the geometric recurrence gives (14). Because its right-hand side tends to $\eta/(1-\lambda)$ as $R \to \infty$, every smaller h is eventually reached. □

The proposition formalizes a property that is only heuristic in many moving-score detectors: evidence below the offset decays at least geometrically, whereas sufficiently persistent evidence accumulates toward a positive limit. For the principal experiment we fix $(\lambda,\delta) = (0.92, 0.18)$ before testing and calibrate only h on stationary streams.

4.2 Exact fuzzy attribution

Because (11) contains interactions, assigning the full pair mass to either member would double-count or privilege one component. We allocate each pair term equally and define

$$\phi_{j,t} = m_j u_{j,t} + \frac{1}{2}\sum_{k\neq j} m_{jk} \min(u_{j,t}, u_{k,t}), \tag{15}$$

where absent pair masses are zero.

Proposition 3 (Conservation of alarm evidence) *For every t, $\phi_{j,t} \geq 0$ and*

$$\sum_{j=1}^{4} \phi_{j,t} = q_t. \tag{16}$$

Proof. Nonnegativity follows from $m_j, m_{jk}, u_j \geq 0$. Summing (15) over j counts each singleton once and each pair twice with factor $1/2$, reproducing (11). □

The attribution is descriptive, not causal. At an alarm, ϕ_1, ϕ_2, and ϕ_3 are treated as location, scale, and tail specialists, while ϕ_4 is global transport support. To reduce confusion between heavy-tail and robust-scale evidence, a prespecified rule labels tail drift when $\phi_3 \geq 0.60\phi_2$ and $\phi_1 < 0.60\max(\phi_2,\phi_3)$; otherwise, comparable location and scale contributions ($\min(\phi_1,\phi_2) \geq 0.65\max(\phi_1,\phi_2)$) imply a mixed label, and the largest remaining specialist determines the class.

4.3 Algorithmic form and complexity

Algorithm 1: Persistent Choquet Fuzzy Drift Monitor (PCFDM)

1. From independent stationary calibration streams, compute a_j, b_j in (7); fix capacity masses m_j, m_{jk}.
2. Calibrate h as the chosen quantile of the *per-stream maximum* persistence state under stationarity.
3. For each evaluation time t, form adjacent windows (A_t, B_t) and compute d_t using (3)–(6).
4. Map $d_t \mapsto u_t$ with (8); aggregate $u_t \mapsto q_t$ with (11).
5. Update $S_t = [\lambda S_{t-1} + q_t - \delta]_+$.
6. If $S_t \geq h$, report an alarm and compute ϕ_t from (15); otherwise continue.

Median, MAD, and fixed quantiles are $O(w)$ with selection algorithms, while the direct empirical transport calculation sorts two windows and is $O(w\log w)$. The implementation therefore costs $O(w\log w)$ per evaluation and stores $O(w)$ observations. Evaluating every s samples gives amortized cost $O(w\log w/s)$ per incoming observation. No $n\times n$ distance matrix or evolving rule base is required.

5. EXPERIMENTAL DESIGN

5.1 Controlled streams and drift scenarios

A controlled design is used because exact onset and drift type are required for both detection delay and attribution. Each stream has length $L = 2400$. The principal monitor uses $w = 96$ and is evaluated every $s = 8$ samples. Except for gradual drift, the change begins at $t_0 = 1400$; a detection is counted when no pre-change alarm has occurred and $T_h \in [t_0, t_0 + 700]$. All pseudo-random generators use fixed seeds, and replicate-level output is retained.

Seven scenarios are generated. The stationary case is $z_t \sim N(0,1)$ throughout. Abrupt mean drift changes $N(0,1)$ to $N(1.2,1)$; scale drift changes it to $N(0,2^2)$; and mixed drift changes it to $N(0.9,1.7^2)$. Gradual location drift starts at $t_0 = 1400$ and linearly moves the mean from 0 to 1.5 over 300 samples before remaining at 1.5. Tail drift changes $N(0,1)$ to a standardized Student distribution $T_{2.5}/\sqrt{2.5/(2.5-2)}$, preserving unit variance while changing shape. Finally, a negative-control shock replaces only 12 observations by $N(0,5^2)$ and then immediately returns to $N(0,1)$. For this last scenario, a lower alarm frequency is preferable because the post-shock regime is unchanged.

Calibration uses 120 independent stationary streams. The fuzzy anchors are estimated from all calibration-window discrepancies, yielding the values in Table 1. For PCFDM and the two fuzzy ablations, h is the empirical 90th percentile of

Table 1. Stationary fuzzy calibration for the four evidence components.

Evidence j	a_j	b_j	Singleton m_j
Location	0.2094	0.2364	0.10
Scale	0.1922	0.2006	0.10
Tail/shape	0.1934	0.2056	0.10
Transport	0.2191	0.1558	0.20

Nonzero interactions: $m_{14} = m_{24} = m_{34} = 0.15$ and $m_{12} = 0.05$; total singleton plus pair mass equals one.

each calibration stream's maximum S_t after warm-up. Transport, Page–Hinkley, and CUSUM thresholds are calibrated analogously from their per-stream stationary maxima. This targets approximately 10% stream-wise stationary alarm probability without using any drifted test data.

5.2 Comparators and numerical criteria

Five comparators isolate different design choices. *FuzzyMean* uses the same u_t and persistence recursion but replaces (11) by $q_t = \frac{1}{4}\sum_j u_{j,t}$. *MaxGrade* uses $q_t = \max_j u_{j,t}$. *Wasserstein* thresholds $d_{4,t}$ directly. A two-sided Page–Hinkley-style cumulative deviation uses offset 0.02, and a two-sided standardized CUSUM uses the first 256 samples for baseline mean/scale with allowance $k = 0.25$. The latter two connect the experiment to classical sequential monitoring [6, 7]; all thresholds are nevertheless independently stationary-calibrated rather than copied from unrelated defaults.

The test experiment contains 220 independent replications for each of seven scenarios, or 1540 streams. For persistent drifts we report valid-detection probability and conditional median delay. Stationarity and the transient shock are evaluated by alarm frequency. Because all methods see the same generated stream within a replication, event-rate contrasts are paired. For methods M_1, M_2, the replicate statistic is $\Delta_r = I_r(M_1) - I_r(M_2)$, where $I_r = 1$ denotes successful drift detection, while for the shock $I_r = 1$ denotes *no alarm*. We report $\bar{\Delta}$ with percentile intervals from 3000 paired bootstrap resamples.

6. NUMERICAL RESULTS

6.1 Stationary behavior and transient shocks

The independently calibrated test false-alarm frequencies are shown in Table 2. Finite calibration and test samples do not make all empirical rates exactly 0.10; they range from 0.0545 to 0.1500. PCFDM obtains 0.1136, close to the intended stream-wise level. More importantly for the negative control, its shock alarm rate is 0.4045. This is 8.64 percentage points below FuzzyMean, 36.36 points below MaxGrade, and 25.45 points below transport-only monitoring. Thus the persistence-plus-interaction construction does not merely maximize sensitivity; it suppresses a substantial fraction of short-lived alarms relative to the more reactive fuzzy/transport alternatives.

Page–Hinkley is the most shock-resistant comparator at 0.1091, and CUSUM has the lowest stationary alarm rate. These results are retained because they expose the design trade-off: PCFDM is not intended to replace a specialized mean-shift control chart when the only objective is rejecting short transients.

Table 2. Alarm rates on stationary streams and a 12-sample transient shock (220 test replications each). Lower is better in both columns.

Method	Stationary	Shock
PCFDM	0.1136	0.4045
FuzzyMean	0.1273	0.4909
MaxGrade	0.1500	0.7682
Wasserstein	0.1273	0.6591
Page–Hinkley	0.0864	0.1091
CUSUM	0.0545	0.4818

6.2 Persistent drift detection and delay

Table 3 reports detection rate and median delay as "rate/delay." PCFDM detects mean, scale, mixed, and gradual drift with rates 0.9455, 0.8955, 0.9500, and 0.9227, respectively. The corresponding median delays are 72, 88, 72, and 192 samples. These four scenarios therefore satisfy

$$\min\{p_{\text{mean}}, p_{\text{scale}}, p_{\text{mixed}}, p_{\text{gradual}}\} = 0.8955,$$

showing a relatively even response across changes that affect different functionals.

For strong location movement, classical cumulative methods are faster: CUSUM reaches median delay 19 under abrupt mean drift, whereas PCFDM requires 72. Under scale drift, however, Page–Hinkley falls to 0.4227 detection while PCFDM remains at 0.8955. Heavy-tail drift reverses the comparison with the fuzzy ablations: MaxGrade detects 0.9136 and FuzzyMean 0.6591, whereas PCFDM detects 0.4818. The reason is structural. A tail-only event can produce one dominant u_j; MaxGrade passes that evidence intact, while the Choquet design deliberately rewards corroboration and the persistence offset δ requires sustained aggregate evidence. This is a limitation for isolated shape change but also contributes to the lower transient-shock response in Table 2.

Across the five persistent-drift scenarios, mean detection rates are 0.8391 (PCFDM), 0.8700 (FuzzyMean), 0.9091 (MaxGrade), 0.8936 (Wasserstein), 0.6764 (Page–Hinkley), and 0.7955 (CUSUM). Reporting this average prevents selective emphasis on PCFDM's strongest cases. Conversely, the shock-rejection probabilities $1 - p_{\text{shock}}$ are 0.5955, 0.5091, 0.2318, 0.3409, 0.8909, and 0.5182, respectively. Sensitivity and transient rejection therefore form a genuine two-objective problem rather than a single ranking.

6.3 Paired uncertainty in event-rate contrasts

Selected paired-bootstrap contrasts in Table 4 quantify those trade-offs. Positive values mean PCFDM has the preferable event outcome: successful detection for a persistent drift or rejection of a transient shock. Under mixed drift, PCFDM exceeds MaxGrade by 0.0455 with a 95% interval $[0.0045, 0.0819]$. Under the shock, its advantage over MaxGrade is 0.3636 and over Wasserstein 0.2545, both well separated from zero. Against FuzzyMean, shock rejection improves by 0.0864 $[0.0455, 0.1273]$.

The negative tail contrasts are equally informative: PCFDM is 0.1773 below FuzzyMean and 0.3318 below transport-only monitoring. The proposed interaction structure is therefore supported as a *persistence/confirmation mechanism*, not as

Table 3. Valid detection rate / conditional median delay in samples. Detection requires no pre-change alarm and an alarm within 700 samples of onset.

Method	Mean	Scale	Mixed	Gradual	Tail/shape
PCFDM	0.945/72	0.895/88	0.950/72	0.923/192	0.482/152
FuzzyMean	0.932/72	0.905/88	0.927/80	0.927/192	0.659/112
MaxGrade	0.909/80	0.914/88	0.905/88	0.905/184	0.914/120
Wasserstein	0.914/48	0.891/64	0.936/48	0.914/160	0.814/312
Page–Hinkley	0.968/56	0.423/402	0.986/69	0.950/164	0.055/330
CUSUM	0.977/19	0.936/117	0.968/26	0.982/125	0.114/362

Table 4. Selected paired event-rate contrasts $\bar{\Delta}$ (PCFDM minus comparator) with 95% paired-bootstrap intervals. For shock, success means no alarm.

Scenario	Comparator	$\bar{\Delta}$	95% interval
Mean	FuzzyMean	0.0136	[0.0000, 0.0318]
Mixed	MaxGrade	0.0455	[0.0090, 0.0864]
Scale	Page–Hinkley	0.4727	[0.3955, 0.5455]
Tail	FuzzyMean	-0.1773	[-0.2364,-0.1273]
Tail	Wasserstein	-0.3318	[-0.4091,-0.2545]
Shock	FuzzyMean	0.0864	[0.0455, 0.1273]
Shock	MaxGrade	0.3636	[0.2955, 0.4318]
Shock	Wasserstein	0.2545	[0.1818, 0.3273]

Table 5. PCFDM attribution accuracy among valid detections.

Scenario	Detected runs	Correct attribution
Abrupt mean	208	0.8462
Gradual location	203	0.9064
Scale	197	0.7411
Mixed	209	0.6842
Tail/shape	106	0.5660

a universal replacement for a detector specialized to isolated tail change.

6.4 Attribution at the alarm time

The exact decomposition in (16) allows the alarm's fuzzy evidence to be inspected rather than treated as a black-box score. Table 5 evaluates the prespecified specialist rule only on PCFDM runs with valid detections. Gradual location drift is treated as the same attribution class as abrupt mean drift, because both change location rather than scale or shape.

The strongest attribution is gradual location at 90.64%; abrupt location reaches 84.62%. Scale and mixed cases are lower because transport contributes to both and location–scale evidence can overlap near an alarm. Tail attribution improves over a naive largest-specialist rule by allowing ϕ_3 to remain diagnostic when robust scale evidence ϕ_2 is also elevated, yet it remains the hardest class. The result supports using ϕ_t as an explanatory decomposition, while discouraging interpretation as a perfectly identified causal drift label.

7. ROBUSTNESS AND ABLATION ANALYSIS

7.1 Interaction ablations

FuzzyMean and MaxGrade are not generic external methods; they are deliberate ablations of the same four fuzzy grades. Thus differences among PCFDM, FuzzyMean, and MaxGrade isolate aggregation. At abrupt mean drift, the rates are 0.9455, 0.9318, and 0.9091; at mixed drift they are 0.9500, 0.9273, and 0.9045. The interaction-based score is therefore favorable in these corroborated scenarios. However, at tail drift the ordering becomes $0.4818 < 0.6591 < 0.9136$. Mathematically, this is consistent with

$$q_{\max} = \max_j u_j \geq u_3,$$

whereas the singleton tail contribution inside PCFDM is only $m_3 u_3 = 0.10 u_3$ unless other evidence activates pairwise terms. The ablation therefore identifies the exact mechanism behind both the benefit and the cost of interaction-based confirmation.

7.2 Window and persistence sensitivity

Table 6 varies w, λ, and δ using independent 60-stream calibrations and 90 test replications per scenario. The column "persistent mean" averages detection over mean, scale, mixed, gradual, and tail drift; "shock" is the undesirable transient alarm rate. These are diagnostic sensitivity runs rather than replacements for the principal experiment.

The window width has the clearest structural effect. Relative to the $w = 96$ sensitivity row, increasing w to 128 reduces the shock alarm rate from 0.411 to 0.122 and raises scale and gradual detection from 0.889 and 0.922 to 0.956 and 0.978, respectively, but tail detection falls from 0.567 to 0.289. The wider comparison window therefore improves sustained low-frequency separation while smoothing a shape change that is expressed mainly through a specialized tail statistic. In the other direction, $w = 64$ raises shock alarms to 0.622 and reduces the five-drift mean to 0.636, with gradual detection only 0.378. Raising λ from 0.92 to 0.97 increases memory and raises shock alarms to 0.500, consistent with the accumulation limit $\eta/(1-\lambda)$ in Proposition 2. Lowering λ to 0.85 gives a lower shock rate 0.322 but slightly reduces persistent mean detection. The base setting is therefore a prespecified compromise rather than an optimized universal constant.

7.3 Calibration interpretation

The stationary grade anchors in Table 1 are similar but not identical: $a_j \in [0.1922, 0.2191]$ and $b_j \in [0.1558, 0.2364]$. This matters because Equation (8) calibrates each evidence on its own null distribution. Without that transformation, a raw sum $\sum_j d_j$ would mix quantities whose stationary scales are arbitrary. The fuzzy map instead compares exceedance relative to a_j and b_j, making the interaction terms $\min(u_j, u_k)$ dimensionally meaningful.

Table 6. PCFDM sensitivity. Every row is recalibrated under stationarity before testing.

w	λ	δ	Mean	Scale	Mixed	Gradual	Tail	Persistent mean	Shock
64	.92	.18	.889	.700	.844	.378	.367	.636	.622
96	.85	.18	.944	.844	.922	.922	.489	.824	.322
96	.92	.12	.944	.878	.911	.878	.478	.818	.378
96	.92	.18	.944	.889	.967	.922	.567	.858	.411
96	.97	.18	.967	.911	.933	.956	.378	.829	.500
128	.92	.18	.989	.956	1.000	.978	.289	.842	.122

8. DISCUSSION

PCFDM separates three questions that are often merged in stream monitoring. The first is *what changed?*, represented by $d_t \in \mathbb{R}^4_+$. The second is *how strongly do the changes jointly support drift?*, represented by the non-additive fuzzy score $q_t \in [0,1]$. The third is *has that evidence persisted long enough to alarm?*, represented by $S_t \geq 0$. This factorization is useful because a change to the capacity ν need not alter the raw evidence definitions, and a change to temporal persistence (λ, δ) need not alter the interpretation of u_t.

The numerical findings also clarify where that decomposition helps. In mean, scale, mixed, and gradual scenarios, PCFDM maintains detection between 0.8955 and 0.9500. It is much less vulnerable than Page–Hinkley to the scale and tail scenarios, while reacting less often than MaxGrade and Wasserstein to the short shock. Pairwise interactions are especially defensible when two components corroborate one another; for example, location and transport evidence jointly increase through $m_{14}\min(u_1, u_4)$. This is different from an additive score in which the derivative of the total with respect to u_1 is fixed regardless of u_4.

The same mechanism explains the principal weakness. If only the tail grade is strongly elevated, interaction reinforcement can be small, and the persistence offset can prevent accumulation. MaxGrade is designed for precisely the opposite philosophy: $q_{\max} = 1$ as soon as any one grade reaches one. The resulting 91.36% tail-detection rate comes with a 76.82% shock-alarm rate in this experiment. No single ordering resolves this trade-off without specifying the relative cost of missed persistent drift and false transient alarms.

The method should also be distinguished from adaptive fuzzy predictors. Evolving fuzzy systems modify a model as data arrive [18, 21, 22], and recent methods explicitly adapt to nonstationary distributions [23, 24]. PCFDM instead returns (T_h, q_t, ϕ_t) for a monitored scalar signal. A practical architecture could place it above an evolving predictor by taking z_t to be a residual or loss, but the present experiment does not test closed-loop adaptation and makes no claim about post-alarm retraining accuracy.

Four limitations are material. First, the capacity masses are prespecified rather than learned. This preserves interpretability and avoids tuning on drifted test cases, but domain-specific data could support constrained estimation of m_j, m_{jk} subject to nonnegativity and normalization. Second, fixed windows create a resolution trade-off, as Table 6 shows. Third, the controlled experiment is univariate; multivariate streams would require componentwise summaries, sliced transport, or a multivariate distribution distance. Fourth, the attribution rule is diagnostic rather than causal, and tail/scale ambiguity remains substantial. These limitations suggest extensions in which ν is learned under monotonicity constraints, w is multiscale, or d_t is built from vector-valued residuals.

9. CONCLUSION

A fuzzy drift monitor has been developed for streams in which evidence of nonstationarity may arise through location, scale, tail shape, or global distributional transport. Four robust discrepancies are independently calibrated to grades $u_t \in [0,1]^4$, after which a normalized 2-additive capacity forms the bounded evidence

$$q_t = \sum_j m_j u_{j,t} + \sum_{j<k} m_{jk} \min(u_{j,t}, u_{k,t}).$$

Temporal confirmation is then introduced separately through $S_t = [\lambda S_{t-1} + q_t - \delta]_+$, for which geometric decay and persistent-accumulation bounds were derived. The pairwise interaction terms are also exactly decomposable through $q_t = \sum_j \phi_{j,t}$, providing an interpretable alarm profile.

In 220-replication test scenarios, PCFDM detects abrupt mean, scale, mixed, and gradual location changes at rates from 89.55% to 95.00%, while reducing transient-shock alarms relative to fuzzy mean, fuzzy maximum, and transport-only evidence. The method is intentionally less aggressive toward isolated tail change; its 48.18% heavy-tail detection rate is below MaxGrade and Wasserstein, making the sensitivity–persistence trade-off explicit rather than hidden. These results support PCFDM as a transparent multi-evidence monitoring layer whose value lies in bounded fuzzy interaction, persistence control, and decomposable attribution. Future work should learn monotone capacities from domain data, extend d_t to multivariate streams, and couple alarms to evolving fuzzy predictors so that detection, explanation, and adaptation can be evaluated jointly.

REFERENCES

[1] J. a. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Computing Surveys*, vol. 46, no. 4, pp. 44:1–44:37, 2014.

[2] G. I. Webb, R. Hyde, H. Cao, H. L. Nguyen, and F. Petitjean, "Characterizing concept drift," *Data Mining and Knowledge Discovery*, vol. 30, no. 4, pp. 964–994, 2016.

[3] J. Lu, A. Liu, F. Dong, F. Gu, J. a. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12, pp. 2346–2363, 2019.

[4] T. S. Sethi and M. Kantardzic, "On the reliable detection of concept drift from streaming unlabeled data," *Expert*

Systems with Applications, vol. 82, pp. 77–99, 2017.

[5] F. Bayram, B. S. Ahmed, and A. Kassler, "From concept drift to model degradation: An overview on performance-aware drift detectors," *Knowledge-Based Systems*, vol. 245, p. 108632, 2022.

[6] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1–2, pp. 100–115, 1954.

[7] D. V. Hinkley, "Inference about the change-point in a sequence of random variables," *Biometrika*, vol. 57, no. 1, pp. 1–17, 1970.

[8] J. a. Gama, P. Medas, G. Castillo, and P. Rodrigues, "Learning with drift detection," in *Advances in Artificial Intelligence – SBIA 2004*, ser. Lecture Notes in Computer Science, vol. 3171, Springer, pp. 286–295, 2004.

[9] A. Bifet and R. Gavaldà, "Learning from time-changing data with adaptive windowing," in *Proceedings of the 2007 SIAM International Conference on Data Mining*, SIAM, pp. 443–448, 2007.

[10] G. J. Ross, N. M. Adams, D. K. Tasoulis, and D. J. Hand, "Exponentially weighted moving average charts for detecting concept drift," *Pattern Recognition Letters*, vol. 33, no. 2, pp. 191–198, 2012.

[11] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[12] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measures and the choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets and Systems*, vol. 29, no. 2, pp. 201–227, 1989.

[13] M. Grabisch, "The application of fuzzy integrals in multicriteria decision making," *European Journal of Operational Research*, vol. 89, no. 3, pp. 445–456, 1996.

[14] M. Grabisch, "k-order additive discrete fuzzy measures and their representation," *Fuzzy Sets and Systems*, vol. 92, no. 2, pp. 167–189, 1997.

[15] W. N. Street and Y. Kim, "A streaming ensemble algorithm (SEA) for large-scale classification," in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 377–382, 2001.

[16] A. Bifet, G. Holmes, R. Kirkby, and B. Pfahringer, "MOA: Massive online analysis," *Journal of Machine Learning Research*, vol. 11, no. 52, pp. 1601–1604, 2010.

[17] H. M. Gomes, J. Read, A. Bifet, J. P. Barddal, and J. a. Gama, "Machine learning for streaming data: State of the art, challenges, and opportunities," *ACM SIGKDD Explorations Newsletter*, vol. 21, no. 2, pp. 6–22, 2019.

[18] E. Lughofer and P. Angelov, "Handling drifts and shifts in on-line data streams with evolving fuzzy systems," *Applied Soft Computing*, vol. 11, no. 2, pp. 2057–2068, 2011.

[19] R. D. Baruah and P. Angelov, "Evolving fuzzy systems for data streams: a survey," *WIREs Data Mining and Knowledge Discovery*, vol. 1, no. 6, pp. 461–476, 2011.

[20] I. Škrjanc, J. A. Iglesias, A. Sanchis, D. Leite, E. Lughofer, and F. Gomide, "Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A survey," *Information Sciences*, vol. 490, pp. 344–368, 2019.

[21] D. Ge and X.-J. Zeng, "Learning data streams online—an evolving fuzzy system approach with self-learning/adaptive thresholds," *Information Sciences*, vol. 507, pp. 172–184, 2020.

[22] X. Gu and Q. Shen, "A self-adaptive fuzzy learning system for streaming data prediction," *Information Sciences*, vol. 579, pp. 623–647, 2021.

[23] H. Li and T. Zhao, "A dynamic similarity weighted evolving fuzzy system for concept drift of data streams," *Information Sciences*, vol. 659, p. 120062, 2024.

[24] M. M. Ferdaus, T. Dam, S. Alam, and D.-T. Pham, "X-Fuzz: An evolving and interpretable neuro-fuzzy learner for data streams," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 8, pp. 4001–4012, 2024.