ICML
International Conference On Machine Learning

DCI
Data-Centric Intelligence Lab

Beyond Euclidean Summaries: Online Change Point Detection for Distribution-Valued Data

- Presented at ICML 2026

Joint Work with Dr. Xiaoyu Chen, Zipan (Yujing) Huang

Yingyan Zeng, zengyy@ucmail.uc.edu

Assistant Professor, Department of Mechanical, Materials, and Industrial Engineering

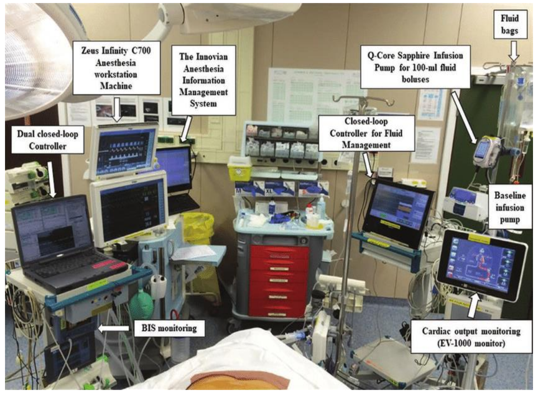
University of Cincinnati

Website: https://yyzeng43.github.io

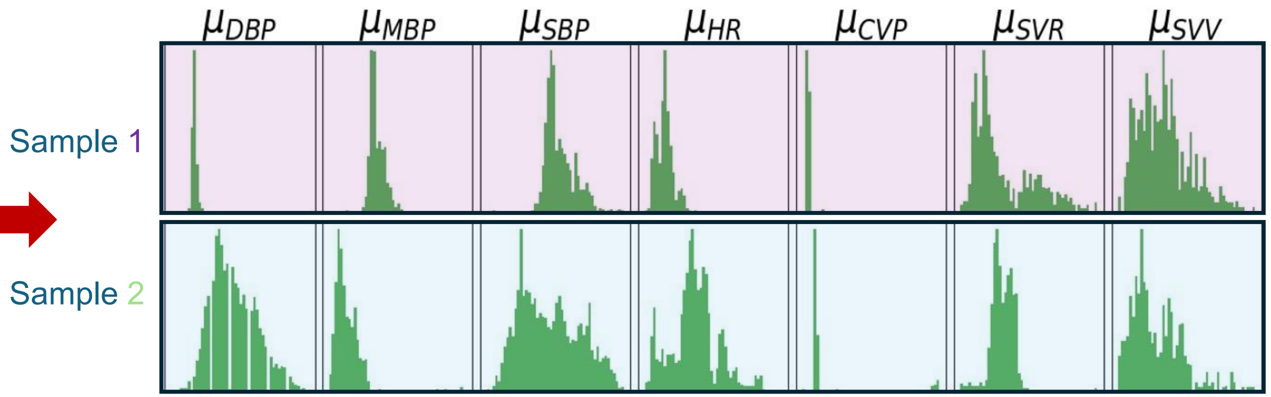
Motivation

From Single Measurements to Distributions

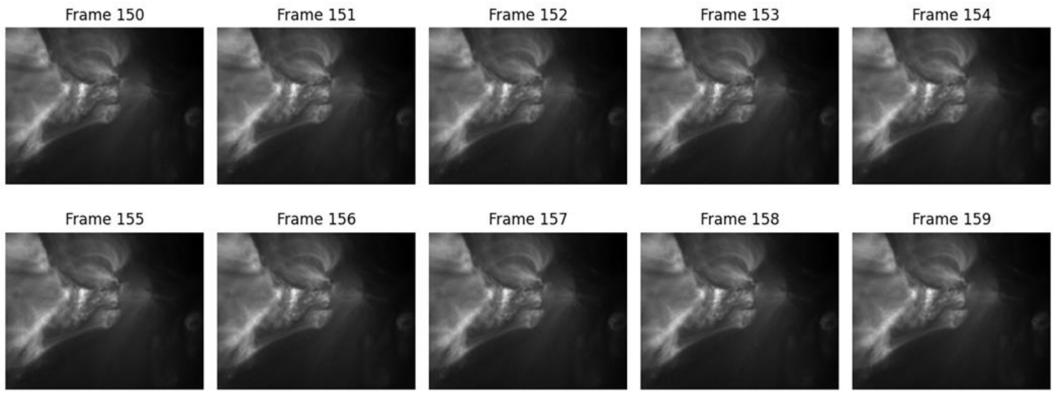
Operation Room Monitoring



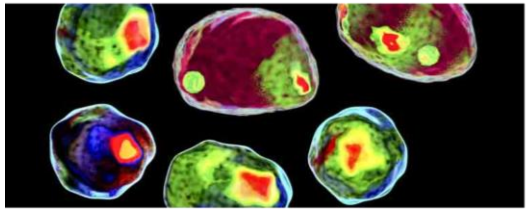
Distribution-value Samples

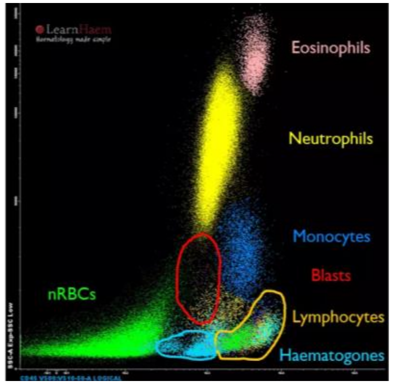


Sunspot Flare



Flow Cytometry





Each step:
an empirical measure
$\mu_t \in \mathcal{P}_2(\mathcal{X})$

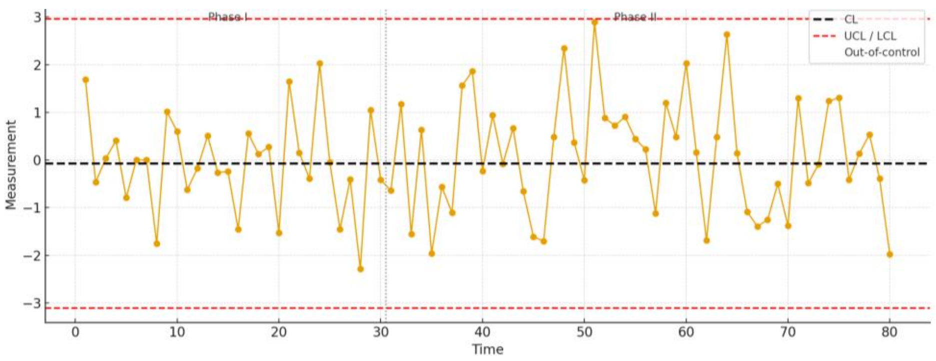
Reference in the Notes

Related Work

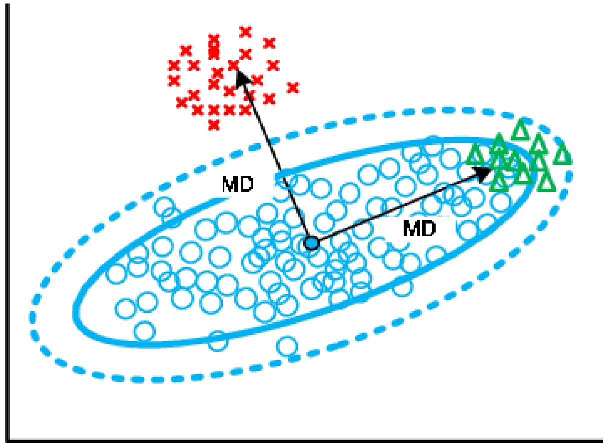
- ***Statistical Process Control***
  - Shewhart Control Chart [Shewhart, W.A. , (1926)]
    - $\bar{X}$ chart
    - X and MR chart (individuals)
    - Only applied to univariate process monitoring
  - **Limitation**: Only focus on the first and second moments (i.e., mean and variance) without the change in distributional shape.
  - Hotelling T^2 Control Chart [Hotelling, H., (1947)]

$$T_t^2 = (\mathbf{X}_t - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\mathbf{X}_t - \boldsymbol{\mu}_0) \qquad T_t^2 \sim \frac{p(n-1)}{n-p} F_{p,n-p}$$

    - Detect the shifts in the mean of multivariate samples.
  - **Limitation**: Gaussian assumption; linear covariance structure; focusing on first and second moments.



UCL/LCL = $\bar{X} \pm 3\,\hat{\sigma}$



Mahalanobis Distance

Related Work

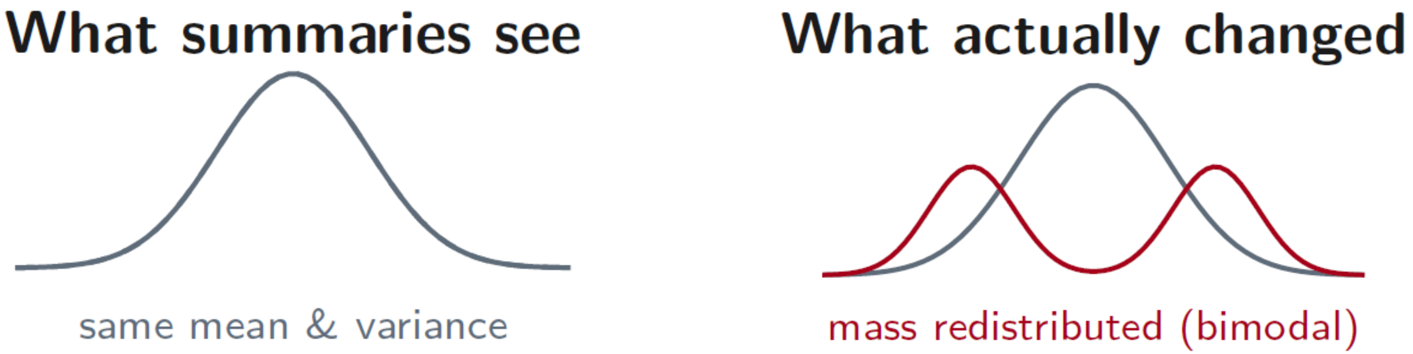
- ***Change Point Detection (CPD) for Distribution Streams***
  - Euclidean (moment based) Method
    - Shewhart, CUSUM, EWMA [Shewhart, W.A. , (1926); Crosier, R. B., (1988); Roberts, S. W. (2000)]
    - **Limitation**: Only focus on low-order moments without the change in distributional shape.
  - Model-free (distance based) Method
    - Kernel two sample tests [Gretton et al., 2012]
    - Scan-B (Maximum Mean Discrepancies) [Li et al., 2019]
    - NEWMA [Keriven et al., 2020]
    - **Limitation**: Reducing a batch to a single scalar discrepancy, only flagging “something changed”.
  - Object-valued CPD
    - Fréchet change point detection [Dubey & Müller(2020)]
    - Self-normalization-based inference [Zhang et al. (2025); Jiang et al. (2024)]
    - L^2-Wasserstein distances [Horvath et al. (2021)]
    - **Limitation**: Built on scalar distances or operate in univariate settings, ignoring the direction of distributional shift

Objective and Challenges

- **Objective**

  To develop an online change point detector that can directly monitor random distribution-valued in the 2-Wasserstein space of probability measures.

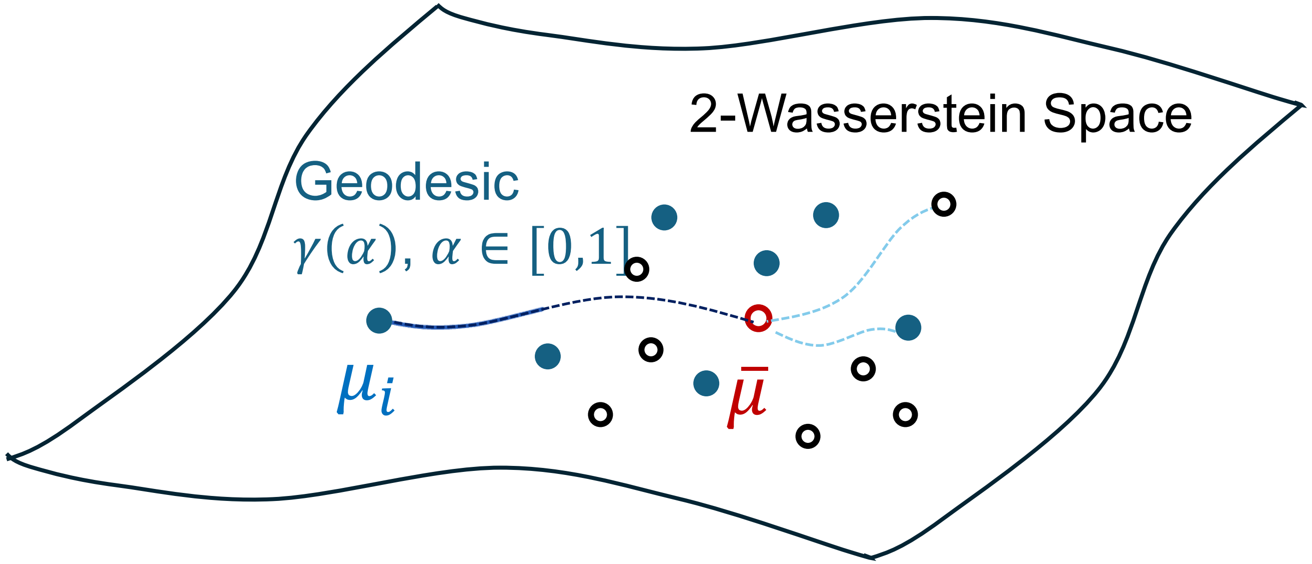
- **Challenges:**
  - Probability measures under the Wasserstein metric form a curved manifold, where the methods based on Euclidean space cannot be applied.
  - Low-order moments or scalar distances fail to capture the change in shape.
  - How to enable distribution-free monitoring?



- **Key Gap:**
  - When mass is redistributed across the support (translations, shape changes, multimodal reweighting), moment-based and model-free detectors have little power as they ignore the **geometry**.

Problem Setup and Preliminaries

Treat the stream $\{\mu_t\}$ as a stochastic process living in the **2-Wasserstein space** $(\mathcal{P}_2(\mathcal{X}))$



Barycenter $\bar{\mu} = \arg\min_{\nu \in \mathcal{W}2(\mathbb{R}^d)} \frac{1}{n} \sum_{i=1}^{n} W_2^2(\nu, \mu_i)$

Optimal Transport Map from $\bar{\mu}$ to μ_i:

$$T_{\#}\bar{\mu} = \mu \Rightarrow \mu_i = T_{\bar{\mu},\mu_i\#}\bar{\mu}$$

2-Wasserstein distance in Kantorovich formulation:

$$W_2^2(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} \|x - y\|^2 \, \pi(dx, dy)$$

Objective: detect the potential change in the distribution-valued sequence $\{\nu_t\}_{t=1}^{n}$ online.

Null Hypothesis: $H_0 : \nu_t \overset{iid}{\sim} \mathbb{P}_1, \quad t = 1, \ldots, n,$

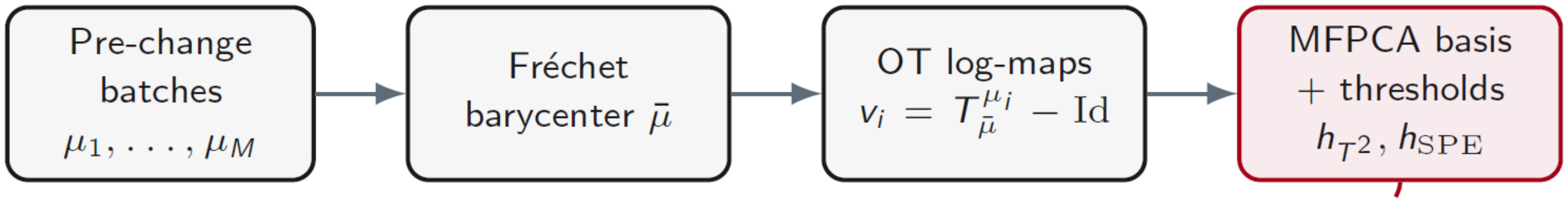
Alternative Hypothesis: H_1: there exists an unknown change point $\kappa \in \{1, \ldots, n\}$

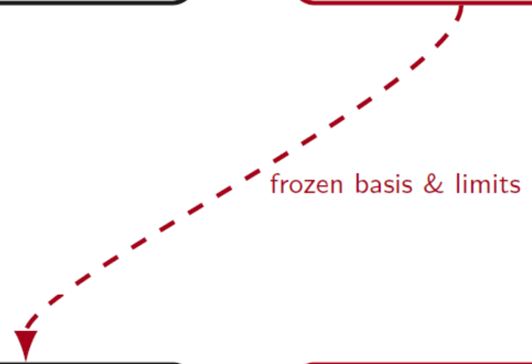
$$t < \kappa,\ \nu_t \sim \mathbb{P}_1, \quad t \geq \kappa,\ \nu_t \sim \mathbb{P}_2, \quad \text{where}\ \mathbb{P}_1 \neq \mathbb{P}_2.$$

IDD — Intrinsic Distribution-valued CPD

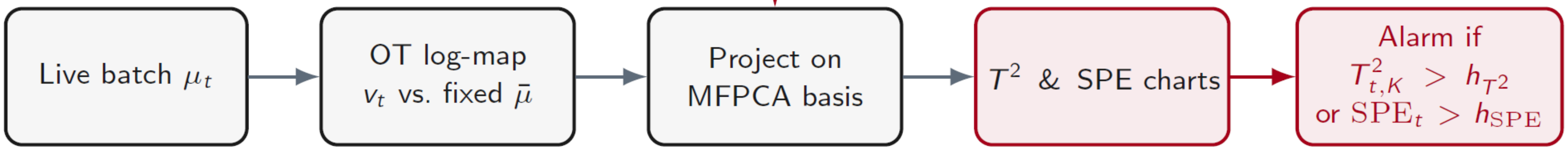
Overview:

- **Pre-Change Calibration**:



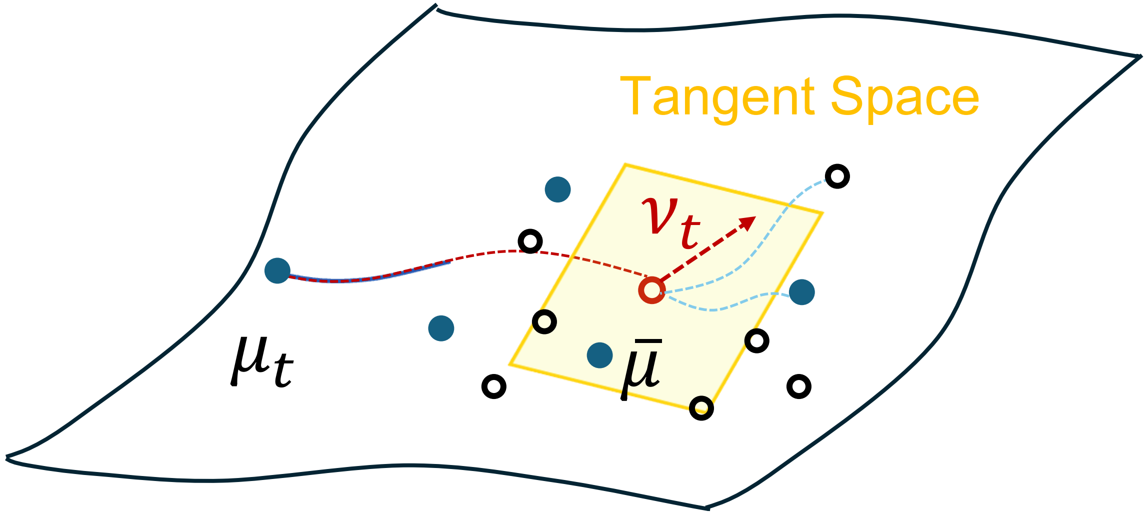


- **Online Monitoring (per incoming batch)**



Linearize $\mathcal{W}2$ via Optimal Transport

- Map the empirical measure μ_t into a common Hilbert space
- *Monitoring Distribution-valued Samples* ⇨ *Monitoring Optimal Transport Maps* ⇨ *Monitoring Tangent Vector Fields*



Denote $T_{\bar{\mu},\mu_t\#}\bar{\mu}$ as $T_{\bar{\mu}}^t$.

OT Displacement Field $\quad v_t := T_{\bar{\mu}}^{\mu_t} - \mathrm{Id} \in L^2(\bar{\mu};\mathbb{R}^d)$

⇩ *Elements on the tangent space located at $\bar{\mu}$*

Linearized embedding of μ_t, exactly preserve the Wasserstein distance from $\bar{\mu}$

Proposition 3.4 (Radial Isometry)

If $\bar{\mu} \in \mathcal{P}_2(\mathcal{X})$ is absolutely continuous, then for any μ the optimal map $T_{\bar{\mu}}^{\mu}$ is $\bar{\mu}$-a.e. unique, and the log-map $v(\mu) = T_{\bar{\mu}}^{\mu} - \mathrm{Id}$ satisfies

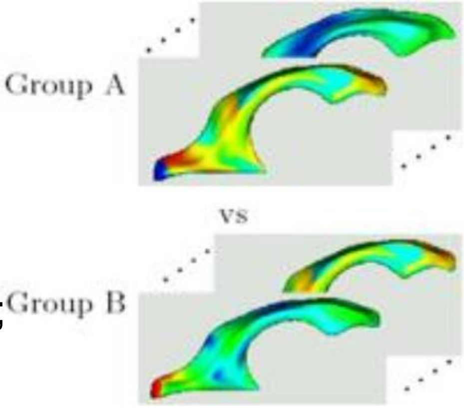
$$\mathcal{W}_2^2(\bar{\mu},\mu) = \int_{\mathcal{X}} \|v(\mu)(x)\|^2 d\bar{\mu}(x) = \|v(\mu)\|_{L^2(\bar{\mu})}^2.$$

Two Complementary Control Charts

- **Key Idea from *Shape Analysis* and *Functional Data Analysis communities*:**
  - Mapping observations to a common tangent space enables vector operations and allows estimation of a covariance in that space.

[Muralidharan & Fletcher, (2012); Dryden & Mardia, (2016)]

[Horv´ath & Kokoszka, (2012); Joseph et al., (2015); Berrendero et al., (2020)]



- **Proposed Approach:**

Apply MFPCA to the pre-change tangent fields. Split variation into a principal subspace and its orthogonal residual.

Hotelling's T^2 (in-subspace)

$$T^2_{t,K} = \sum_{m=1}^{K} \frac{\hat{\xi}^2_{tm}}{\hat{\lambda}_m}$$

Detects shifts *along* the dominant pre-change modes of variation.

SPE / Q (out-of-subspace)

$$\mathrm{SPE}_t = \left\| (I - \hat{P}_K)\Delta_t \right\|^2_H$$

Detects *new* geometry the reference period never showed.

Guarantee I — False-alarm Control & Null Distribution

Theorem 3.10 (Sequential false-alarm control)

Let $(h_{T^2}, h_{\mathrm{SPE}})$ be fixed thresholds. If the monitoring statistics are i.i.d. under $\mathbb{P}_\infty$, the run-length τ is geometric with success probability $p_\infty = \mathbb{P}_\infty(T^2_{1,K} > h_{T^2} \cup \mathrm{SPE}_1 > h_{\mathrm{SPE}})$, so

$$\mathrm{ARL}_0 = n_0 + 1 + \frac{1}{p_\infty} \geq n_0 + 1 + \frac{1}{\alpha_{T^2} + \alpha_{\mathrm{SPE}}}.$$

Proposition 3.9 (Asymptotic null distribution)

As the pre-change size $n_0 \to \infty$, for fixed K,

$$T^2_{t,K} \xrightarrow{d} \chi^2_K, \qquad \mathrm{SPE}_t \xrightarrow{d} \sum_{m>K} \lambda_m Z_m^2, \quad Z_m \sim \mathcal{N}(0,1).$$

Empirical quantiles of the calibration set give data-driven thresholds.

Guarantee II — ϵ-Isometry Guarantee

- The regularity of optimal transport maps forces a rapid spectral decay of the covariance operator, guaranteeing that the tangent space is faithfully captured by a few principal components (ε-isometry).

Theorem 3.14 (ε-Isometry)

Let $\mathcal{X} \subset \mathbb{R}^d$ and the covariance kernel $\mathcal{K}$ satisfy the conditions of Prop. 3.13 (Lipschitz kernel). Let $\{\lambda_m\}_{m\geq 1}$ be the eigenvalues of the covariance operator Γ in non-increasing order. Then there exists a constant $A_{\mathcal{X}}$ depending only on d such that

$$\sum_{m>K} \lambda_m \leq A_{\mathcal{X}}\, C_K\, K^{-1/d}.$$

Consequently, to achieve an ε-isometry in mean square (relative reconstruction error $\leq \varepsilon^2$), it suffices that

$$\boxed{K \geq \left(\frac{A_{\mathcal{X}}\, C_K}{\varepsilon^2\, \mathrm{tr}(\Gamma)}\right)^d.}$$

The number of principal components needed to achieve ϵ −isometry

Experiments

➢ Data Generation

Type	Setting	Change Mechanisms
Continuous (synthetic)	d ∈ {1, 5, 10, 50}, N ∈ {50, 100, 300}	barycenter change, multimodal reweight, copula shift.
Discrete (synthetic)	d ∈ {1, 6}, N ∈ {50, 100, 300}	Poisson spike (rare burst outliers), ordered-categorical drift (mass migrates to adjacent ranks).
Real	$\mathbb{R}^7$ / $\mathbb{R}^{20}$, N ∈ {50-500}	AML FlowCAP-II (cytometry, $\mathbb{R}^7$), Reddit vaccine sentiment (SBERT → $\mathbb{R}^{20}$).

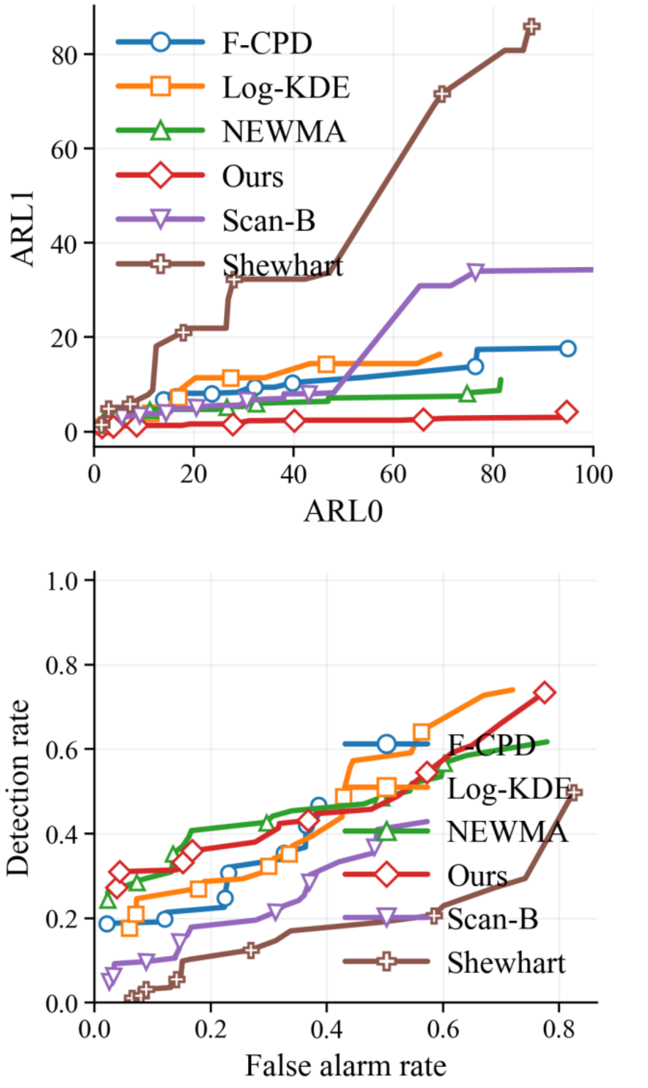
➢ Baselines

- Euclidean: Shewhart / Hotelling T_2
- Distance: Scan-B (MMD), NEWMA (RFF)
- Manifold: Log-KDE (L_2 log-density), F-CPD (Fréchet)

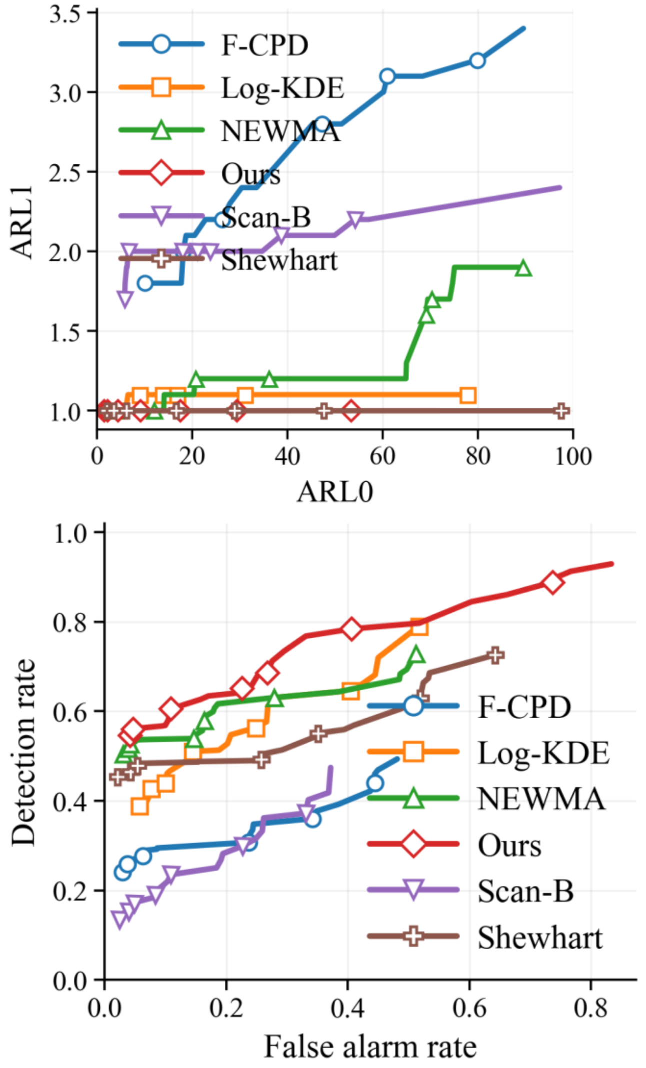
➢ Metrics

- Calibrate all methods to a **matched** ARL_0 (time-to-false-alarm).
- Compare ARL_1 = detection delay (smaller is better); F1 for AML.

Synthetic Continuous Streams



Multimodal reweight (d=10): delay vs ARL0 / detection vs false-alarm

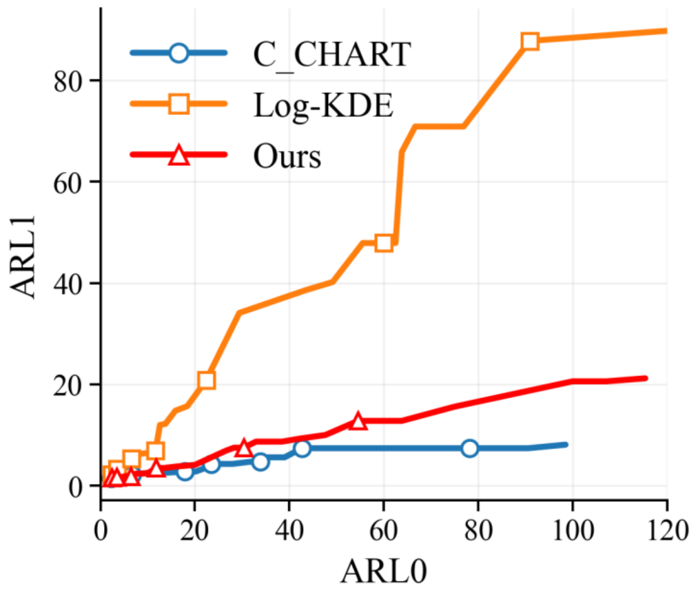


Copula shift (d=50): delay vs ARL0 / detection vs false-alarm

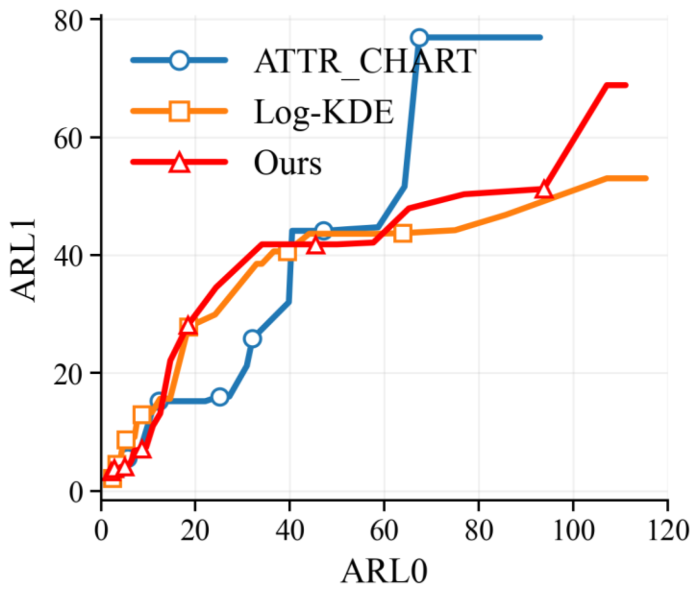
➢ **Finding**

- IDD attains the lowest ARL1 and the best detection/false-alarm trade-off.
- Euclidean Shewhart misses the mean-preserving multimodal shift
- F-CPD weakens in the sparse high-dimensional copula setting.

Discrete Streams



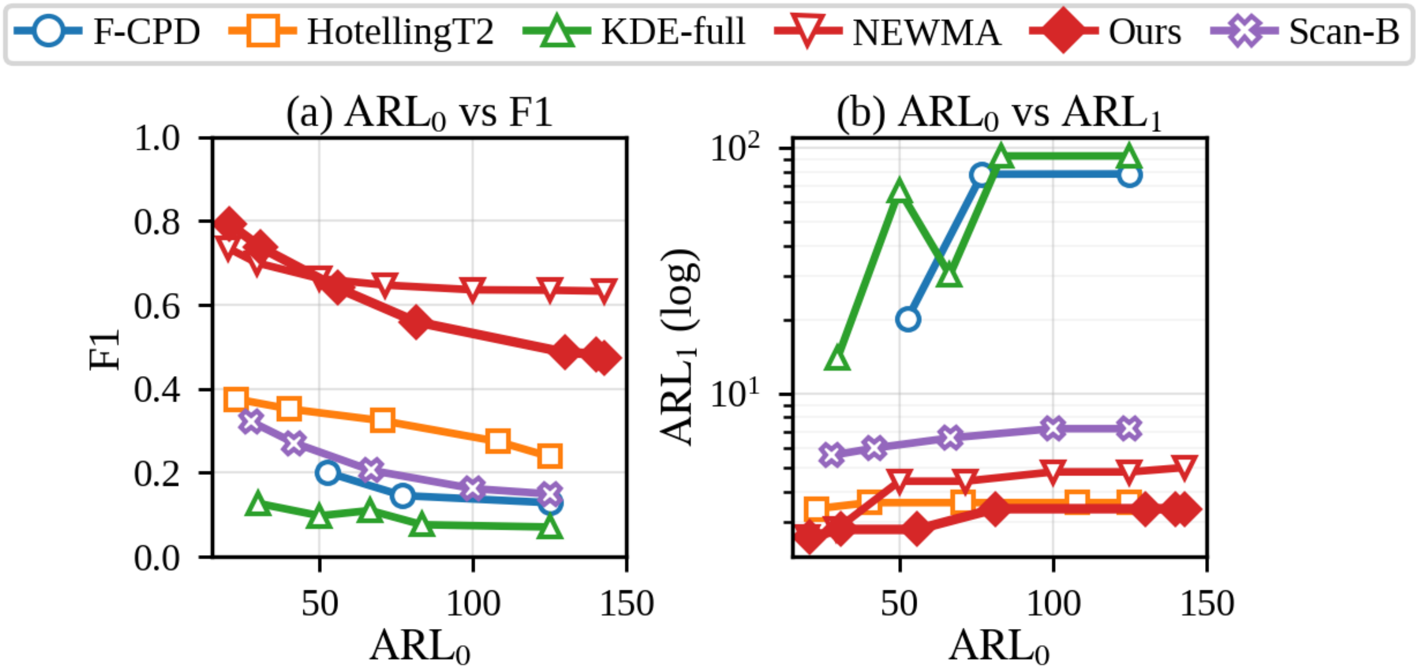
Poisson Spike



Ordered Categorical Drift

- **Finding**
  - IDD significantly outperforms Log-KDE in passion spike scenario. As log-KDE kernel smoothing blurs distinct spikes on discrete grids, IDD preserves signal fidelity by exploiting the full discrete geometry.
  - IDD beats Shewhart attribute charts by exploiting the underlying ordinal metric rather than treating classes as nominal.
  - Leveraging intrinsic discrete/ordinal structure lets IDD detect coherent mass migration between adjacent ranks more efficiently.

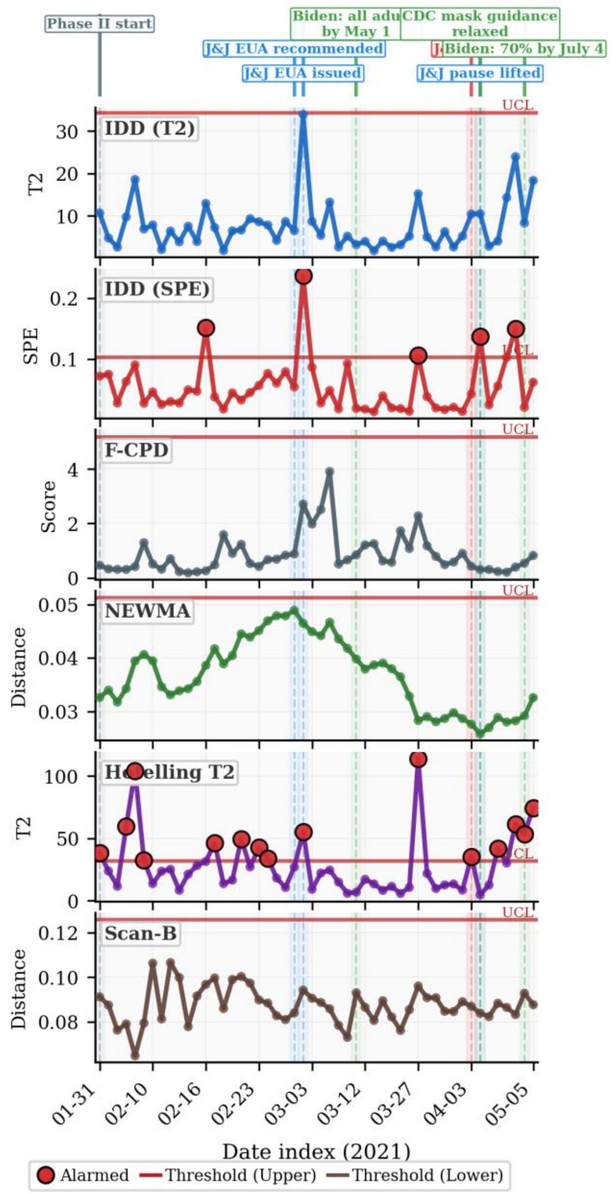
Case Study: AML detection (FlowCAP-II)



- **Setup:**
  - each patient = thousands of single cells in $\mathbb{R}^7$ (fluorescence markers)
  - calibrate on healthy donors, monitor a stream with injected AML-positive subjects.

- **Findings:**
  - IDD: best F1 (≈ 0.75) and near-instant detection (ARL1≈1).
  - Hotelling T2 collapses sub-populations ⇒ precision < 0.4.
  - Captures the intrinsic shape of the leukemic shift.

Case Study: Reddit Vaccine Sentiment



Date	Event	Description
Feb 27	**J&J EUA recommended**	FDA advisory committee recommends Emergency Use Authorization for the Johnson & Johnson vaccine.
Mar 02	**J&J EUA issued**	FDA issues Emergency Use Authorization for the J&J vaccine.
Mar 11	All-adults-by-May-1	President Biden directs states to make all adults eligible for vaccination by May 1.
Apr 13	**J&J Pause**	FDA/CDC recommend pausing J&J administration due to rare clotting reports.
Apr 23	**J&J Pause Lifted**	CDC safety panel recommends resuming J&J vaccinations; pause lifted.
Apr 27	Mask Relaxation	CDC relaxes outdoor mask guidelines for vaccinated people.
May 04	70% by July 4 goal	President Biden announces a 70% adult-vaccination goal by July 4.

- **Setup:**
  - daily comment batches (2021), SBERT→ R20. Phase I (Jan 1–Feb 26, 50 d): calibration. Phase II (Feb 27–May 5, 28 d): monitoring.
- **Findings:**
  - IDD's ***T_2* & SPE alarms (red)** align sharply with the **J&J pause (Apr 13)** and the subsequent reopening.
  - NEWMA drifts monotonically; F-CPD is noisy. Both miss the specific triggers.

Summary and Future Work

- **Summary**:
  - Distribution-valued streams need **geometry-aware** detection — Euclidean summaries hide shape, support, and multimodal changes.
  - A distributional-free online CPD is proposed to monitor distribution-valued samples in the 2-Wasserstein space.
  - T2 statistics and SPE statistics in the tangent space is developed with isometry property.
  - The proposed IDD is validated in the simulation study and case study and shows superior performance in faster, more accurate detection.
- **Future Work:**
  - Expediate the estimation of barycenter and optimal transport plans.
  - Enhance the robustness of the framework by improving the optimal transport plan estimation for multivariate distributions with high dimensionality.
  - Generalized to other applications.