CHANGE-POINT DETECTION USING FOUNDATION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Foundation models have achieved remarkable success across a broad range of tasks, including time-series forecasting and video analysis, yet their potential for change detection remains largely unexplored. We investigate whether pretrained foundation-model representations encode change-relevant information and how such representations can be used for change detection across different contexts without task-specific training. This question is studied in two settings: change-point detection (CPD) in multivariate time series and generic event boundary detection (GEBD) through temporal video segmentation. Our model-agnostic framework extracts frozen foundation-model representations and projects them onto a change-sensitive univariate signal using a data-driven transformation, enabling the direct application of established univariate CPD methods. Across multiple foundation models and benchmarks, the resulting training-free approach is competitive with specialized CPD and GEBD methods, achieving strong performance on CPD benchmarks and matching or exceeding several GEBD baselines at moderate boundary tolerances. It also eliminates the model-training cost required by many self-supervised approaches. Our results suggest that pretrained foundation-model representations can provide a fast and general-purpose basis for change detection across both time-series and video data.

1 INTRODUCTION

Change-point detection (CPD) is an important problem in time-series analysis, with applications spanning sensor networks (Nitzan et al., 2020), financial markets (Cho & Korkas, 2022), and medical monitoring (Fiecas et al., 2024). A closely related problem in video understanding is Generic Event Boundary Detection (GEBD), which aims to identify temporal boundaries between distinct events in a video sequence (Shou et al., 2021). More generally, both problems seek to identify locations in sequential data where the underlying data distribution changes. This becomes particularly challenging for complex, high-dimensional observations such as multivariate time series and video sequences, where changes may be difficult to identify directly in the original input space (Stepikin et al., 2025). These challenges are further amplified for long and high-frequency sequences, which can impose substantial computational and memory costs prohibiting traditional approaches. This is even more apparent in settings where a model is trained on this data.

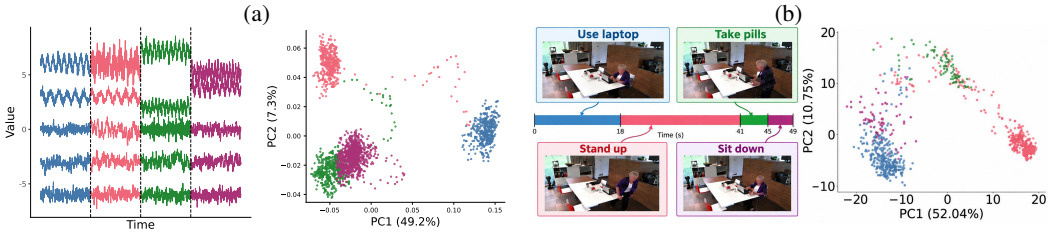


Figure 1: Latent representations from foundation models reveal multiple change points without task-specific training. (a) Multivariate time series with change points and the first two PCA components of latent representations from Chronos-Bolt Tiny. (b) Video from the Toyota Smart home dataset and the first two PCA components of latent representations from DINOv2 ViT-S/14.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

To address these challenges, recent deep learning methods learn latent representations in which temporal changes are easier to identify. Self-supervised and representation-learning methods have achieved strong performance in both CPD and GEBD (Bazarova et al., 2024; Deldari et al., 2021; Kang et al., 2022), but typically require task-specific representation learning under specialized training objectives. Recent advances in foundation models offer a new perspective on this problem (Liang et al., 2024): their pretrained representations can be reused without downstream representation learning or fine-tuning. This raises a fundamental question:

Can pretrained representations be effectively reused for change-point detection?

Figure 1 shows that pretrained foundation-model representations exhibit distinct clusters in the principal component analysis projections (PCA; Jolliffe (2002)), corresponding to different regimes and with transitions near annotated change points and event boundaries. As illustrated in Figure 1a, the latent space separates different distributional regimes, and similar changes are observed in the video example (Figure 1b), where the representation space exhibits changes in the event boundaries. This provides initial evidence that we can use pretrained representations for CPD, even without task-specific training as they seem to retain structure associated with changes in the underlying sequence.

Contributions. While visually distinct clusters provide important insights, they are not sufficient on their own to serve as a change-point detector. We address this gap by proposing a unified framework called GAS-FM, that transforms frozen representations from foundation models (FMs) into a univariate signal via the Generalized Autosimilarity Statistics (GAS), allowing for computationally efficient change-point detection using established statistical methods. In this work we:

- Propose a unified framework applicable across foundation-model families and modalities, and evaluate it on 13 time-series foundation-models and four vision foundation models. We demonstrate competitive unsupervised CPD and GEBD performance across five benchmarks and a real-world case study.
- Introduce *change-encoding maps*, which characterize whether distributional changes in the input series remain distinguishable in representation space, together with an empirical-null maximum mean discrepancy (MMD; Gretton et al. (2012)) framework for quantifying change encoding under temporal dependence.
- Define *change-preserving scalarizations* and establish conditions under which *change-encoding maps* yield a scalar process suitable for statistical CPD under temporal dependence. Motivated by anisotropy in foundation-model representations, we further propose the *Generalized Autosimilarity Statistic* (GAS), a covariance-aware family of efficient scalarizations.

2 Background and Setup

2.1 Related Work

Representation learning for CPD and GEBD. Representation learning has become a prominent approach for detecting changes in complex sequential data. For CPD, methods such as TS-CP^2 (Deldari et al., 2021), SN-TS2Vec (Bazarova et al., 2024), and TIRE (De Ryck et al., 2021) learn representations through self-supervised or unsupervised objectives designed to make temporal changes easier to identify. Related approaches in GEBD, similarly learn representations or boundary-sensitive features from video data (Shou et al., 2021; Kang et al., 2022), including recent methods such as EfficientGEBD and DiffGEBD (Zheng et al., 2024; Hwang et al., 2025). While effective, these approaches require additional task-specific representation learning.

Foundation models. Foundation models provide a natural alternative by offering pretrained representations that can be reused across downstream tasks without fine-tuning. Time-series foundation models have recently emerged for both univariate and multivariate sequences (Ansari et al., 2024; Das et al., 2024; Woo et al., 2024; Ansari et al., 2025), alongside general-purpose vision foundation models with transferable visual representations (Oquab et al., 2024). Recent studies further suggest that these latent spaces contain information relevant to temporal changes (Choi et al., 2026; Uray et al., 2026); for example, Uray et al. (2026) observe separation between operating regimes and pronounced responses near regime boundaries. However, the presence of change-related structure in a

high-dimensional representation does not itself provide a change-point detector: the representation must still be converted into a signal on which boundaries can be identified reliably. GraphGEBD (Meiraz et al., 2025) demonstrates that frozen DINOv2 (Oquab et al., 2024) and ResNet-50 (He et al., 2016) representations can be used for training-free GEBD applying normalized cuts. Although this removes task-specific learning, graph construction is computationally expensive as sequence length increases.

Statistical change-point detection. Statistical CPD provides a mature framework for detecting structural changes in sequential data, ranging from classical CUSUM procedures (Page, 1954) to nonparametric methods capable of detecting general distributional changes (Matteson & James, 2014; Chen & Zhang, 2015; Arlot et al., 2019; McGonigle & Cho, 2025); see Aue & Horváth (2013); Aminikhanghahi & Cook (2017); Truong et al. (2020) for reviews. Direct nonparametric comparison, however, can become computationally demanding for long and high-dimensional sequences. A common strategy is therefore to reduce multivariate observations to a univariate statistic and apply established statistical change-point procedures (Horváth & Hušková, 2012; Cho & Fryzlewicz, 2015; Jirak, 2015; Wang & Samworth, 2018; Enikeeva & Harchaoui, 2019; Grundy et al., 2020). This provides an attractive route for foundation-model representations, provided that projected statistics contain the change-point-specific information.

Despite these developments, a unified framework connecting frozen foundation-model representations to scalable statistical change-point detection across both time series and video sequences remains missing. Our approach addresses this gap by characterizing the information to retain through representation and scalarization, followed by nonparametric change-point detection along with theoretical guarantees.

2.2 Problem Formulation

Let $\{\mathbf{x}_t\}_{t=1}^T$, with $\mathbf{x}_t \in \mathbb{R}^p$, be a multivariate time series and we assume that there exists an unknown number of change-points $k \geq 0$ at $\tau_0 = 1, \ldots, \tau_{k+1} = T$. Consider $\tau_0 = 1$ and $\tau_{k+1} = T$, and define the jth regime as $\mathbb{I}_j = \{t : \tau_{j-1} \leq t < \tau_j\}$, where $j = 1, \ldots, k+1$. Within each regime, the observations are generated by a strictly stationary process with distribution $\mathrm{P}^{(j)}$, and adjacent regimes satisfy

$$\mathrm{P}^{(j-1)} \neq \mathrm{P}^{(j)}, \qquad j = 1, \ldots, k+1.$$

Thus, a change-point may correspond to a change in any distributional properties of the underlying process, including its mean, variance, covariance or higher-order distributional properties. Given a realization $\{\boldsymbol{x}_t\}_{t=1}^T$, the objective is to estimate both the number of change-points k and their locations $\{\tau_1, \ldots, \tau_k\}$. We regard $\{\mathbf{x}_t\}_{t=1}^T$ as an observed realization of the underlying stochastic process $\{\mathbf{x}_t\}_{t\in\mathbb{Z}}$. For the theoretical analysis, we consider α-mixing (Bradley, 2005) with respect to this underlying process, as in Definition 1, permitting strong dependence at short lags, while requiring the dependence to decay exponentially with temporal separation.

Definition 1 (Exponentially α-mixing)**.** *A multivariate time series $\{\mathbf{x}_t\}_{t\in\mathbb{Z}}$ with $\mathbf{x}_t \in \mathbb{R}^p$ is* exponentially α-mixing *if for some constants $c_1, c_2 > 0$,*

$$\alpha_X(\ell) := \sup_{t\in\mathbb{Z}} \sup_{A,B} |\Pr(\mathbf{x}_{\leq t} \in A, \mathbf{x}_{\geq t+\ell} \in B) - \Pr(\mathbf{x}_{\leq t} \in A)\Pr(\mathbf{x}_{\geq t+\ell} \in B)| \leq c_1 e^{-c_2 \ell}.$$

3 Proposed Methodology

3.1 Change-point detection

Figure 2 summarizes the proposed framework. Given a sequential input, a frozen foundation model f maps finite-context windows of size w to high-dimensional latent representations. Because directly applying multivariate change-point detection in this high-dimensional space can be computationally expensive, we optionally denoise the representations and transform them with the Generalized Autosimilarity Statistic (GAS) into a univariate sequence that captures distributional changes in the latent space. We then apply a statistical change-point detector to this scalar sequence, enabling the same framework to support both CPD and GEBD tasks.

Univariate change-point detection: The scalar sequence produced by GAS may contain general distributional changes and inherit temporal dependence from overlapping context windows. We

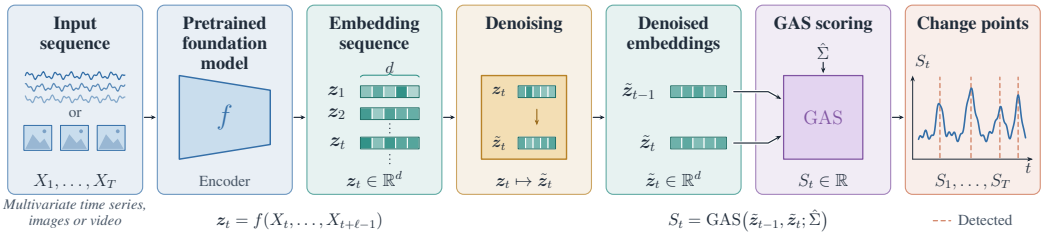


Figure 2: **Overview of the GAS-FM framework.** A multivariate time series, video is mapped to latent representations using a frozen foundation model. The representations are denoised and transformed into a univariate sequence using the Generalized Autosimilarity Statistic (GAS), after which a statistical change-point detector estimates the change-point locations.

therefore use SCAN (Prabashwara et al., 2026), a nonparametric window-based detector using integral probability metrics (IPMs; Müller, 1997) that accommodates exponentially α-mixing observations. SCAN combines detections across multiple window sizes through an ensemble procedure and uses data-driven detection thresholds to identify an unknown number of change points without requiring a prespecified change type. In Section 4.3, we compare SCAN with alternative univariate detectors, including Binary Segmentation (Sen & Srivastava, 1975), PELT (Killick et al., 2012), and KernelCPD (Celisse et al., 2018).

The following sections formalize two requirements: a *change-encoding map* that preserves input changes in the input within the representation space, and a *change-preserving scalarization* that ensures those changes remain detectable in the univariate signal.

3.2 Change-encoding maps

A representation is useful for change-point detection only if it preserves the distributional differences between adjacent regimes. If two consecutive regimes induce the same distribution in the representation space, the change between them cannot be identified from distributional differences in the representations. This observation motivates a minimal requirement on the representation map, which we formalize as the *change-encoding* property, and we further propose an empirical framework to evaluate whether a given encoder and its latent representations satisfy this property.

Definition 2 (Change-encoding map)**.** *Letting $f : (\mathbb{R}^p)^w \to \mathbb{R}^d$ be a measurable foundation-model encoder with context window w, $\mathbf{z}_t = f(\mathbf{x}_{t-w+1:t})$, and $\mathrm{P}_{\mathbf{z}}^{(i)}$ denote the distribution of $\mathbf{z}_t$ for $t \in \mathcal{I}_i$ with $t \geq \tau_{i-1} + w$, then f is a* change-encoding map *if $\mathrm{P}_{\mathbf{z}}^{(i)} \neq \mathrm{P}_{\mathbf{z}}^{(i+1)}$ for all $i = 1, \ldots, k$.*

Definition 2 imposes only an identifiability requirement and places no restrictions on how the representations are obtained or whether they are trained specifically for change-point detection, although we assume that the latent representations are bounded for the theoretical analysis. For subsequent analysis, we must also characterize the temporal dependence of the resulting representation sequence. Since $\mathbf{z}_t$ is constructed from a finite context window, consecutive representations share overlapping observations and are therefore dependent. Theorem 1 shows that finite-context encoding by f preserves exponential α-mixing, which covers a broad class of temporal dependence structures and includes independence as a special case.

Theorem 1. *If $\{\mathbf{x}_t\}_{t\in\mathbb{Z}}$ is exponentially α-mixing, then $\{\mathbf{z}_t\}_{t\in\mathbb{Z}}$ is also exponentially α-mixing.*

This, however, does not ensure that changes between regimes remain distinguishable in the representation space. We therefore empirically assess the change-encoding ability of foundation models using the framework described in Appendix B. At each annotated change point, we compute a change-encoding signal-to-noise ratio (CE-SNR), which measures the distributional discrepancy between the pre- and post-change representations relative to their within-regime variance. We then define the change-encoding rate, CER_α, as the proportion of annotated changes for which the CE-SNR exceeds the $(1-\alpha)$ quantile of an empirical null distribution constructed from no-change pseudo-splits. Empirical evidence that pretrained foundation models exhibit this change-encoding property is presented in Section 4.

3.3 FROM MULTIVARIATE REPRESENTATIONS TO UNIVARIATE DETECTION

Standard dimensionality-reduction methods can map multivariate time series to lower-dimensional representations, but are not typically designed to preserve change-related information. Methods such as PCA prioritize directions of high global variance, which need not coincide with directions that distinguish adjacent regimes, and are restricted to linear projections. Learning a nonlinear projection requires additional task-specific training, contrary to our training-free objective. Instead, we construct a scalar sequence *directly from latent representations produced by the foundation model*.

For a projection to be compatible with the downstream detector, it must satisfy two properties. First, it should preserve the broad class of temporal dependence structures accommodated by SCAN. Second, it should retain sufficient distributional information to distinguish adjacent regimes.

Corollary 1. *Under the conditions of Theorem 1, for any bounded measurable function $h : \mathbb{R}^d)^m \to \mathbb{R}$, the sequence $q_t = h(\mathbf{z}_{t-m+1}, \ldots, \mathbf{z}_t)$ is also exponentially α-mixing.*

Corollary 1 shows that finite-context scalarization preserves the mixing dependence structure inherited by the latent sequence. However, preserving temporal dependence does not guarantee preservation of change-related information. In general, no scalarization can retain every possible distribution in a high-dimensional representation. Therefore, we characterize the class of scalarization functions that preserve the regime differences relevant for downstream change-point detection.

Definition 3 (Change-Preserving Scalarization)**.** *Let $d_{\mathcal{G}}(P, Q) := \sup_{g \in \mathcal{G}} |\mathbb{E}_P[g] - \mathbb{E}_Q[g]|$ denote the integral probability metric (IPM) induced by the function class $\mathcal{G}$ on two distributions P and Q. A bounded measurable function $h : (\mathbb{R}^d)^m \to \mathbb{R}$ is* change-preserving *with respect to $d_{\mathcal{G}}$, for $q_t = h(\mathbf{z}_{t-m+1}, \ldots, \mathbf{z}_t)$, if there exists $\vartheta_{\mathcal{G}} > 0$ such that $d_{\mathcal{G}}(\mathrm{P}_q^{(j)}, \mathrm{P}_q^{(j+1)}) \geq \vartheta_{\mathcal{G}}$, for all $j = 1, \ldots, k$, where $P_q^{(j)}$ is the probability distribution of q_t.*

No scalarization can preserve every possible distributional distinction in the latent space, so the choice of h is problem-dependent. Theorem 2 does not claim validity for any arbitrary function h; it shows that any scalarization satisfying the change-preservation condition yields a SCAN-valid univariate process when a change-encoding map is used.

Theorem 2 (Validity under change-preserving scalarization)**.** *Let $\{\mathbf{x}_t\}_{t=1}^T$ be exponentially α-mixing, $\mathbf{z}_t = f(\mathbf{x}_{t-w+1:t})$ where f is a change-encoding map, and $q_t = h(\mathbf{z}_{t-m+1}, \ldots, \mathbf{z}_t)$ for some change-preserving and bounded $h : (\mathbb{R}^d)^m \to \mathbb{R}$. Suppose $L = w + m - 1$ satisfies $L = o(\tau_{\min})$ where $\tau_{\min} = \min_{0 \leq j \leq k}(\tau_{j+1} - \tau_j)$. Then $\{q_t\}_{t=1}^T$ satisfies the exponential α-mixing, minimum-spacing, and change-preservation conditions required by SCAN.*

Change-encoding map f + Change-preserving h $\implies$ SCAN-valid scalar process

These results characterize what a valid scalarization must preserve but do not specify choice of h. Kernel statistics such as MMD capture general distributional differences but require pairwise kernel evaluations, which can be costly for long sequences. Inner-product similarities are computationally inexpensive and have also been effective in representation-based CPD, where prior work has primarily used cosine similarity (see Deldari et al., 2021; Bazarova et al., 2024). Since each representation already summarizes a finite context window, the similarity between consecutive representations measures local change at a cost linear in T. Inner products, however, depend on the geometry of the latent space, and foundation-model representations are strongly anisotropic (Appendix C). We therefore introduce a covariance-aware family of similarity statistics that includes common choices as special cases.

Definition 4 (**G**eneralized **A**utosimilarity **S**tatistic (GAS))**.** *Let $\mathbf{z}_t \in \mathbb{R}^d$ denote the representation at time t, $\mathbf{\Sigma} \in \mathbb{R}^{d \times d}$ be positive definite, and $\rho : \mathbb{R}^d \to (0, \infty)$ be measurable. The GAS statistic is*

$$S_t = \frac{\mathbf{z}_{t-1}^\top \mathbf{\Sigma}^{-1} \mathbf{z}_t}{\rho(\mathbf{z}_{t-1})\, \rho(\mathbf{z}_t)}.$$

Within the general scalarization framework above, GAS corresponds to setting $m = 2$ and $q_t = S_t = h_{\mathrm{GAS}}(\mathbf{z}_{t-1}, \mathbf{z}_t)$. The matrix $\mathbf{\Sigma}$ determines the feature-space geometry, while ρ controls the degree of normalization. Different choices of ρ recover several commonly used similarity

measures. In particular, normalization by a constant yields scaled dot-product similarity, whereas $\rho(\mathbf{z}) = \|\mathbf{z}\|_2$ gives cosine similarity. More generally, regularized and norm-constrained variants of cosine similarity are obtained with $\rho(\mathbf{z}) = (\|\mathbf{z}\|_2^\beta)$ where $0 < \beta < 1$, and $\rho(\mathbf{z}) = \max\{c, \|\mathbf{z}\|_2\}$, respectively. Choosing $\rho(\mathbf{z}) = \sqrt{\mathbf{z}^\top \boldsymbol{\Sigma}^{-1}\mathbf{z}}$ instead yields cosine similarity under a Mahalanobis feature-space geometry.

Thus, GAS is not a single statistic, but a unified family of computationally inexpensive similarity projections that can adapt to the anisotropic structure of foundation-model representations. Proposition 1 (Appendix A.4) shows that covariance-adjusted SDP-GAS preserves changes in the latent mean, covariance, or lag-one dependence whenever such changes induce a difference in its first or second moment. Analogous arguments extend to other statistics within the GAS family. Together with Theorem 2, this provides sufficient conditions under which the resulting scalar process remains suitable for change-point detection with SCAN.

Choice of metric. Covariance-adjusted GAS is useful for anisotropic representations, but its sensitivity depends on how the change aligns with the eigendirections of the covariance matrix, so no single geometry uniformly dominates. Among the available choices, we adopt the scaled dot product (SDP) as the default, because it retains both radial and directional information. This follows from the radial-directional decomposition (Jennrich & Port, 1988) defined as $\mathbf{y}_t = \boldsymbol{\Sigma}^{-1/2}\mathbf{z}_t = r_t \boldsymbol{u}_t$, $r_t = \sqrt{\mathbf{z}_t^\top \boldsymbol{\Sigma}^{-1}\mathbf{z}_t}$, with $\|\boldsymbol{u}_t\|_2 = 1$. Then the covariance-adjusted cosine and SDP are

$$S_{cos,\boldsymbol{\Sigma}}(\mathbf{z}_i, \mathbf{z}_j) = \boldsymbol{u}_i^\top \boldsymbol{u}_j, \qquad S_{SDP,\boldsymbol{\Sigma}}(\mathbf{z}_i, \mathbf{z}_j) = \frac{r_i r_j}{\sqrt{d}} \boldsymbol{u}_i^\top \boldsymbol{u}_j.$$

Both account for anisotropy through the covariance-adjusted geometry, but cosine normalization discards radial variation. We therefore use SDP as the default GAS statistic. In high dimensions, direct estimation of $\boldsymbol{\Sigma}^{-1}$ can be unstable. Therefore, in such cases we use the Ledoit-Wolf shrinkage estimator (Ledoit & Wolf, 2004) to obtain a well-conditioned covariance estimate.

Optional denoising. Latent representations can contain substantial local noise, increasing the variability of the GAS sequence. We model the representation as $\mathbf{z}_t = \mathbf{s}_t + \boldsymbol{\varepsilon}_t$ and optionally apply a temporal denoiser before computing GAS, yielding a smoother projected signal and reducing the likelihood that changes are masked by noise (see equation 4, Appendix A.4). This approach accommodates standard denoising approaches, including total-variation denoising (Rudin et al., 1992), median filtering (Fried & George, 2025) and wavelet shrinkage (Donoho & Johnstone, 1994). We evaluate the effect of denoising separately through ablation experiments.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate on three time-series and two video benchmarks. Yahoo S5-A4 (Yahoo! Inc., 2015) contains 100 synthetic time series with changes in trend, seasonality, and noise. USC-HAD (Zhang & Sawchuk, 2012) contains 12 daily activities from 14 subjects recorded at 100 Hz, with activity transitions used as change-points, while HASC (Kawaguchi et al., 2011) consists of wearable accelerometer recordings with activity transitions treated as change-points. For video, we use the validation sets of TAPOS and Kinetics-GEBD (Shou et al., 2021).

Foundation models and evaluation. We evaluate 13 foundational models for time-series and four vision foundation models of varying sizes (Table 3, Appendix D.1), all kept frozen. The models span encoder-only, encoder-decoder, and decoder-only architectures, with representation extraction detailed in Appendix D.2. Time-series CPD is evaluated using F1-score with one-to-one matching at dataset-specific tolerances adopted from Deldari et al. (2021), while GEBD uses micro-F1 at standard relative-distance thresholds. For CPD, each benchmarking dataset is split into a validation (30%) and test set (70%). The context-window length, scalarization statistic, and denoising coefficient are selected on the validation set using grids in Appendix D.3, and the selected configuration is evaluated once on the test set. For GEBD benchmarks, we follow the same protocol, reserving 5% of the training data for hyperparameter tuning and evaluating the selected configuration on the validation set. All experiments use a single NVIDIA L40S GPU (8c vGPU, 12 GB VRAM).

4.2 RESULTS

Change-encoding rate. Frozen foundation-model representations preserve most annotated changes: at calibration level $\alpha = 0.10$, CER ranges from 0.74-0.88 on Yahoo A4 and 0.67-0.83 on USC-HAD, with false change-encoding rates near the nominal 0.10 level (Figure 5). CER is similarly high on Kinetics-GEBD, but lower on HASC and TAPOS (Figure 6). Across these comparisons, foundation models achieve higher CER than TS-CP^2 (Figure 7).

Time series change-point detection. Table 1 compares our foundation-model-based framework with unsupervised and self-supervised CPD methods with hyperparameters selected on the validation set, with the held-out test set used only for final evaluation. Results for all 13 TSFMs are reported in Appendix D.4, Table 5. Overall, the proposed unsupervised approach via foundation models remains competitive with existing self-supervised and unsupervised methods.

Table 1: CPD benchmark results on Yahoo A4, USC-HAD, and HASC (F_1 across detection margins and their averages). Best and second-best per column are **bold** and underlined, respectively.

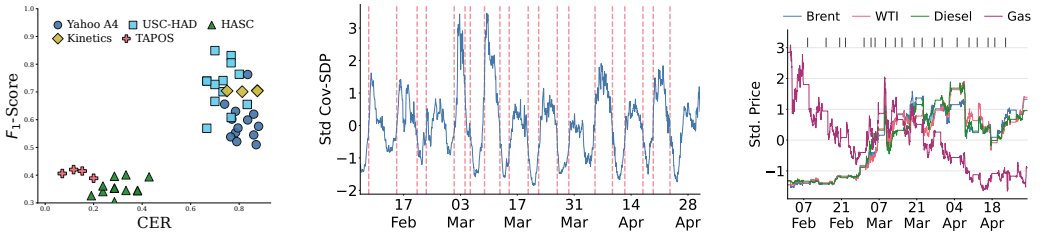
	Yahoo A4				USC-HAD				HASC			
Method	**24**	**50**	**75**	**Avg.**	**100**	**200**	**400**	**Avg.**	**60**	**100**	**200**	**Avg.**
Self-supervised												
TS-CP^2 (Deldari et al., 2021)	.640	.810	.843	.764	.824	.857	.833	.838	.400	.438	.632	.490
SN-TS2Vec (Cos.) (Bazarova et al., 2024)	.726	.775	.852	.784	.736	.909	**1.000**	.882	.476	.306	.663	.482
SN-TS2Vec (MMD) (Bazarova et al., 2024)	.692	.810	.851	.784	**.909**	.809	**1.000**	.906	.476	.467	.444	.462
Unsupervised												
RuLSIF (Liu et al., 2013)	.610	.612	.614	.612	.398	.398	.415	.404	**.482**	**.518**	.590	**.530**
ESPRESSO (Deldari et al., 2020)	.224	.340	.444	.336	.633	.833	.833	.766	.288	.423	**.693**	.468
MC-TIRE (Cao et al., 2024)	.197	.213	.229	.213	.783	.783	.783	.783	.383	.426	.497	.435
GAS-Chronos-2	**.732**	**.836**	**.861**	**.810**	.741	.932	.955	.876	.394	.446	.529	.456
GAS-Moirai-Base	.580	.656	.710	.649	.849	**.950**	.950	**.916**	.396	.508	.603	.502

Chronos-2 and Moirai-Base, both designed for multivariate time series, perform particularly well on Yahoo A4 and USC-HAD, respectively, possibly because their architectures can model cross-variable dependencies in multivariate data (especially on USC-HAD). However, most TSFMs remain competitive, suggesting that multivariate capability alone does not determine performance and that architecture and pretraining also matter. Notably, HASC appears to be a particularly challenging benchmark as the self-supervised baselines also fail to achieve uniformly high detection performance, with comparatively lower F_1-scores.

Comparison across time-series foundation models. Considering the validation and test results across TSFMs, we find no consistent benefit from increasing model size. Larger variants outperform their smaller counterparts in only two of the five model families, and parameter count is not meaningfully associated with performance across the 13 TSFMs. We also observe an architectural pattern: decoder-only models achieve lower F_1-scores than encoder-based models on Yahoo A4 and USC-HAD, while the two groups show similar performance on HASC (see Figure 10, Appendix D.4.1). These results suggest that architecture and pretraining characteristics may matter more than model or representation size for preserving change-relevant information.

Runtime comparison. We compare the proposed foundation-model (Chronos-2) approach with the unsupervised RuLSIF and self-supervised TS-CP^2 baselines. Our method is about $80\times$ faster than TS-CP^2, since it requires no training; it is on average about $13\times$ slower than RuLSIF run on CPU. When the series length increases by about $11\times$ and the number of channels by $3\times$, total runtime grows by about $33\times$ for RuLSIF, $14\times$ for TS-CP^2, and $12.6\times$ for our method on GPU ($9.4\times$ on CPU). Our framework therefore scales more favorably than RuLSIF and TS-CP^2, without its training cost (Table 7, Appendix D.5).

Generic event boundary detection. Table 2 shows that the proposed framework, using pretrained representations, is competitive with unsupervised GEBD methods on Kinetics-GEBD and TAPOS. It achieves the highest average F_1 on Kinetics-GEBD and the best performance at stricter tolerances on TAPOS (F_1@0.05 and F_1@0.10). The critical-difference analysis (Figure 9, Appendix D.6)



(a) (b) (c)

Figure 3: (a) Relationship between the change-encoding rate (CER) and downstream F_1-score across the time-series and video benchmarks. (b) Example projected univariate signal obtained from foundation-model representations, with detected change points indicated by vertical dashed lines. (c) Standardized Brent, WTI, diesel, and natural-gas prices over the corresponding evaluation period, together with the detected change points.

shows significant improvement over GraphGEBD and no significant difference from FlowGEBD, while avoiding normalized-cut graph partitioning. Complete results across tolerances are provided in Appendix D.4, Tables 8 and 9.

Table 2: **GEBD benchmark results across unsupervised methods.** Comparison of methods on Kinetics-GEBD and TAPOS using micro-F1. Best and second-best results are **bold** and underlined.

	Kinetics-GEBD					TAPOS				
Method	**.05**	**.10**	**.25**	**.50**	**Avg.**	**.05**	**.10**	**.25**	**.50**	**Avg.**
PA (Shou et al., 2021)	.396	.488	.544	.564	.527	.360	.459	.567	.615	.543
FlowGEBD-Ensemble (Gothe et al., 2024)	.713	.828	.862	.869	.845	.375	.502	**.658**	**.717**	**.623**
GraphGEBD-Res50 (Meiraz et al., 2025)	**.732**	.813	.842	.853	.830	.398	.493	.595	.661	.582
GAS-DINO-ViT B/16	.705	**.830**	**.890**	**.898**	**.849**	**.420**	**.529**	.615	.637	.569

Change-encoding and downstream detection. We examine how change encoding relates to downstream detection performance. Figure 3 compares the change encoding rate (CER) with the corresponding F_1-scores across the time-series and video benchmarks. A consistent pattern is that representations with low CER do not achieve high downstream F_1, as observed particularly on HASC and TAPOS (Figures 5 and 6, Appendix B). This indicates that preserving distributional changes in the representation space is an important prerequisite for successful detection. However, high CER alone does not guarantee high F_1, since the encoded change must also be preserved by the scalarization and accurately localized by the downstream detector. CER therefore provides a diagnostic of whether a frozen representation preserves detectable change information and can be used for downstream detection, rather than a direct criterion for ranking foundation models.

Effect of scalarization. More importantly, for time series benchmarks when evaluated on the validation set, the covariance-adjusted SDP gives the highest F_1-score. Moreover, we also carry out an oracle analysis (Table 6, Appendix D.4) to identify the best performance under each scalarization for each TSFM. This analysis shows that SDP-based statistics rank first across all five benchmarks, with covariance-adjusted SDP on Yahoo A4 and USC-HAD, while plain SDP performs best on HASC, Kinetics-GEBD, and TAPOS. Cosine similarity does not rank first on any benchmark (see Table 12 in Appendix D.6). These findings are consistent with the theoretical discussion in Section 3.3. For short video segments in GEBD benchmarks, $T \ll d$, making estimation of $\mathbf{\Sigma}$ unstable in high-dimensional settings. We therefore use SDP rather than the covariance-adjusted version.

Real-data case study. Among our benchmarks, only Yahoo A4 appears in a TSFM pre-training corpus (MOMENT's); to our knowledge, no other model was trained on our benchmark data. We therefore further evaluate the framework on 2136 hourly observations of Brent, WTI, diesel, and natural-gas prices from 1 February to 30 April 2026, which post-date the pre-training data of all models considered. The projected series and detected change points are shown in Figures 3b and 3c. The detected change points appear to align closely with abrupt drops and spikes in the projected series, as well as with corresponding Reuters-reported events affecting energy prices, listed in Table 11 in Appendix D.7.

4.3 Ablation Study

We ablate the main components of the framework to assess the contribution of the representation, denoising, and downstream detector, and its robustness to input noise.

Representation. To isolate the effect of foundation-model representations, we apply each scalarization directly to the raw input (Table 13, Appendix E.1). FM representations outperform raw inputs in all comparisons on Yahoo A4 and HASC, and in 154 of the 156 comparisons on USC-HAD; the only exceptions are Cov-SDP variants from the MOMENT family. One-sided paired Wilcoxon tests with Holm correction (Holm, 1979) confirm the improvement from FM representations.

Denoising. Now, we compare TV denoising with no denoising and coordinate-wise temporal median filtering. Across Chronos-2 and Chronos-T5 Tiny, median filtering provides only a small improvement, while TV denoising increases average F_1 from 0.546 to 0.757 (Figure 4a). Unlike local median filtering, TV denoising suppresses longer-term fluctuations while preserving abrupt changes. Its benefit is therefore largest when latent noise is high relative to the change magnitude and smaller when representations are already smooth.

Detector. After replacing SCAN with PELT, Binary Segmentation, and KernelCPD while fixing the Chronos-2 representation and Cov-SDP scalarization, we find that SCAN achieves the highest F_1 at all tolerances, followed by KernelCPD (Figure 4b). The Gaussian-cost methods perform worse because they primarily capture mean and variance changes, whereas KernelCPD and SCAN can detect broader distributional changes. SCAN further benefits from aggregating evidence using multiple window sizes with a data-driven threshold.

Noise robustness. We add Gaussian noise with $\sigma^2 \in \{0, 0.25, 0.5, 1, 2, 4\}$ to the standardized Yahoo A4 series and evaluate Chronos-T5 Tiny and Chronos-2 across all four scalarizations. Performance drops most sharply with the initial increase in noise and then largely plateaus, indicating slow degradation as the noise level increases further (Figure 4c).

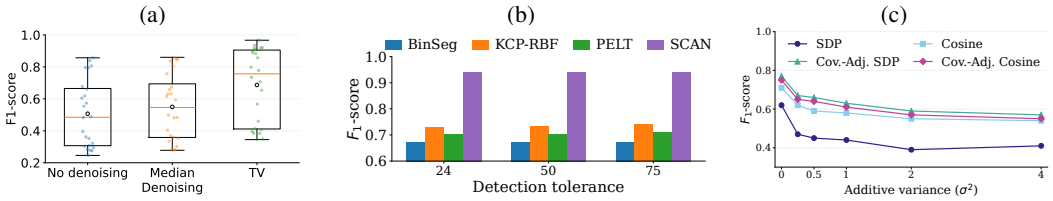


Figure 4: (a) Distribution of F_1-scores under no denoising, median filtering, and total-variation (TV) denoising across four scalarizations using Chronos-2 and Chronos-T5 Tiny. (b) Macro-F_1 of SCAN, PELT, Binary Segmentation, and KernelCPD across Yahoo A4 detection tolerances (oracle) on Chronos-2 embeddings. (c) Robustness of the four scalarization strategies to increasing additive noise variance at a detection tolerance of 24.

5 Conclusion

In this work, we investigate whether frozen foundation-model representations can support change-point detection (CPD) and generic event boundary detection (GEBD) without task-specific representation learning. Our GAS-FM framework combines a frozen encoder, covariance-aware GAS scalarization, and a nonparametric change-point detector in a common pipeline for multivariate time series and video. Across both modalities, pretrained representations preserve a large proportion of the annotated changes according to the proposed CER measure, and the resulting pipeline is competitive with self-supervised CPD and unsupervised GEBD baselines.

Encoder-based architectures generally outperform decoder-only models on the considered benchmarks. Low CER is consistently associated with low downstream F_1, supporting its use as a screening criterion for change-relevant representations. Theoretically, Proposition 1 provides sufficient conditions under which covariance-adjusted SDP-GAS preserves changes in the mean, covariance, or lag-one dependence of the latent process. These guarantees are not universal, since transformations that leave the scalarized distribution unchanged may remain undetectable. Future work includes data-driven denoising, broader classes of change-preserving scalarizations, and online extensions.

AI USE STATEMENT

The authors used `OpenAI ChatGPT (GPT-5.6)` solely for the purpose of language refinement, text proofreading, and improving clarity of expression during the manuscript preparation. All core concepts, technical formulations, and experimental analyses were conceived and written entirely by the human authors. The final text was thoroughly reviewed, verified, and edited by the authors, who maintain full responsibility for the content and accuracy of this work.

REPRODUCIBILITY STATEMENT

We have taken several steps to ensure reproducibility. Source code for the full GAS-FM framework, including representation extraction, denoising, GAS scalarization, and change-point detection, together with Python scripts for reproducing all experiments and their processing steps, are available at `https://anonymous.4open.science/r/anonymous_cpd-964F/`. All benchmark datasets are publicly available, and the energy-price data were obtained through the Dukascopy Bank API. Additional implementation and experimental details are provided in the Appendix E.2.

REFERENCES

Samaneh Aminikhanghahi and Diane J. Cook. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2):339–367, 2017.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024. arXiv:2403.07815.

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, Mononito Goswami, Shubham Kapoor, Danielle C. Maddix, Pablo Guerron, Tony Hu, Junming Yin, Nick Erickson, Prateek Mutalik Desai, Hao Wang, Huzefa Rangwala, George Karypis, Yuyang Wang, and Michael Bohlke-Schneider. Chronos-2: From univariate to universal forecasting, 2025. URL `https://arxiv.org/abs/2510.15821`.

Sylvain Arlot, Alain Celisse, and Zaid Harchaoui. A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research*, 20(162):1–56, 2019.

Alexander Aue and Lajos Horváth. Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16, 2013.

Alexandra Bazarova, Evgenia Romanenkova, and Alexey Zaytsev. Normalizing self-supervised learning for provably reliable change point detection. *arXiv preprint arXiv:2410.13637*, 2024.

Richard C. Bradley. Basic properties of strong mixing conditions: A survey and some open questions. *Probability Surveys*, 2:107–144, 2005. doi: 10.1214/154957805100000104.

Zhenxiang Cao, Nick Seeuws, Maarten De Vos, and Alexander Bertrand. Change point detection in multi-channel time series via a time-invariant representation. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):7743–7756, 2024. doi: 10.1109/TKDE.2023.3347356.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers, 2021. URL `https://arxiv.org/abs/2104.14294`.

Alain Celisse, Guillemette Marot, Morgane Pierre-Jean, and Guillem Rigaill. New efficient algorithms for multiple change-point detection with reproducing kernels. *Computational Statistics & Data Analysis*, 128:200–220, 2018.

Hao Chen and Nancy Zhang. Graph-based change-point detection. *The Annals of Statistics*, 43(1): 139–176, 2015.

Haeran Cho and Piotr Fryzlewicz. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):475–507, 2015.

Haeran Cho and Karolos K. Korkas. High-dimensional garch process segmentation with an application to value-at-risk. *Econometrics and Statistics*, 23:187–203, 2022. doi: 10.1016/j.ecosta.2021.07.009.

Jinmyeong Choi, Brad Shook, and Artur Dubrawski. Non-stationarity in the embedding space of time series foundation models. *arXiv preprint arXiv:2604.16428*, 2026. doi: 10.48550/arXiv.2604.16428.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, 2024. arXiv:2310.10688.

Tim De Ryck, Maarten De Vos, and Alexander Bertrand. Change point detection in time series data using autoencoders with a time-invariant representation. *IEEE Transactions on Signal Processing*, 69:3513–3524, 2021. doi: 10.1109/TSP.2021.3087031.

Shohreh Deldari, Daniel V. Smith, Amin Sadri, and Flora D. Salim. ESPRESSO: Entropy and shape aware time-series segmentation for processing heterogeneous sensor data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3):77:1–77:24, 2020. doi: 10.1145/3411832.

Shohreh Deldari, Daniel V. Smith, Hao Xue, and Flora D. Salim. Time series change point detection with self-supervised contrastive predictive coding. In *Proceedings of the Web Conference 2021 (WWW '21)*, pp. 3124–3135, 2021.

David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994. doi: 10.1093/biomet/81.3.425.

Farida Enikeeva and Zaid Harchaoui. High-dimensional change-point detection under sparse alternatives. *The Annals of Statistics*, 47(4):2051–2079, 2019.

Mark B. Fiecas, Kathryn R. Cullen, and Rebecca Killick. Generalised mixed effects models for changepoint analysis of biomedical time series data. *arXiv preprint arXiv:2410.00183*, 2024.

Roland Fried and Ann Cathrice George. *Median Filters and Extensions*, pp. 1450–1452. Springer Berlin Heidelberg, Berlin, Heidelberg, 2025. ISBN 978-3-662-69359-9. doi: 10.1007/978-3-662-69359-9_356. URL https://doi.org/10.1007/978-3-662-69359-9_356.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, 2024. arXiv:2402.03885.

Sourabh Vasant Gothe, Vibhav Agarwal, Sourav Ghosh, Jayesh Rajkumar Vachhani, Pranay Kashyap, and Barath Raj Kandur Raja. What's in the flow? exploiting temporal motion cues for unsupervised generic event boundary detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 6941–6950, 2024.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

Thomas Grundy, Rebecca Killick, and Gueorgui Mihaylov. High-dimensional changepoint detection via a geometrically inspired mapping. *Statistics and Computing*, 30(4):1155–1166, 2020.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

Lajos Horváth and Marie Hušková. Change-point detection in panel data. *Journal of Time Series Analysis*, 33(4):631–648, 2012.

Jaejun Hwang, Dayoung Gong, Manjin Kim, and Minsu Cho. Generic event boundary detection via denoising diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 14084–14094, 2025.

Robert I. Jennrich and Sidney C. Port. Radial and directional parts of a random vector. *Statistics & Probability Letters*, 6(3):155–158, February 1988. doi: None. URL https://ideas.repec.org/a/eee/stapro/v6y1988i3p155-158.html.

Moritz Jirak. Uniform change point tests in high dimension. *The Annals of Statistics*, 43(6):2451–2483, 2015.

Ian T. Jolliffe. *Principal Component Analysis*. Springer, New York, 2 edition, 2002.

Hyolim Kang, Jinwoo Kim, Taehyun Kim, and Seon Joo Kim. Uboco: Unsupervised boundary contrastive learning for generic event boundary detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20073–20082, 2022.

Nobuo Kawaguchi, Nobuhiro Ogawa, Yohei Iwasaki, Katsuhiko Kaji, Tsutomu Terada, Kazuya Murao, Sozo Inoue, Yoshihiro Kawahara, Yasuyuki Sumi, and Nobuhiko Nishio. Hasc challenge: Gathering large scale human activity corpus for the real-world activity understandings. In *Proceedings of the 2nd Augmented Human International Conference (AH '11)*, pp. 27:1–27:5. ACM, 2011. doi: 10.1145/1959826.1959853.

Rebecca Killick, Paul Fearnhead, and Idris A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012. doi: 10.1080/01621459.2012.737745.

Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004. doi: 10.3905/jpm.2004.110.

Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, and Qingsong Wen. Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 6555–6565. ACM, August 2024. doi: 10.1145/3637528.3671451. URL http://dx.doi.org/10.1145/3637528.3671451.

Song Liu, Makoto Yamada, Nigel Collier, and Masashi Sugiyama. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 43:72–83, 2013. doi: 10.1016/j.neunet.2013.01.012.

Xu Liu, Juncheng Liu, Gerald Woo, Taha Aksu, Yuxuan Liang, Roger Zimmermann, Chenghao Liu, Silvio Savarese, Caiming Xiong, and Doyen Sahoo. Moirai-moe: Empowering time series foundation models with sparse mixture of experts, 2024. URL https://arxiv.org/abs/2410.10469.

David S. Matteson and Nicholas A. James. A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345, 2014.

E T McGonigle and H Cho. Nonparametric data segmentation in multivariate time series via joint characteristic functions. *Biometrika*, 112(2), 2025. ISSN 1464-3510. doi: 10.1093/biomet/asaf024. URL http://dx.doi.org/10.1093/biomet/asaf024.

Ori Meiraz, Or Rubin, and Shlomi Laufer. *GraphGEBD: Zero-Shot Generic Event Boundary Detection Using Graph Representation of Videos*, pp. 243–258. 08 2025. ISBN 978-3-031-99964-2. doi: 10.1007/978-3-031-99965-9_16.

Alfred Müller. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, 29(2):429–443, 1997. doi: 10.2307/1428011.

Eyal Nitzan, Topi Halme, and Visa Koivunen. Bayesian methods for multiple change-point detection with reduced communication. *IEEE Transactions on Signal Processing*, 68:4871–4886, 2020. doi: 10.1109/TSP.2020.3016139.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision, 2024. URL https://arxiv.org/abs/2304.07193.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

Ashoka Prabashwara, Patricia Menéndez, Liam Hodgkinson, and Stuart Lee. Scan: Sequentially detecting change-points via adaptive nonparametric inference, 2026. URL https://arxiv.org/abs/2608.28110.

Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992. doi: 10.1016/0167-2789(92)90242-F.

Ashish Sen and Muni S. Srivastava. On tests for detecting change in mean. *The Annals of Statistics*, 3(1):98–108, 1975. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/2958081.

Mike Zheng Shou, Stan Weixian Lei, Weiyao Wang, Deepti Ghadiyaram, and Matt Feiszli. Generic event boundary detection: A benchmark for event segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8075–8084, 2021.

Alexander Stepikin, Evgenia Romanenkova, and Alexey Zaytsev. Wwaggr: A window wasserstein-based aggregation for ensemble change point detection, 2025. URL https://arxiv.org/abs/2506.08066.

Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

Martin Uray, Saverio Messineo, Roland Kwitt, and Stefan Huber. Exploring zero-shot foundation models for multivariate time series anomaly detection, 2026. URL https://arxiv.org/abs/2607.12454.

Tengyao Wang and Richard J. Samworth. High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):57–83, 2018.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, 2024. arXiv:2402.02592.

Yahoo! Inc. Yahoo! Webscope Program, Dataset S5: A Labeled Anomaly Detection Dataset. http://webscope.sandbox.yahoo.com, 2015. Version 1.0.

Mi Zhang and Alexander A. Sawchuk. Usc-had: A daily activity dataset for ubiquitous activity recognition using wearable sensors. In *ACM International Conference on Ubiquitous Computing (Ubicomp) Workshop on Situation, Activity and Goal Awareness (SAGAware)*, Pittsburgh, Pennsylvania, USA, September 2012.

Ziwei Zheng, Zechuan Zhang, Yulin Wang, Shiji Song, Gao Huang, and Le Yang. Rethinking the architecture design for efficient generic event boundary detection. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 1215–1224, 2024. doi: 10.1145/3664647.3681513.

A Theoretical Details and Proofs

A.1 Proof of Theorem 1

Proof. Let $(\Omega, \mathbb{F}, \mathbb{P})$ be a probability space, and let $\{\mathbf{x}_t\}_{t\in\mathbb{Z}}$ be a stochastic process defined on it, and $\mathbf{z}_t$ be the corresponding latent representations from problem definition under Section 2.2. For $n \in \mathbb{Z}$ and $\ell \geq 1$, define the past and future σ-algebras of the process by

$$\mathbb{F}^X_{-\infty,n} = \sigma(\mathbf{x}_t : t \leq n), \qquad \mathbb{F}^X_{n+\ell,\infty} = \sigma(\mathbf{x}_t : t \geq n + \ell).$$

For any two sub-σ-algebras $\mathbb{A}, \mathbb{B} \subseteq \mathbb{F}$, the α-mixing (strong mixing) coefficient between them is defined as

$$\alpha(\mathbb{A}, \mathbb{B}) = \sup_{A\in\mathbb{A},\, B\in\mathbb{B}} |\, \mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B)\,|\,.$$

We record a monotonicity property that will be used repeatedly: if $\mathbb{A}_1 \subseteq \mathbb{A}_2$ and $\mathbb{B}_1 \subseteq \mathbb{B}_2$, then

$$\alpha(\mathbb{A}_1, \mathbb{B}_1) \leq \alpha(\mathbb{A}_2, \mathbb{B}_2).$$

The α-mixing coefficient of the process $\{\mathbf{x}_t\}$ at lag ℓ is defined as $\alpha_X(\ell) = \sup_{n\in\mathbb{Z}} \alpha\left(\mathbb{F}^X_{-\infty,n},\, \mathbb{F}^X_{n+\ell,\infty}\right)$.

Since f is Borel measurable, $\mathbf{z}_t = f(\mathbf{x}_{t-w+1}, \ldots, \mathbf{x}_t)$ is measurable with respect to the σ-algebra $\sigma(\mathbf{x}_{t-w+1}, \ldots, \mathbf{x}_t)$. Therefore, for every $n \in \mathbb{Z}$, it satisfies $\mathbb{F}^Z_{-\infty,n} \subseteq \mathbb{F}^X_{-\infty,n}$, since every $\mathbf{z}_t$ with $t \leq n$ is constructed from observations $\mathbf{x}_j$ with indices $j \leq t \leq n$.

Similarly, for the future σ-algebra, if $t \geq n + \ell$, then the earliest observation entering $\mathbf{z}_t$ is $\mathbf{x}_{t-w+1}$, whose index satisfies $t - w + 1 \geq n + \ell - w + 1$. Hence it satisfies,

$$\mathbb{F}^Z_{n+\ell,\infty} \subseteq \mathbb{F}^X_{n+\ell-w+1,\infty}.$$

Then similarly from monotonicity, for every n,

$$\begin{aligned}\alpha\big(\mathbb{F}^Z_{-\infty,n}, \mathbb{F}^Z_{n+\ell,\infty}\big) &\leq \alpha\left(\mathbb{F}^X_{-\infty,n}, \mathbb{F}^X_{n+\ell-w+1,\infty}\right)\\ &\leq \alpha_X(\ell - w + 1).\end{aligned}$$

Taking the supremum over n gives

$$\alpha_Z(\ell) \leq \alpha_X(\ell - w + 1).$$

For $\ell \geq w$, we have $\ell - w + 1 \geq 1$, and under the assumed exponential α-mixing definition of $\{\mathbf{x}_t\}$,

$$\begin{aligned}\alpha_Z(\ell) &\leq c_1 \exp\{-c_2(\ell - w + 1)\}\\ &= c_1 e^{c_2(w-1)} e^{-c_2\ell}.\end{aligned}$$

Thus, with $\bar{c}_1 = c_1 e^{c_2(w-1)}$, we obtain $\alpha_Z(\ell) \leq \bar{c}_1 e^{-c_2\ell}$ for $\ell \geq w$. Enlarging $\bar{c}_1$ if necessary to cover the finitely many lags $1 \leq \ell < w$, $\{\mathbf{z}_t\}$ is exponentially α-mixing. □

A.2 Proof of Corollary 1

Proof. Consider the setting of Theorem 1 and given h is measurable, and the definition $q_t = h(\mathbf{z}_{t-m+1}, \ldots, \mathbf{z}_t)$, q_t is measurable with respect to $\sigma(\mathbf{z}_{t-m+1}, \ldots, \mathbf{z}_t)$. Additionally, each $\mathbf{z}_j$ is also a measurable function of $\mathbf{x}_{j-w+1}, \ldots, \mathbf{x}_j$. Therefore, the collection $\mathbf{z}_{t-m+1}, \ldots, \mathbf{z}_t$ depends only on the observations $\mathbf{x}_{t-m-w+2}, \ldots, \mathbf{x}_t$, such that, $q_t \in \sigma(\mathbf{x}_{t-w-m+2}, \ldots, \mathbf{x}_t)$. Therefore, it follows that

$$\mathbb{F}^q_{-\infty,n} \subseteq \mathbb{F}^X_{-\infty,n}.$$

Similarly for $t \geq n + \ell$, as in the proof of Theorem 1 we show that it satisfies,

$$\mathbb{F}^q_{n+\ell,\infty} \subseteq \mathbb{F}^X_{n+\ell-w-m+2,\infty}.$$

By monotonicity of the α-coefficient, these two inclusions give, for every n,

$$\alpha\left(\mathbb{F}^q_{-\infty,n}, \mathbb{F}^q_{n+\ell,\infty}\right) \leq \alpha\left(\mathbb{F}^X_{-\infty,n}, \mathbb{F}^X_{n+\ell-w-m+2,\infty}\right) \leq \alpha_X(\ell - w - m + 2).$$

Then taking the supremum over n yields $\alpha_q(\ell) \leq \alpha_X(\ell - w - m + 2)$. Thus, defining $\bar{c}_q = c_1 e^{c_2(w+m-2)}$, we obtain $\alpha_q(\ell) \leq \bar{c}_q\, e^{-c_2\ell}$ for all $\ell \geq w + m - 1$. Enlarging $\bar{c}_q$ if necessary to cover the finitely many lags $1 \leq \ell < w + m - 1$, $\{q_t\}$ is also exponentially α-mixing. □

A.3 Proof of Theorem 2

Proof. We verify the three conditions stated in SCAN (Prabashwara et al., 2026).

Exponential α-mixing. By Corollary 1, $q_t = h(\mathbf{z}_{t-m+1}, \ldots, \mathbf{z}_t)$ is exponentially α-mixing.

Minimal spacing. Each q_t is a measurable function of $\mathbf{x}_{t-L+1:t}$, where $L = w + m - 1$. Hence, a change at τ_j can affect only those scalarized observations whose effective input blocks intersect both adjacent regimes. This produces a transition region of length at most $L - 1$ following each change-point. Consequently, the distance between consecutive stationary interiors is at least $\tau_{\min} - 2(L-1)$. Since $L = o(\tau_{\min})$, we obtain $\tau_{\min} - 2(L-1) = \tau_{\min}\{1 - o(1)\}$. Thus, the finite-context transformation preserves the minimum-spacing requirement up to the transition regions.

Minimal signal magnitude. By assumption, there exists $\vartheta_{\mathcal{G}} > 0$ such that $d_{\mathcal{G}}(\mathrm{P}_q^{(j)}, \mathrm{P}_q^{(j+1)}) \geq \vartheta_{\mathcal{G}}$ for every $j = 1, \ldots, k$. Hence adjacent stationary regimes remain uniformly distinguishable under the IPM used by SCAN.

Therefore, the scalarized process satisfies the exponential α-mixing, minimum-spacing, and minimal-signal conditions. □

A.4 Proof of Proposition 1

Proposition 1. *Let $\{\mathbf{z}_t\}$ follows Definition 2 with $\mathbf{z}_t = \boldsymbol{\mu}_j + \boldsymbol{\varepsilon}_t$ within regime j. Suppose that noise $(\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_t)$ is Gaussian within each regime, with $\mathbb{E}_j[\boldsymbol{\varepsilon}_t] = 0$, $\mathbf{C}_j = \mathrm{Cov}_j(\mathbf{z}_t)$, and $\mathbb{E}_j[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-1}^\top] = \boldsymbol{\Gamma}_j(1)$. Let $\boldsymbol{\Sigma} \succ 0$ denote the covariance matrix of the latent representations $\{\mathbf{z}_t\}$. Then,*

(i) $\mathbb{E}_j[S_t] = \frac{1}{\sqrt{d}} \left[\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma}_j(1)\big)\right]$.

(ii)
$$\mathbb{E}_j[S_t^2] = \frac{1}{d}\Big[d\big(\mathbb{E}_j[S_t]\big)^2 + 2\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1} \left(\mathbf{C}_j + \boldsymbol{\Gamma}_j(1)\right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1} \mathbf{C}_j \boldsymbol{\Sigma}^{-1} \mathbf{C}_j\big) + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma}_j(1) \boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma}_j(1)\big)\Big].$$

Hence, changes in $\boldsymbol{\mu}_j$, $\mathbf{C}_j$, or $\boldsymbol{\Gamma}_j(1)$ that alter either (i) *or* (ii) *induce a change in the covariance-adjusted SDP series.*

Proof. **(i)** For a regime j, consider $\mathbf{z}_{t-1} = \boldsymbol{\mu}_j + \boldsymbol{\varepsilon}_{t-1}$, and $\mathbf{z}_t = \boldsymbol{\mu}_j + \boldsymbol{\varepsilon}_t$. Substituting into the definition of S_t gives

$$\begin{aligned}\sqrt{d}\, S_t &= (\boldsymbol{\mu}_j + \boldsymbol{\varepsilon}_{t-1})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_j + \boldsymbol{\varepsilon}_t) \\ &= \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j + \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t + \boldsymbol{\varepsilon}_{t-1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j \\ &\quad + \boldsymbol{\varepsilon}_{t-1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t. \end{aligned} \tag{1}$$

Taking expectation of equation 1 and given that $\mathbb{E}_j[\boldsymbol{\varepsilon}_{t-1}] = \mathbb{E}_j[\boldsymbol{\varepsilon}_t] = 0$ and $\mathbb{E}_j[\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}^\top] = \mathbb{E}_j[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top] = \mathbf{C}_j$, and $\mathbb{E}_j[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-1}^\top] = \boldsymbol{\Gamma}_j$.

$$\begin{aligned}\mathbb{E}_j[\boldsymbol{\varepsilon}_{t-1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t] &= \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} \mathbb{E}_j[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-1}^\top]\right) \\ &= \mathrm{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma}_j(1)).\end{aligned}$$

Therefore,

$$\mathbb{E}_j[S_t] = \frac{1}{\sqrt{d}} \left[\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j + \mathrm{tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma}_j(1))\right]$$

(ii) For notational simplicity we define $Q_t = \mathbf{z}_{t-1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{z}_t$, and $S_t = \frac{Q_t}{\sqrt{d}}$. For the second moment, we define Q_t^2, such that $\mathbb{E}_j[S_t^2] = \frac{\mathbb{E}_j[Q_t^2]}{d}$. Then,

$$\begin{aligned}\mathbb{E}_j[Q_t^2] &= \left(\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j\right)^2 + 2\left(\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j\right) \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma}_j(1)\big) \\ &\quad + \mathbb{E}_j\left[\left(\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t + \boldsymbol{\varepsilon}_{t-1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j\right)^2\right] \\ &\quad + \mathbb{E}_j\left[\left(\boldsymbol{\varepsilon}_{t-1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t\right)^2\right]. \end{aligned} \tag{2}$$

Then $\mathbb{E}_j\Big[\big(\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\big)^2\Big] = \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\mathbf{C}_j\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j$, and similarly, $\mathbb{E}_j\Big[\big(\boldsymbol{\varepsilon}_{t-1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j\big)^2\Big] = \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\mathbf{C}_j\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j$. Additionally, $\mathbb{E}_j\big[\big(\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\big)\big(\boldsymbol{\varepsilon}_{t-1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j\big)\big] = \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_j(1)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j$. Therefore, $\mathbb{E}_j\Big[\big(\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t + \boldsymbol{\varepsilon}_{t-1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j\big)^2\Big] = 2\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\left(\mathbf{C}_j + \boldsymbol{\Gamma}_j(1)\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j$.

Moreover, the fourth Gaussian moment simplification yields,

$$\begin{aligned}\mathbb{E}_j\Big[\big(\boldsymbol{\varepsilon}_{t-1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t\big)^2\Big] = &\big[\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_j(1)\big)\big]^2 \\ &+ \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\mathbf{C}_j\boldsymbol{\Sigma}^{-1}\mathbf{C}_j\big) \\ &+ \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_j(1)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_j(1)\big)\,.\end{aligned} \tag{3}$$

Then, by equation 2 and equation 3, $\mathbb{E}_j[Q_t^2]$ follows,

$$\begin{aligned}\mathbb{E}_j[Q_t^2] = &\big(\mathbb{E}_j[Q_t]\big)^2 \\ &+ 2\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\left(\mathbf{C}_j + \boldsymbol{\Gamma}_j(1)\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j \\ &+ \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\mathbf{C}_j\boldsymbol{\Sigma}^{-1}\mathbf{C}_j\big) \\ &+ \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_j(1)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_j(1)\big)\,.\end{aligned}$$

Finally, since $S_t = Q_t/\sqrt{d}$,

$$\begin{aligned}\mathbb{E}_j[S_t^2] = \frac{1}{d}\Big[&d\big(\mathbb{E}_j[S_t]\big)^2 + 2\boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\left(\mathbf{C}_j + \boldsymbol{\Gamma}_j(1)\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j \\ &+ \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\mathbf{C}_j\boldsymbol{\Sigma}^{-1}\mathbf{C}_j\big) + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_j(1)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_j(1)\big)\Big].\end{aligned} \tag{4}$$

Thus, any regime change that alters either the first or second moment of the Covariance-adjusted SDP scalar process induces a change in its distribution. Whenever $d_{\mathcal{G}}$ separates the corresponding scalar distributions,

$$d_{\mathcal{G}}\left(\mathrm{P}_q^{(j)}, \mathrm{P}_q^{(j+1)}\right) \geq \vartheta_{\mathcal{G}}$$

for some $\vartheta_{\mathcal{G}}$. □

B Empirical Evaluation of Change Encoding

B.1 Change-encoding signal-to-noise ratio (CE-SNR)

Let $k(\mathbf{z}, \mathbf{z}') = \exp\left(-\|\mathbf{z} - \mathbf{z}'\|^2/(2h^2)\right)$ be a Gaussian RBF kernel with reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ and a canonical feature map $\phi(\mathbf{z}) = k(\mathbf{z}, \cdot)$, so that $\langle\phi(\mathbf{z}), \phi(\mathbf{z}')\rangle_{\mathcal{H}} = k(\mathbf{z}, \mathbf{z}')$. For an annotated change point τ_j, let $Z_j^- = \{\mathbf{z}_t : t = \tau_j - n, \ldots, \tau_j - 1\}$ and $Z_j^+ = \{\mathbf{z}_t : t = ta_j + w - 1, \ldots, ta_j + w + n - 2\}$ denote the n latent representations immediately before and after the change. We define the change-encoding signal-to-noise ratio (CE-SNR) as

$$S_j = \frac{\widehat{\mathrm{MMD}}_k^2(Z_j^-, Z_j^+)}{\frac{1}{2}\left[\widehat{V}_{\mathcal{H}}(Z_j^-) + \widehat{V}_{\mathcal{H}}(Z_j^+) + \xi\right]},$$

where $\xi > 0$ is a small stabilizing constant avoiding division by 0 and $\widehat{\mathrm{MMD}}_k^2$ is the squared maximum mean discrepancy between the two windows computed using a Gaussian RBF kernel, which is positive semi-definite, with the biased empirical estimator

$$\widehat{\mathrm{MMD}}^2 = \frac{1}{n}\sum_{i,j=1}^{n} k(x_i, x_j) + \frac{1}{n}\sum_{i,j=1}^{n} k(y_i, y_j) - \frac{2}{n}\sum_{i,j=1}^{n} k(x_i, y_j),$$

thereby ensuring that $\widehat{\mathrm{MMD}}^2 \geq 0$. For a set $Z = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$,

$$\widehat{V}_{\mathcal{H}}(Z) = \frac{1}{n-1} \sum_{i=1}^{n} \left\| \phi(\mathbf{z}_i) - \widehat{\boldsymbol{\mu}}_Z \right\|_{\mathcal{H}}^2, \qquad \widehat{\boldsymbol{\mu}}_Z = \frac{1}{n} \sum_{i=1}^{n} \phi(\mathbf{z}_i),$$

is the empirical variance of the embedded points in $\mathcal{H}$. This can be written as $\widehat{V}_{\mathcal{H}}(Z) = \frac{1}{n-1}\left[\sum_i k(\mathbf{z}_i, \mathbf{z}_i) - \frac{1}{n}\sum_{i,l} k(\mathbf{z}_i, \mathbf{z}_l)\right]$, so it can be computed from kernel evaluations alone. Thus S_j measures the between-regime discrepancy relative to the average within-regime variation, with larger values indicating that the inter-regime discrepancy is higher than the intra-regime.

Since the scale of S_j depends on the representation, we calibrate it using an empirical null distribution. We define a time index u as an *admissible no-change pseudo-split* if the sampled latent representations $Z_u^- = \{\mathbf{z}_t : u-n, \ldots, u-1\}$, and $Z_u^+ = \{\mathbf{z}_t = u, \ldots, u+n-1\}$ are both within the same segment, so that all latent representations are computed from inputs from the same segment. We sample a sufficiently large number of admissible pseudo-splits (s), chosen to ensure adequate resolution for the calibration level α, and compute $S_u^{(0)}$ using the same window length. These values form the empirical null distribution $\widehat{F}_0$. For a null exceedance level α, we define the threshold $\widehat{c}_\alpha = \widehat{F}_0^{-1}(1-\alpha)$ and regard an annotated change point τ_j as encoded if $S_j > \widehat{c}_\alpha$.

Definition 5 (Change-encoding rate)**.** *For the set of annotated change points $\{\tau_1, \ldots, \tau_k\}$, the* change-encoding rate *is*

$$\mathrm{CER}_\alpha(f) = \frac{1}{k} \sum_{j=1}^{k} \mathbf{1}\{S_j > \widehat{c}_\alpha\}.$$

Thus, $\mathrm{CER}_\alpha(f)$ is the proportion of annotated changes whose CE-SNR exceeds the $(1-\alpha)$ null quantile. We also report the *false change-encoding rate* (FCER), defined as the proportion of the set of all admissible pseudo-splits exceeding the same threshold. CER measures sensitivity to annotated changes, while FCER quantifies spurious change encoding. We use this framework to assess whether pretrained foundation models act as change encoders.

B.2 Change-encoding rate (CER)

Algorithm 1 summarizes the empirical computation of the Change-Encoding Rate (CER). CE-SNR is evaluated at each annotated change point and compared with a null distribution estimated from admissible no-change pseudo-splits. CER is the proportion of annotated changes exceeding the resulting null threshold.

Remark 1. *Considering $\mathbf{z}_t = f(x_{t-w+1:t})$, representations in the contamination region $\mathcal{C}_j = \{t : \tau_j \leq t \leq \tau_j + w - 2\}$ mix pre- and post-change observations. Therefore, we exclude $\mathcal{C}_j$ and take the post-change window as $Z_j^+ = \{z_t : t = \tau_j + w - 1, \ldots, \tau_j + w + n - 2\}$, so that each CE-SNR compares representations computed entirely within a single regime.*

B.3 Testing procedure

We empirically evaluate whether pretrained foundation-model encoders satisfy the change-encoding property by applying Algorithm 1 to five benchmark datasets. For TSFMs, we use context length $w = 128$ and $n = 40$ representations per side, while for vision foundation models we use $w = 10$ frames and $n = 10$. In both cases, contaminated regions due to context windows are around each annotated change point are excluded to avoid weakening the MMD signal. We report CER at calibration level $\alpha = 0.1$. As shown in Figures 5 and 6, CER is consistently high on Yahoo A4, USC-HAD, and Kinetics-GEBD, but substantially lower on HASC and TAPOS-GEBD, indicating that the extent to which pretrained representations preserve change information is dataset-dependent.

As a calibration check, we also evaluate CE-SNR at admissible no-change pseudo-splits and report the false change-encoding rate (FCER). FCER remains around the nominal level of 0.10 across models and datasets.

Algorithm 1 Empirical Change-Encoding Evaluation

Input: Representation sequences $\mathbf{Z}^{(r)} = \{\mathbf{z}_t^{(r)}\}_{t=w}^{T_r}$; annotated change-point sets $\mathcal{T}^{(r)} = \{\tau_1^{(r)}, \ldots, \tau_{k_r}^{(r)}\}$, $r = 1, \ldots, R$; context length w; sample size n; calibration level α
Output: Change-encoding rate CER_α;

`//` $\mathcal{S}_{\mathrm{chg}}$`:` `CE-SNR scores at annotated change points`
$\mathcal{S}_{\mathrm{chg}} \leftarrow \emptyset$
`//` $\mathcal{S}_{\mathrm{null}}$`:` `CE-SNR scores at no-change pseudo-splits`
$\mathcal{S}_{\mathrm{null}} \leftarrow \emptyset$
for $r = 1, \ldots, R$ **do**
  `// Evaluate annotated change points`
  foreach $\tau_j^{(r)} \in \mathcal{T}^{(r)}$ **do**
    Select n clean representations on both sides of $\tau_j^{(r)}$
    $Z_j^{-,(r)} \leftarrow \{\mathbf{z}_t^{(r)} : t = \tau_j^{(r)} - n, \ldots, \tau_j^{(r)} - 1\}$
    $Z_j^{+,(r)} \leftarrow \{\mathbf{z}_t^{(r)} : t = \tau_j^{(r)} + w - 1, \ldots, \tau_j^{(r)} + w + n - 2\}$
    Compute the CE-SNR at the annotated change point $\tau_j^{(r)}$; Section 3.2
    $$S_j^{(r)} \leftarrow \frac{\widehat{\mathrm{MMD}}_k^2\left(Z_j^{-,(r)}, Z_j^{+,(r)}\right)}{\frac{1}{2}\left[\widehat{V}_{\mathcal{H}}\left(Z_j^{-,(r)}\right) + \widehat{V}_{\mathcal{H}}\left(Z_j^{+,(r)}\right)\right]}$$
    $\mathcal{S}_{\mathrm{chg}} \leftarrow \mathcal{S}_{\mathrm{chg}} \cup \{S_j^{(r)}\}$
  `// Construct empirical no-change scores`
  Sample admissible pseudo-splits $\mathcal{U}_0^{(r)}$ from annotated no-change regions
  foreach $u \in \mathcal{U}_0^{(r)}$ **do**
    Extract n clean representations Z_u^- and Z_u^+ on either side of u, with both samples contained within the same regime
    $S_u^{(0)} \leftarrow S(Z_u^-, Z_u^+)$
    $\mathcal{S}_{\mathrm{null}} \leftarrow \mathcal{S}_{\mathrm{null}} \cup \{S_u^{(0)}\}$
`// Estimate the empirical null distribution`
Let $\widehat{F}_0$ be the empirical distribution of $\mathcal{S}_{\mathrm{null}}$: $\widehat{c}_\alpha \leftarrow \widehat{F}_0^{-1}(1-\alpha)$
`// Compute change-encoding rate`
$$\mathrm{CER}_\alpha \leftarrow \frac{\sum_{S \in \mathcal{S}_{\mathrm{chg}}} \mathbf{1}\{S > \widehat{c}_\alpha\}}{|\mathcal{S}_{\mathrm{chg}}|}$$ **return** CER_α

B.4 Evaluation results

The results from the experiment in Section B.3 for 13 TSFMs and vision foundation models are shown in Figures 5 and 6.

Additionally, we conduct a comparison study by evaluating the CER of latent representations obtained from self-supervised architectures trained specifically for change-point detection and from foundation models. The results indicate that foundation-model representations exhibit stronger change-encoding ability than those learned by the self-supervised methods.

C Anisotropy of Foundation-Model latent representations

Foundation-model representations do not exhibit approximately isotropic geometry. Figure 8 shows the ordered eigenvalue spectra of the empirical covariance matrices for representative time-series and vision foundation models. In both cases, the eigenvalues span several orders of magnitude, with a small number of directions accounting for substantially greater variation than the remaining dimensions. This pronounced spectral decay indicates strong anisotropy in the representation space.

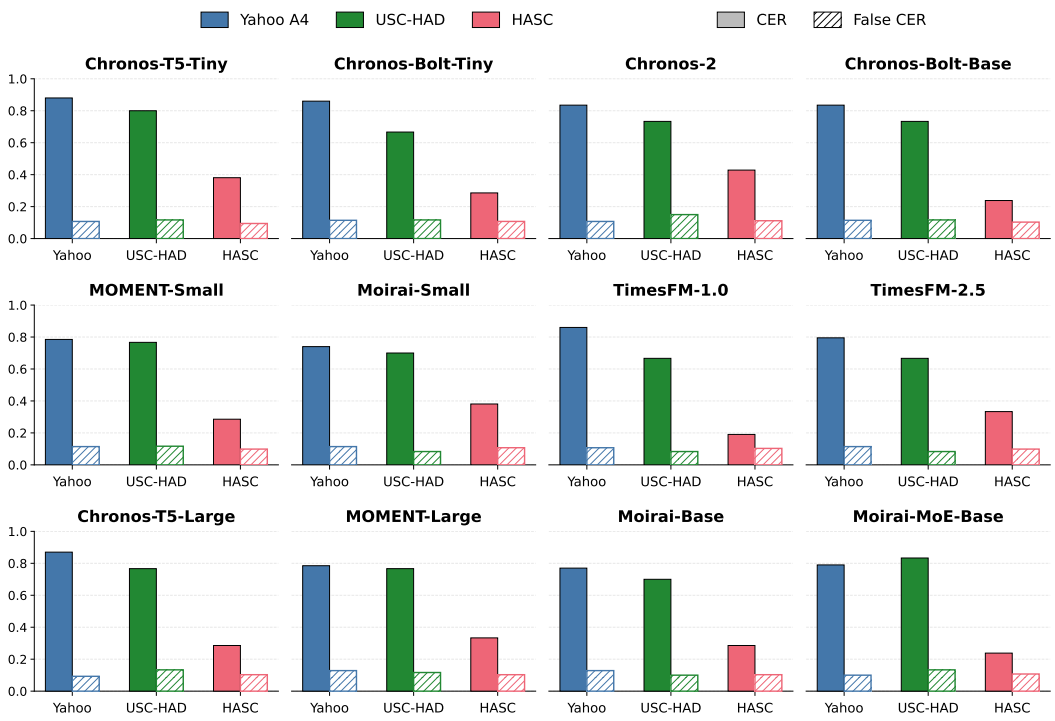


Figure 5: **Change-encoding evaluation of time-series foundation models.** Change encoding rate (CER; solid bars) and False CER (hatched bars) at calibration level $\alpha = 0.10$ across Yahoo A4, USC-HAD, and HASC. Higher CER indicates that a larger proportion of annotated changes remain statistically distinguishable in the frozen representation space, whereas False CER measures the corresponding rejection rate at non-change locations.

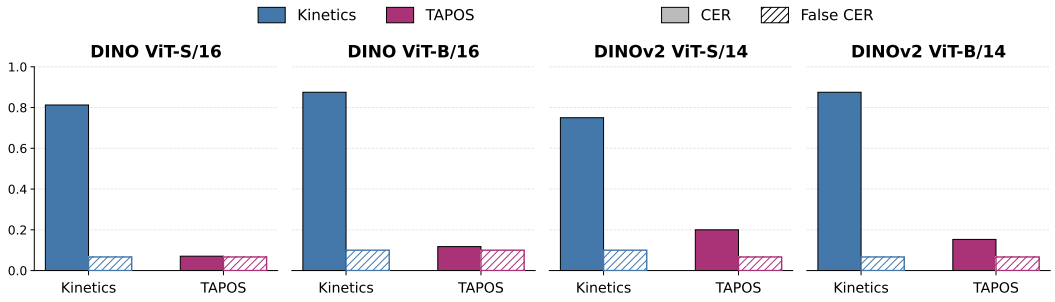


Figure 6: **Change-encoding evaluation of vision foundation models.** CER and False CER at calibration level $\alpha = 0.10$ on Kinetics-GEBD and TAPOS-GEBD. Results are shown for DINO ViT-S/16, DINO ViT-B/16, DINOv2 ViT-S/14, and DINOv2 ViT-B/14.

Consequently, treating all dimensions in the latent space as equally scaled can obscure the underlying representation geometry, motivating the covariance-aware statistics considered in Section 3.3.

D Experimental Details

D.1 Evaluated foundation models

Table 3 lists the frozen foundation models evaluated in the experiments. Model names are reported using the canonical architecture names used by the corresponding model families.

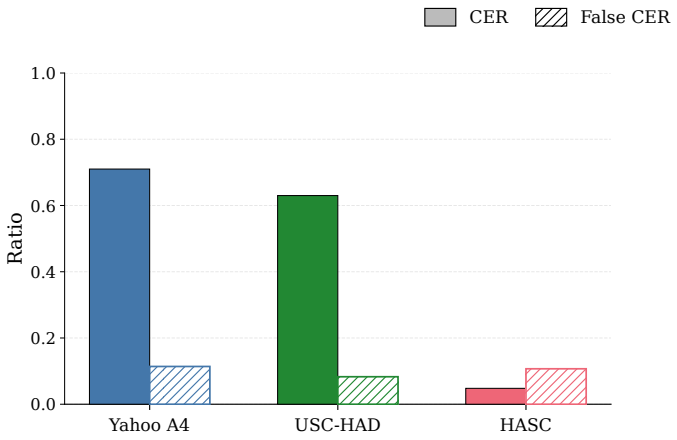


Figure 7: **Change-encoding evaluation of the TS-CP^2 self-supervised change-point detection method.** CER and False CER at calibration level $\alpha = 0.10$ are reported on Yahoo A4, USC-HAD, and HASC. The resulting CER values are lower than those obtained from the foundation-model representations.

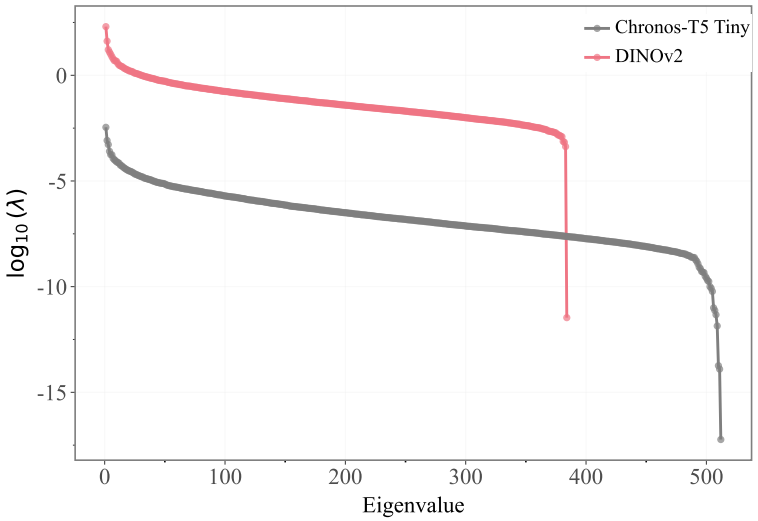


Figure 8: **Eigenvalue spectra of foundation-model representations.** Ordered eigenvalues of the empirical covariance matrix of latent representations from Chronos-T5 Tiny and DINOv2, shown on a $\log_{10}$ scale.

D.2 Latent representation extraction

Time-series foundation models. We consider TSFMs spanning three architecture types: encoder-only (Chronos-2, MOMENT, Moirai), encoder-decoder (Chronos-T5, Chronos-Bolt), and decoder-only (Moirai-MoE, TimesFM). For encoder-only and encoder-decoder models, we use the final encoder hidden states and for decoder-only models, we use the final hidden state before the forecasting head. Token- or patch-level representations are mean-pooled over the sequence dimension, to obtain a single representation for each context window. Inputs are passed through each model's native scaling. For univariate TSFMs applied to multivariate data, each channel is encoded independently, and the channel representations are combined by mean or max pooling, selected on the validation split.

Vision foundation models. For GEBD, we use encoder-only vision transformers (DINO and DINOv2). All frames are processed at each video's native frame rate without temporal subsampling and are resized to 224×224. For each frame, we extract the final-layer CLS-token representation.

Table 3: Evaluated foundation models considered in the study with model size (number of parameters).

Modality	Model family	Model (parameters)
Time series	Chronos-T5 (Ansari et al., 2024)	Tiny (8M) Large (710M)
	Chronos-Bolt (Ansari et al., 2024)	Tiny (9M) Base (205M)
	Chronos-2 (Ansari et al., 2025)	120M
	MOMENT (Goswami et al., 2024)	Small (40M) Large (385M)
	Moirai (Woo et al., 2024)	Small (14M) Base (91M)
	Moirai-MoE (Liu et al., 2024)	Small (117M total; 11M active) Base (935M total; 86M active)
	TimesFM (Das et al., 2024)	1.0 (200M) 2.5 (200M)
Vision	DINO (Caron et al., 2021)	ViT-S/16 (21M) ViT-B/16 (85M)
	DINOv2 (Oquab et al., 2024)	ViT-S/14 (21M) ViT-B/14 (86M)

For GEBD, we use a context window of 5 frames; we obtain the representation by applying mean or max pooling over the frame-level CLS embeddings, with the pooling operation selected on the validation set.

D.3 Hyperparameter Settings

During evaluation, we use a validation-set protocol because the framework contains two dataset-dependent hyperparameters: the context-window length and the denoising coefficient. We also include the four scalarization statistics in the validation grid. The resulting hyperparameter search space is summarized in Table 4.

Table 4: **Hyperparameter search space.** Parameters are selected using the validation set.

Hyperparameter	Search space
Context length w	32, 64, 128, 256, 512
Scalarization statistic	Scaled Dot Product (Raw), Cosine similarity (Raw), Covariance-adjusted SDP, Covariance-adjusted Cosine
TV regularization λ	0, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0

The hyperparameters reported in the table are dataset-specific. However, the downstream detector SCAN is applied separately to each projected series and implemented via `scan-cpd` library, and the minimum and maximum window sizes used in the ensemble are determined by the length of the corresponding input series.

D.4 Results

D.4.1 Test results across TSFMs

For each foundation model and dataset, we select the context-window length w, the scalarization statistic, and the denoising parameter λ on the validation split using the parameter grid in Table 4. Covariance-adjusted SDP was selected for every model, and the resulting test-set results are reported in Table 5.

Table 5: **Change-point detection performance across time-series foundation models on the test set with validation-selected hyperparameters.** F1-scores are reported at three tolerance levels and their average for Yahoo A4, USC-HAD, and HASC. Covariance-adjusted SDP is used for all models, as it was the best-performing on the validation set.

	Yahoo A4				USC-HAD				HASC			
Model	**24**	**50**	**75**	**Avg.**	**100**	**200**	**400**	**Avg.**	**60**	**100**	**200**	**Avg.**
Chronos-T5-Tiny	.485	.630	.719	.611	.764	.932	.932	.876	.346	.447	.543	.445
Chronos-Bolt-Tiny	.616	.661	.674	.650	.740	**.977**	**.977**	.898	.304	.400	.491	.398
Chronos-2	**.732**	**.836**	**.861**	**.810**	.741	.932	.955	.876	.394	.446	.529	.456
Chronos-Bolt-Base	.572	.662	.680	.638	.701	.950	.955	.868	.341	.408	.510	.419
Chronos-T5-Large	.458	.637	.713	.603	.831	.936	.936	.901	.353	.472	.548	.458
MOMENT-Small	.518	.610	.622	.583	.608	.704	.816	.709	.353	.400	.510	.421
Moirai-Small	.646	.712	.771	.710	.666	.715	.737	.706	.345	.440	.554	.446
Moirai-MoE-Small	.522	.619	.666	.603	.727	.958	.958	.881	.355	.450	.535	.447
TimesFM-1.0	.490	.595	.617	.567	.739	.881	.950	.856	.326	.360	.475	.387
TimesFM-2.5	.523	.613	.673	.603	.569	.712	.840	.707	.373	.478	.552	.468
MOMENT-Large	.604	.667	.692	.654	.805	.955	**.977**	.912	.346	.471	.562	.459
Moirai-Base	.580	.656	.710	.649	**.849**	.950	.950	**.916**	.396	.508	**.603**	**.502**
Moirai-MoE-Base	.493	.575	.621	.563	.655	.859	.864	.793	.361	.453	.529	.448

D.5 Runtime Comparison

Table 7: **Runtime comparison.** Per-series wall-clock time in seconds on Yahoo A4 (T=1,680, 1 channel) and USC-HAD (T=18,900, 3 channels). *Increase* is the ratio of USC-HAD to Yahoo A4 total time. GPU: NVIDIA L40S. Our method and RuLSIF require no training.

		Yahoo A4			USC-HAD			
Method	Device	Fit	Infer.	Total	Fit	Infer.	Total	Increase
RuLSIF	CPU	-	0.09	0.09	-	2.96	2.96	32.9×
Ours (Chronos-2 + Cov.-Adj. SDP)	GPU	-	1.72	1.72	-	21.60	21.60	12.6×
Ours (Chronos-2 + Cov.-Adj. SDP)	CPU	-	17.48	17.48	-	164.40	164.40	9.4×
TS-CP^2 (100 epochs)	GPU	130.51	0.65	131.16	1874.59	2.17	1876.76	14.3×
TS-CP^2 (100 epochs)	CPU	179.89	1.66	181.55	2556.61	14.19	2570.80	14.2×

D.6 Additional GEBD Results

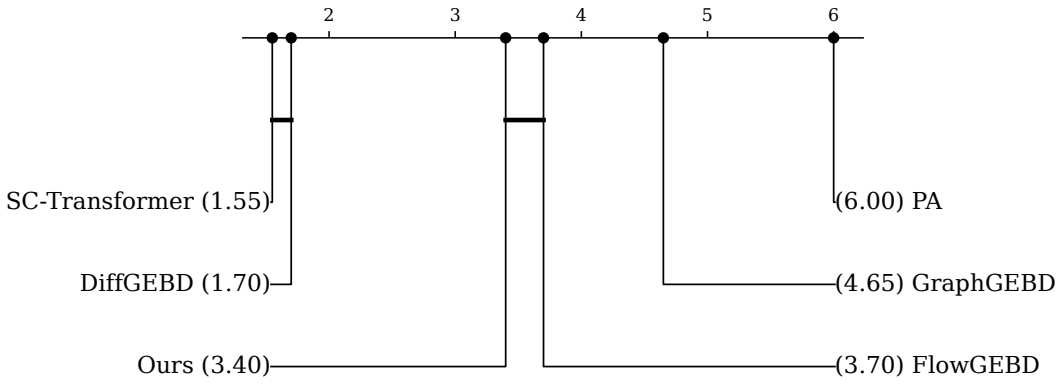


Figure 9: Critical-difference diagram comparing GEBD methods across the evaluation thresholds used in the benchmark analysis. Lower average rank is better; horizontal bars connect methods that are not significantly different.

Table 6: **Time-series foundation-model results across scalarization statistics (oracle results).** Best F1 scores on Yahoo A4, USC-HAD, and HASC across the corresponding evaluation tolerances. SDP denotes scaled dot product and Cov.-Adj. denotes covariance-adjusted similarity.

Model	Statistic	Yahoo A4			USC-HAD			HASC		
		24	50	75	100	200	400	60	100	200
Chronos-T5 Tiny	SDP	.567	.601	.616	.857	.857	.857	.460	.460	.540
	Cosine	.778	.782	.782	.967	.967	.967	.384	.416	.480
	Cov.-Adj. SDP	.707	.712	.727	.918	.951	.984	.407	.513	.566
	Cov.-Adj. Cosine	.654	.665	.670	.921	.921	.921	.349	.403	.443
Chronos-Bolt Tiny	SDP	.600	.604	.615	.862	.862	.862	.441	.500	.529
	Cosine	.840	.840	.840	.915	.915	.915	.386	.386	.470
	Cov.-Adj. SDP	.825	.825	.825	.921	.921	.921	.411	.434	.549
	Cov.-Adj. Cosine	.861	.861	.861	.933	.933	.933	.411	.440	.553
Chronos-2	SDP	.798	.798	.813	.735	.765	.824	.396	.426	.475
	Cosine	.889	.889	.894	.903	.903	.903	.385	.400	.446
	Cov.-Adj. SDP	.917	.917	.917	.935	.935	.935	.413	.477	.555
	Cov.-Adj. Cosine	.884	.889	.889	.912	.912	.912	.345	.446	.518
Chronos-Bolt Base	SDP	.647	.660	.668	.814	.814	.814	.420	.521	.655
	Cosine	.830	.830	.830	.900	.900	.900	.444	.540	.635
	Cov.-Adj. SDP	.848	.848	.848	.852	.852	.852	.439	.500	.530
	Cov.-Adj. Cosine	.801	.801	.801	.881	.881	.881	.427	.489	.519
MOMENT Small	SDP	.695	.695	.700	.806	.806	.806	.380	.479	.563
	Cosine	.781	.781	.781	.857	.889	.889	.342	.427	.496
	Cov.-Adj. SDP	.816	.816	.816	.915	.949	.949	.409	.439	.485
	Cov.-Adj. Cosine	.830	.835	.835	.897	.897	.897	.362	.456	.550
Moirai Small	SDP	.697	.706	.706	.786	.786	.786	.423	.511	.613
	Cosine	.799	.799	.804	.929	.929	.929	.425	.442	.549
	Cov.-Adj. SDP	.854	.854	.854	.933	.933	.933	.385	.397	.500
	Cov.-Adj. Cosine	.831	.836	.836	.881	.881	.915	.411	.477	.543
Moirai-MoE Small	SDP	.615	.628	.654	.881	.881	.881	.432	.496	.576
	Cosine	.777	.777	.782	.877	.912	.912	.371	.419	.516
	Cov.-Adj. SDP	.777	.781	.781	.909	.909	.909	.400	.419	.457
	Cov.-Adj. Cosine	.803	.803	.808	.867	.867	.867	.357	.393	.464
TimesFM 1.0	SDP	.707	.720	.738	.857	.857	.857	.420	.464	.507
	Cosine	.759	.759	.759	.893	.893	.893	.360	.396	.505
	Cov.-Adj. SDP	.821	.821	.821	.918	.918	.918	.366	.427	.549
	Cov.-Adj. Cosine	.785	.785	.785	.862	.862	.897	.378	.422	.486
TimesFM 2.5	SDP	.683	.692	.706	.862	.862	.862	.435	.503	.571
	Cosine	.706	.711	.711	.931	.931	.931	.400	.424	.570
	Cov.-Adj. SDP	.849	.849	.849	.846	.846	.846	.380	.430	.512
	Cov.-Adj. Cosine	.792	.792	.798	.949	.949	.983	.384	.424	.525
Chronos-T5 Large	SDP	.578	.593	.612	.879	.879	.879	.365	.406	.457
	Cosine	.777	.777	.777	.931	.966	.966	.425	.442	.496
	Cov.-Adj. SDP	.747	.747	.753	.903	.935	.968	.450	.523	.577
	Cov.-Adj. Cosine	.634	.634	.651	.862	.862	.862	.390	.441	.492
MOMENT Large	SDP	.661	.670	.679	.806	.806	.806	.368	.397	.515
	Cosine	.782	.782	.791	.921	.921	.921	.442	.496	.549
	Cov.-Adj. SDP	.795	.800	.805	.754	.812	.841	.455	.504	.520
	Cov.-Adj. Cosine	.792	.792	.797	.877	.912	.912	.400	.508	.585
Moirai Base	SDP	.678	.678	.691	.786	.786	.786	.390	.488	.585
	Cosine	.789	.794	.799	.833	.833	.833	.340	.383	.482
	Cov.-Adj. SDP	.859	.859	.859	.935	.935	.968	.432	.531	.630
	Cov.-Adj. Cosine	.836	.836	.836	.881	.915	.949	.366	.441	.527
Moirai-MoE Base	SDP	.614	.626	.626	.900	.900	.900	.458	.519	.611
	Cosine	.772	.777	.787	.915	.915	.915	.330	.348	.435
	Cov.-Adj. SDP	.810	.810	.819	.947	.947	.947	.348	.398	.484
	Cov.-Adj. Cosine	.813	.813	.818	.912	.912	.912	.361	.403	.462

In Figure 9, we compare the validation-set rankings of our framework with supervised GEBD methods, including SC-Transformer and DiffGEBD, as well as unsupervised GEBD baselines. The proposed foundation-model + SCAN framework significantly outperforms GraphGEBD, a closely related approach that combines DINO or ResNet-50 representations with graph-based boundary detection. In addition to its competitive detection performance, our framework benefits from the near-linear computational complexity of SCAN, providing an efficient alternative to graph-based GEBD approaches.

Table 8: Effect of the scalar similarity statistic across selected vision foundation models on Kinetics-GEBD. We report micro-F1 across relative-distance thresholds. Bold marks the best statistic for each model at each threshold.

Statistic	.05	.06	.07	.10	.15	.20	.25	.30	.35	.40	.45	.50	Avg.
DINO ViT-B/16													
Cosine	.607	.639	.663	.710	.752	.778	.795	.807	.816	.821	.825	.827	.753
Cov.-Adj. Cos.	.527	.558	.580	.628	.670	.695	.711	.723	.731	.736	.740	.743	.670
Cov.-Adj. SDP	.697	.737	.764	.817	.856	.877	.889	.894	**.898**	**.900**	**.901**	**.901**	.844
SDP	**.705**	**.749**	**.777**	**.830**	**.866**	**.882**	**.890**	**.895**	.897	.898	.898	.898	**.849**
DINO ViT-S/16													
Cosine	.596	.628	.652	.697	.739	.766	.783	.796	.805	.811	.815	.818	.742
Cov.-Adj. Cos.	.553	.587	.613	.668	.714	.743	.760	.773	.781	.787	.791	.793	.714
Cov.-Adj. SDP	.688	.728	.756	.808	.850	.872	.884	**.891**	**.895**	**.897**	**.898**	**.899**	.839
SDP	**.701**	**.742**	**.770**	**.822**	**.859**	**.877**	**.886**	**.891**	.894	.895	.896	.896	**.844**
DINOv2 ViT-B/14													
Cosine	.606	.639	.663	.712	.752	.778	.796	.808	.816	.822	.826	.828	.754
Cov.-Adj. Cos.	.503	.530	.551	.596	.635	.660	.676	.688	.696	.702	.706	.709	.638
Cov.-Adj. SDP	.692	.732	.760	.812	.854	.876	.888	.895	**.899**	**.900**	**.901**	**.902**	.843
SDP	**.704**	**.749**	**.779**	**.832**	**.868**	**.885**	**.893**	**.897**	**.899**	**.900**	**.901**	.901	**.851**
DINOv2 ViT-S/14													
Cosine	.584	.615	.638	.686	.728	.756	.773	.786	.795	.801	.805	.808	.731
Cov.-Adj. Cos.	.513	.543	.565	.612	.653	.678	.694	.707	.717	.720	.725	.727	.654
Cov.-Adj. SDP	.688	.729	.756	.809	.852	.874	.886	**.893**	**.896**	**.898**	**.899**	**.900**	.840
SDP	**.704**	**.746**	**.774**	**.826**	**.862**	**.879**	**.888**	.892	.894	.895	.896	.896	**.846**

Table 9: Effect of the scalar similarity statistic across selected vision foundation models on TAPOS. We report micro-F1 across relative-distance thresholds. Bold marks the best statistic for each model at each threshold.

Statistic	.05	.06	.07	.10	.15	.20	.25	.30	.35	.40	.45	.50	Avg.
DINO ViT-B/16													
Cosine	.302	.334	.358	.416	.471	.518	.544	.566	.580	.591	.599	.605	.490
Cov.-Adj. Cos.	.210	.227	.247	.291	.357	.399	.437	.461	.480	.495	.502	.509	.385
Cov.-Adj. SDP	.384	.419	.449	.510	.565	.597	**.616**	**.630**	**.636**	**.643**	**.646**	**.649**	.562
SDP	**.420**	**.456**	**.479**	**.529**	**.575**	**.599**	.615	.625	.630	.632	.635	.637	**.569**
DINO ViT-S/16													
Cosine	.267	.297	.320	.376	.442	.489	.524	.547	.562	.577	.585	.591	.465
Cov.-Adj. Cos.	.266	.289	.307	.362	.418	.461	.497	.518	.533	.545	.552	.558	.442
Cov.-Adj. SDP	.386	.420	.452	.509	.561	.598	.619	.635	.644	.651	.655	.657	.566
SDP	**.406**	**.432**	**.459**	**.517**	**.573**	**.605**	**.626**	**.642**	**.650**	**.655**	**.659**	**.661**	**.574**
DINOv2 ViT-B/14													
Cosine	.311	.341	.363	.420	.481	.518	.547	.566	.581	.594	.600	.605	.494
Cov.-Adj. Cos.	.229	.252	.272	.324	.381	.417	.439	.464	.479	.489	.500	.506	.396
Cov.-Adj. SDP	.386	.417	.443	.507	.563	**.601**	**.625**	**.639**	**.650**	**.656**	**.660**	**.664**	**.568**
SDP	**.415**	**.444**	**.469**	**.526**	**.575**	.597	.609	.617	.623	.626	.629	.630	.563
DINOv2 ViT-S/14													
Cosine	.335	.368	.396	.442	.502	.535	.565	.583	.595	.607	.614	.620	.514
Cov.-Adj. Cos.	.295	.321	.341	.395	.463	.501	.532	.554	.567	.582	.589	.596	.478
Cov.-Adj. SDP	.380	.421	.447	.508	.569	.604	.624	.636	.646	.654	.658	.661	.567
SDP	**.389**	**.425**	**.456**	**.519**	**.588**	**.629**	**.648**	**.659**	**.669**	**.676**	**.679**	**.681**	**.585**

D.7 Detected change points and associated Reuters articles

The flat segments in the multivariate series arise from differences in the trading schedules of the four energy markets, including weekend closures and instrument-specific daily trading breaks, during which no new prices are recorded.

Table 11: Detected change points and corresponding Reuters articles.

Detected change-point	Reuters article
8 Feb 2026, 10:00 UTC	Iran insists on right to enrichment, ready for confidence-building
15 Feb 2026, 07:00 UTC	Iran says potential energy, mining and aircraft deals on table in talks with US
20 Feb 2026, 08:00 UTC	Trump warns Iran of 'bad things' if no deal made, sets deadline of 10-15 days
22 Feb 2026, 14:00 UTC	Tehran is ready for nuclear concessions if US meets demands, Iranian official says
1 Mar 2026, 11:00 UTC	Oil and gas prices surge as Iran war disrupts Middle Eastern output
4 Mar 2026, 03:00 UTC	Oil settles at highest in over a year for second straight day as Iran crisis escalates, snarls Hormuz flows
5 Mar 2026, 10:00 UTC	Oil settles up around 5% on supply concerns as Iran conflict widens
8 Mar 2026, 22:00 UTC	Iraqi oil production collapses with Strait of Hormuz blocked by conflict, sources say
12 Mar 2026, 17:00 UTC	Iran's new supreme leader vows to keep Hormuz shut, Netanyahu issues threat
15 Mar 2026, 00:00 UTC	Oil poised for further gains as Middle East conflict threatens export facilities
19 Mar 2026, 16:00 UTC	Iran war leaves deep, costly scar on Mideast energy
22 Mar 2026, 06:00 UTC	Iran threatens to retaliate against Gulf energy and water after Trump ultimatum
26 Mar 2026, 20:00 UTC	Trump pauses attacks on Iran's energy plants, says talks "going well"
29 Mar 2026, 16:00 UTC	Brent eyes record monthly rise; US crude settles above $100 as Houthis join Iran war
5 Apr 2026, 03:00 UTC	OPEC+ agrees to boost oil output when Strait of Hormuz reopens
9 Apr 2026, 10:00 UTC	Hormuz at near standstill as Iran warns ships to keep to its waters
12 Apr 2026, 11:00 UTC	ADNOC CEO says Strait of Hormuz not Iran's to control
16 Apr 2026, 22:00 UTC	Oil prices rise on doubts US-Iran peace talks will ease Hormuz disruption
19 Apr 2026, 11:00 UTC	Oil settles slightly higher as Iran plays down reported Israeli attack
23 Apr 2026, 14:00 UTC	Stocks, FX drop as Middle East tensions keep crude prices above $100

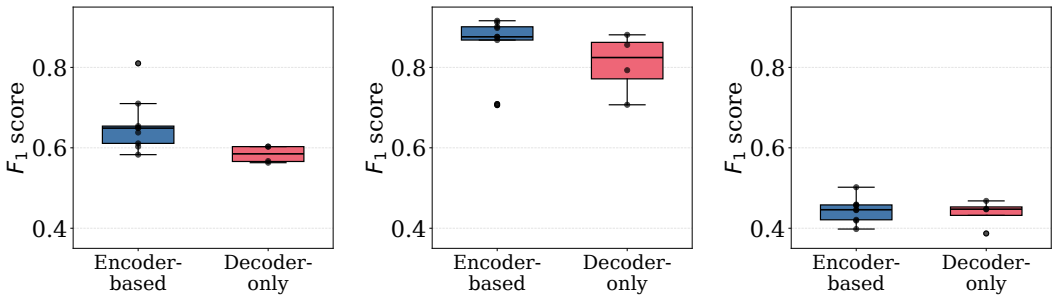


Figure 10: **Comparison of encoder-based and decoder-only time-series foundation models.** Distribution of F_1-scores across encoder-based and decoder-only architectures on (a) Yahoo A4, (b) USC-HAD, and (c) HASC. Each point corresponds to a foundation model.

Table 12: **Comparison of GAS scalarization statistics across modalities.** For time-series datasets, average ranks and model-wise wins are computed across 13 foundation models after averaging F_1 over the three evaluation tolerances. For video datasets, mean F_1 is computed from the *Avg.* column across four vision foundation models. Lower average rank is better.

		Average rank / Mean F_1				**Model-wise wins**			
Modality	**Dataset**	**SDP**	**Cos.**	**Cov.-SDP**	**Cov.-Cos.**	**SDP**	**Cos.**	**Cov.-SDP**	**Cov.-Cos.**
Time series	Yahoo A4	4.00	2.46	**1.54**	2.00	0	2	**7**	4
	USC-HAD	3.69	2.00	**1.85**	2.46	0	4	**7**	2
	HASC	**1.85**	3.00	2.38	2.77	**7**	1	4	1
Video	Kinetics	**0.848**	0.745	0.841	0.669	**4**	0	0	0
	TAPOS	**0.573**	0.491	0.566	0.425	**3**	0	1	0

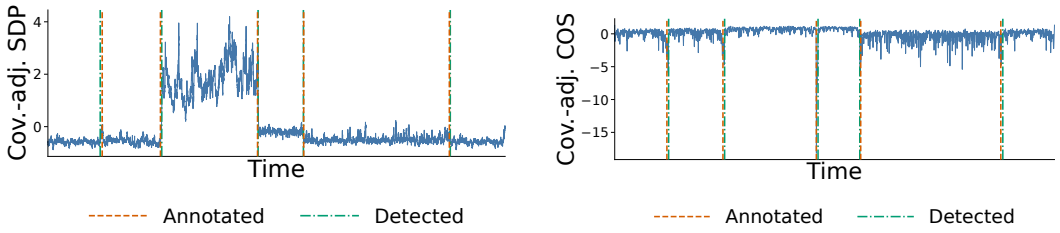
For time-series datasets, the first four columns report average rank (lower is better); for video datasets, they report mean F_1 (higher is better).

E Ablation Studies

E.1 Effect of the foundation model

Table 13: **Raw-input ablation.** F_1-scores obtained by applying the four scalarization statistics directly to the raw input, without foundation-model representations.

	Yahoo A4			**USC-HAD**			**HASC**		
Scalar.	24	50	75	100	200	400	60	100	200
SDP	.310	.339	.375	.754	.787	.787	.205	.205	.205
Cosine	.188	.217	.225	.760	.800	.800	.108	.108	.135
Cov.-SDP	.310	.339	.375	.814	.814	.814	.250	.250	.250
Cov.-Cos.	.188	.217	.225	.735	.776	.776	.086	.086	.114



(a) (b)

Figure 11: **Example change-point detection on USC-HAD. (a)** Covariance-adjusted scaled dot-product scalarization, where the projected series exhibits clear distributional changes around the annotated boundaries. **(b)** Covariance-adjusted cosine scalarization with corresponding SCAN detections.

E.2 Reproducibility

Data preprocessing - Energy data: To obtain hourly series from 1 February to 30 April 2026 (UTC), we aligned all markets to a common hourly grid and imputed missing data via forward-fill imputation.