Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Segmentation Improves Prediction Without Increasing Statistical Separability: Evidence from Industrial Time Series

emilio mastriani

emilio.mastriani@inaf.it

National Institute for Astrophysics

alessandro costa

National Institute for Astrophysics

federico incardona

National Institute for Astrophysics

kevin munari

National Institute for Astrophysics

sebastiano spinello

National Institute for Astrophysics

Research Article

Keywords: Change Point Detection, Predictive Maintenance, Statistical Separability, Sliced Wasserstein Distance, Industrial Time Series, Representation Learning

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10405191/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Segmentation Improves Prediction Without Increasing Statistical Separability: Evidence from Industrial Time Series

Emilio Mastriani[1*], Alessandro Costa[1], Federico Incardona[1], Kevin Munari[1], Sebastiano Spinello[1]

[1*]INAF, Osservatorio Astrofisico di Catania, Via S. Sofia 78, Catania, 95123, Italy.

*Corresponding author(s). E-mail(s): emilio.mastriani@inaf.it;
Contributing authors: alessandro.costa@inaf.it; federico.incardona@inaf.it;
kevin.munari@inaf.it; sebastiano.spinello@inaf.it;

Abstract

Change Point Detection (CPD) techniques are widely applied in predictive maintenance systems to improve performance in detecting anomalies within multivariate industrial time series. Previous studies on a steam turbine monitoring system demonstrated that segmentation based on the ChangeFinder algorithm, combined with heterogeneous ensemble models, substantially enhances classification accuracy. Despite these results, the underlying cause of this improvement remains unclear. A common interpretation is that segmentation increases statistical separability between healthy and faulty operational conditions, thereby facilitating the learning process of classification models.
In this study, we critically examine this hypothesis by employing a classifier-independent approach that relies on the Sliced Wasserstein Distance (SWD). We analyze and compare data representations derived from both raw process variables and segmentation-related features. Statistical robustness are ensured through stratified bootstrap resampling and repetitive distance estimations.
Our findings reveal that segmented representations consistently show reduced statistical separability compared to the original feature spaces. This trend persists even with the addition of extra clustering-derived descriptors. The results suggest that previously observed improvements in predictive performance are not due to an increase in statistical distance between classes. Rather, segmentation appears to assist learning by diminishing temporal ambiguity and coherently structuring the temporal evolution of process dynamics.

Keywords: Change Point Detection, Predictive Maintenance, Statistical Separability, Sliced Wasserstein Distance, Industrial Time Series, Representation Learning

1 Introduction

Predictive maintenance using machine learning has achieved notable success in detecting anomalies and faults early within complex industrial systems [1]. The effectiveness of such methodologies critically depends on the quality of data representations used for training predictive models [2]. Multivariate industrial time series pose specific challenges because of gradual operational transitions, regime changes, and degradation effects, which develop over extended periods [3].

These complex temporal dynamics cause feature overlap between healthy and faulty classes, leading to the introduction of temporal ambiguity

that complicates the training of supervised algorithms to discern optimal classification boundaries [4]. This uncertainty arises because observations with similar instantaneous process values may correspond to different operational states, preventing model learning.

To address these issues, temporal segmentation approaches based on Change Point Detection (CPD) have been proposed [5]. These techniques segment time series into intervals, bringing to light relatively homogeneous dynamics, effectively transforming continuous observations into sequences of operational states.

Prior work [6] applying ChangeFinder-based segmentation [7] on steam turbine monitoring data demonstrated a significant gain in predictive accuracy, with the Area Under the ROC Curve (AUC-ROC) improving from 0.8599 to 0.9760. This improvement highlighted segmentation as a key factor in enhancing classification efficacy. However, the mechanism driving this performance growth remains to be clarified.

One plausible interpretation is that segmentation increases the statistical separability of healthy and faulty classes in the feature space, making them more detectable during model training. This study is motivated by the need to investigate that assumption. Specifically, we examine whether segmentation benefits prediction through increased statistical distance between class distributions or via other structural effects related to temporal dynamics. Confirming the former would associate segmentation with enhancing statistical discrimination, whereas supporting the latter would challenge the common assumption in applied machine learning that better classification performance inevitably implies greater class separability. Therefore, we explicitly formalize the hypothesis and establish an analytical framework aimed at dissecting the structural role of segmentation.

1.1 Theoretical Rationale and Proposed Approach

The primary premise driving this study is that segmentation, as a feature transformation, reshapes the data representation space so that to increase the statistical distance between the distributions corresponding to healthy and faulty operational states [2]. Under this interpretation, integrating segmentation information should move data points associated with different classes into more distinct regions of the feature space, thus improving their discriminability [8–10]. In this work, we adopt the Sliced Wasserstein Distance (SWD), a member of the optimal transport distance family [11, 12] which offers several distinct advantages for this purpose. First, SWD does not require parametric assumptions about the shape or nature of the underlying distributions. Second, it is well-suited for high-dimensional and multivariate data, as it approximates the Wasserstein distance via integrals over projections onto one-dimensional subspaces [13, 14]. This projection-based formulation mitigates the curse of dimensionality common in direct high-dimensional probability comparisons. Our experimental framework compares three data representation modalities:

- The raw, original process variables, denoted X, acting as the baseline.
- The segmentation-derived variables S, representing only the segment identifiers extracted via ChangeFinder [7].
- The combined representation (X, S), which integrates both raw and segmentation features, mirroring the configuration used in prior predictive maintenance classifiers [6].

We also consider an enriched representation (X, S, C) that incorporates additional clustering-based descriptors derived in earlier research, although its use is secondary and intended to confirm the generality of our conclusions.

All SWD calculations employ a stratified bootstrap resampling scheme [15] to generate empirical distributions of the estimated separability metric, ensuring robustness and quantification of variability [16]. This approach provides confidence intervals and statistical significance assessments without depending on the stochastic nature of specific classifiers or training processes.

Our procedure thus enables a rigorous, model-agnostic investigation into whether segmentation explicitly increases the statistical separability of operational classes or facilitates learning through alternative mechanisms, such as dynamic contextualization or temporal restructuring [17].

2 Mathematical Formalization

This section formalizes the core concepts and notation used to analyze the impact of segmentation on the statistical separability of operational classes in industrial time series. Let $X \in \mathbb{R}^d$ denote the vector of continuous process variables observed at a particular time instant in the underlying dynamical system [2]. Each observation is associated with a binary class label $Y \in \{0, 1\}$, where $Y = 0$ corresponds to the healthy operating condition and $Y = 1$ denotes the faulty condition. Observations belonging to these two operational states induce corresponding empirical conditional probability distributions over the feature space:

$$P_0 = P(X \mid Y = 0) \tag{1}$$

and

$$P_1 = P(X \mid Y = 1) \tag{2}$$

These distributions capture the variability and statistical characteristics of data associated with each state, derived directly from observed measurements [9].

The ChangeFinder algorithm [7], based on online autoregressive modeling, generates a sequence of segment identifiers for each observation, which can be collected into a vector $S \in \mathbb{R}^m$. This vector encodes the segmentation-driven contextual state associated with each sample, capturing temporal dynamical information beyond instantaneous process values.

More generally, segmentation can be described by a transformation function:

$$\Phi : \mathbb{R}^d \to \mathbb{R}^{d+m} \tag{3}$$

which maps the original process variables X to an augmented representation $Z = (X, S)$ that combines raw measurements and segment-based descriptors. We introduce the notion of **dynamic contextualization** as follows:

Definition 1 (Dynamic Contextualization). A segmentation operator Φ is said to provide dynamic contextualization if the introduced descriptors S encode information about the recent dynamical evolution of the process, allowing observations with similar instantaneous measurements to be associated with different operational contexts.

To quantify the statistical separability between healthy and faulty classes in these representations, the study employs the **Sliced Wasserstein Distance (SWD)**. This metric leverages principles of optimal mass transport theory to measure the geometric difference between probability distributions, while circumventing the issue of dimensionality inherent in high-dimensional spaces [11, 12].

Formally, given a random unit vector $\theta \in \mathbb{S}^{d-1}$ sampled uniformly from the unit hypersphere, vector observations are projected onto θ to obtain one-dimensional distributions for each class [13]:

$$x_\theta = X\theta \tag{4}$$

The one-dimensional Wasserstein distance between these projected class distributions $P_{0,\theta}$ and $P_{1,\theta}$ is computed via sorted sample quantiles:

$$W_1(P_{0,\theta}, P_{1,\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left| x_{0,\theta}^{(i)} - x_{1,\theta}^{(i)} \right| \tag{5}$$

, where n denotes the number of samples per class Integrating over multiple random directions, the SWD is approximated as:

$$\mathrm{SWD}(P_0, P_1) = \frac{1}{K} \sum_{k=1}^{K} W_1(P_{0,\theta_k}, P_{1,\theta_k}) \tag{6}$$

, with K sufficiently large to sample the unit sphere [13, 14].

In this study, we estimate SWD for several representations:

- $SWD(X)$: the original process variables only,
- $SWD(S)$: segmentation identifiers only,
- $SWD(X, S)$: combination of raw variables and segmentation identifiers.

To evaluate the effect of segmentation, two differential quantities are defined:

$$\Delta_S = \mathrm{SWD}(S) - \mathrm{SWD}(X) \tag{7}$$

and

$$\Delta_{XS} = \mathrm{SWD}(X, S) - \mathrm{SWD}(X) \tag{8}$$

The initial hypothesis tested in this work assumes

$$\Delta_{XS} > 0 \tag{9}$$

, meaning that segmentation should increase the statistical separability between classes.

Alternatively, it is proposed that segmentation may not increase global separability but instead reorganizes the temporal structure of data, reducing ambiguity. The interpretive hypothesis formalizes the last concept.

Proposition 1 (Interpretive Hypothesis). An increase in predictive performance does not necessarily imply an increase in global statistical separability. A representation transformation Φ may improve classification performance even when

$$\Delta_{XS} \leq 0 \tag{10}$$

and transformation introduces dynamic contextualization capable of reducing the ambiguity associated with observations generated under different operating regimes.

This section set up the necessary mathematical formalism to conduct the subsequent empirical investigation exploring which of these scenarios holds in practice.

3 Data description

3.1 Experimental Setup and Dataset Characterization

The monitored physical system consists of an industrial-scale steam turbine coupled directly to an electrical generator in a fully digitalized refining plant environment. The turbine converts the pressure drop from high-pressure (HP) steam to medium-pressure (MP) steam into electrical energy, effectively serving as an energy recovery mechanism. During turbine downtime periods, steam is diverted through a dedicated pressure-reducing valve, ensuring operational continuity. The electrical output of the turbine is tightly coupled with the facility's overall utility steam demand and consequently reflects core production levels within the refining center [18].

The dataset comprises $d = 70$ continuous process variables (telemetry features) that capture essential operational measurements including pressures, temperatures, vibrations, and flow rates. Data encompasses a total of 1,124,820 observations, collected at a consistent sampling rate over a period of 16 months.

The historical data timeline is divided into two main intervals:

- Training horizon: July 9, 2022, to August 3, 2023
- Testing horizon: September 1, 2023, to November 22, 2024

Within the testing interval, a confirmed critical degradation event occurred between August 11 and August 17, 2024, representing a seven-day failure period corresponding to approximately 1.56% of the 448-day test duration.

Operational ground-truth labels were derived from an expert-validated Normal Operating Condition (NoC) log accompanying the telemetry matrix. This reference enables binary classification of each observation as either healthy (label 0) or faulty/anomalous (label 1), providing crucial supervision for subsequent learning algorithms [18].

4 Methodology

4.1 Experimental Design

The experimental workflow, presented in Figure 1, is organized into four consecutive phases: data preparation, representations construction, statistical separability estimation, and comparative analysis. This systematic approach ensures a robust and consistent evaluation of the influence of segmentation on separability [1].

4.2 Data Preparation

The analysis is based on two synchronized datasets characterizing the same industrial process. The first dataset, denoted X, contains the original process variables and establishes a fundamental operational baseline description of the monitored system. The second dataset, denoted S, is derived from applying ChangeFinder-based segmentation to the original variables. It combines both the original continuous telemetry data and the structural segment identifiers generated by the change-point detection routine.

Both datasets have been pre-cleaned, normalized, and feature-selected in previous studies, so no additional preprocessing was applied for this investigation. The operational target variable Y is

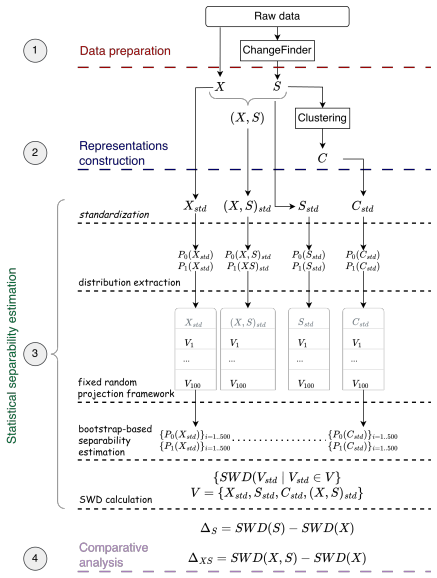


Fig. 1 Workflow of the proposed pipeline, showing data acquisition, temporal segmentation, semantic clustering, feature standardization, random projection for statistical separability estimation, bootstrap analysis, and result visualization.

directly extracted from the original Boolean field indicating normal operation. Samples corresponding to normal operating conditions are labeled as 0, while those associated with anomalies or faults are labeled as 1.

This labeling facilitates direct extraction of empirical distributions for systematic comparisons.

4.3 Representations Construction

A key element of the methodology is the comparison of multiple representations derived from the same underlying process. The baseline representation X comprises strictly the original process variables and serves as the reference for subsequent analyses.

The segmentation-only representation S contains exclusively the segment-level identifiers produced by ChangeFinder, which quantifies the discriminative information contained in the segmentation structure itself.

The combined representation (X, S) integrates the original variables and the segmentation identifiers, corresponding to the dynamics-aware dataset utilized in previous predictive maintenance experiments.

Additionally, an enriched configuration (X, S, C) is considered, which includes descriptors derived from segments and clustering-based features generated in prior research. Although not the primary focus, this representation serves to verify that conclusions remain valid when employing more complex, segmentation-related feature sets. All representations share the same class labels, allowing consistent and unbiased comparative analyses.

4.4 Statistical Separability Estimation

The separability estimation procedure comprises five main processes.

4.4.1 Standardization of Representations

Since the Sliced Wasserstein Distance (SWD) is sensitive to differences in feature scaling, all representations are standardized prior to analysis. Standardization parameters (mean and standard deviation) are estimated exclusively from the training data to prevent information leakage. The fitted scaling transformation is then applied consistently to both training and testing sets, preserving the relative structure of observations [19].

4.4.2 Extraction of Healthy and Faulty Distributions

Following standardization, each representation is partitioned according to the operational label. As already formalized in Section 2 , according to equation 1 and equation 2, for any generic representation R, two empirical distributions are derived ($P_0(R)$ and $P_1(R)$) to represent healthy and faulty operating conditions, respectively. These are the objects of direct comparison using SWD.

4.4.3 Fixed Random Projection Framework

The calculation of SWD relies on averaging Wasserstein distances of distributions projected onto random unit vectors. To avoid variability arising from generating different sets of projection directions in each bootstrap iteration, a fixed set

of 100 random unit projections per representation space is generated once at the beginning of the experiments and reused at all subsequent bootstrap steps. This approach ensures that variation in bootstrap replicates can be attributed exclusively to data resampling rather than randomness in the projection process.

4.4.4 Bootstrap-Based Separability Estimation

The dataset exhibits a severe imbalance between healthy and faulty samples, with the latter representing approximately 1.56% of the test data. Direct computation of distributional distances on the full dataset risks producing unstable estimates dominated by the majority class. To mitigate this inconvenient, we employ a stratified bootstrap procedure that independently samples with replacement an equal number of observations from each class, controlling sample size with an upper bound to maintain manageable computational costs [15, 20]. Balanced bootstrapped samples thus ensure reliable and fair estimation of separability metrics.

4.4.5 Calculation of the Sliced Wasserstein Distance

For each bootstrap replicate, the balanced samples are projected onto the fixed random directions, and Wasserstein-1 distances between the projected healthy and faulty distributions are computed. Averaging across all projections yields the SWD estimate for that replicate. Repeating this procedure across hundreds of bootstrap replicates generates a distribution of SWD values, capturing the statistical uncertainty inherent in the empirical estimations.

Figure 2 summarizes the computational workflow of the SWD estimation, illustrating the mapping from multivariate representations through projections to scalar distance metrics and their aggregation.

4.5 Comparative Statistical Analysis

After performing all bootstrap iterations, we calculate empirical distributions of SWD measures for each representation. Our analysis focuses on systematic comparison of baseline raw data $SWD(X)$, segmentation identifiers $SWD(S)$, combined dynamics-aware data $SWD(X, S)$, and, where applicable, the enriched representation $SWD(X, C)$, with C denoting clustering-based descriptors.

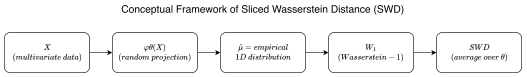


Fig. 2 Conceptual framework of the Sliced Wasserstein Distance (SWD): transformation from multivariate observations X through random projections $\varphi_\theta(X)$, empirical distributions, Wasserstein-1 distance, and final aggregation into SWD.

Primary attention is devoted to differential quantities 7 and 8 defined in Section 2, which measure the influence of segmentation information on class separability.

Representations involving clustering-derived descriptors C may be conditionally available, considering that such features require sufficient historical data and stable unsupervised training. Consequently, in ablation analyses to isolate segmentation effects, representations containing C are excluded.

4.6 Interpretive Framework

The interpretive phase forms the core of this study. Should segmentation genuinely enhance class separability, bootstrap distributions of Δ_{XS} would be predominantly positive, reflecting an increase in statistical distance when segmentation features are added.

Alternatively, if segmentation improves predictive capability without increasing statistical separability, distributions of Δ_{XS} would be near zero or negative, indicating that performance gains emerge from reorganization of temporal structure and reduction of ambiguity rather than from increased geometric separability.

This framework explicitly challenges the assumption that better classifier performance "must" coincide with greater statistical separability, directing focus toward temporal dynamics and representation structural effects as key factors in learning improvements.

Table 1 Mean results of the Sliced Wasserstein Distance (SWD) bootstrap analysis across different operational representations.

Representation Space	Mean SWD
X (raw baseline)	1.354
$X + S$ (segmented-baseline)	1.037
$X + S + C$ (segmented-enriched)	0.534

5 Results

Table 1 summarizes the mean results obtained from the bootstrap analysis of the Sliced Wasserstein Distance (SWD) computed across the different data representations of this study. Contrary to the initial hypothesis, the inclusion of segmentation-driven dynamic information does not increase the statistical separability between healthy and faulty operating conditions; conversely, a consistent reduction in the distance between the two class distributions is observed across all experiments. Specifically, in the first experiment, the dynamics-aware segmented representation (X, S) produces a mean SWD value of approximately 1.037 compared to the baseline raw feature space X, which has a mean SWD of 1.354. This corresponds to a decrease $\Delta SWD = -0.317$, representing a relative reduction of about 23.4%. Importantly, among the 500 stratified bootstrap replicates, not a single instance showed an increase in SWD when segmentation features were added, reinforcing the statistical robustness of the observed effect with an empirical probability $P(\Delta SWD > 0) = 0$.

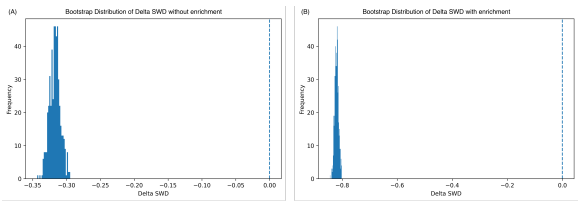


Fig. 3 Bootstrap distribution of Delta SWD over the dynamics-aware (A), and dynamics-aware-enriched (B) representations. The values have been obtained across 500 stratified bootstrap replications.

The bootstrap distribution of ΔSWD for the dynamics-aware representation, depicted in Figure 3A, is entirely situated below zero, indicating a systematic contraction rather than fluctuation-induced variability. The absence of positive values suggests that the reduction is an intrinsic property of the segmentation transformation applied to the data.

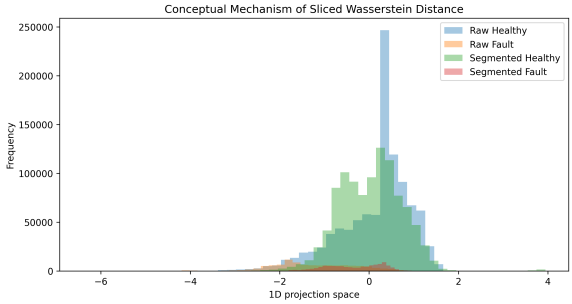


Fig. 4 Visual comparison of RAW vs SEGMENTED representations through independent random projections. The plot illustrates how class distributions differ in the induced one-dimensional projection spaces.

To gain better insight into the geometric underpinnings of this phenomenon, Figure 4 provides a visualization comparing the raw and segmented representations through independent random projections. These projections demonstrate that the reduction in SWD corresponds to a modified structure of the one-dimensional induced distributions, rather than a uniform contraction or simple scaling effect in the feature space. In a second experiment, we consider an enriched representation (X, S, C) that augments the dynamics-aware feature set with clustering-based descriptors derived from previous research efforts. Once again, the statistical separability decreases further to a mean SWD of 0.534, representing a substantial relative reduction of 60.6% compared to the baseline raw representation. The empirical bootstrap distribution illustrated in Figure 3B further corroborates this pronounced contraction and confirms the robustness of the finding across all replicates.

A comparison of the bootstrap SWD distributions for raw versus dynamics-aware and dynamics-aware-enriched representations is shown in Figure 5. This evaluation clearly reveals a progressive decrease in statistical separability as the representational complexity and dynamic awareness increase.

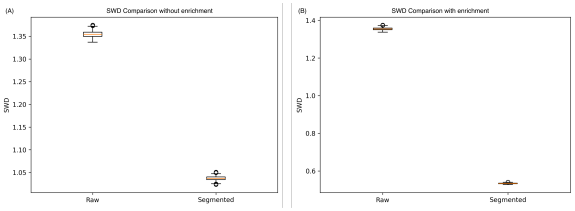


Fig. 5 Comparison between the bootstrap distributions of SWD computed on raw vs dynamics-aware (A), and raw vs dynamics-aware-enriched (B) representations.

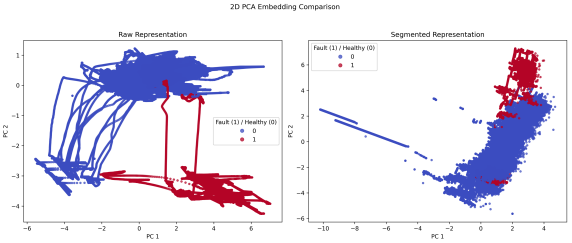


Fig. 6 2D PCA embedding comparison between raw and segmented representations. Segmentation induces a more locally structured organization of the samples in the embedding space.

Complementing the bootstrap analysis, a principal component analysis (PCA) based dimensionality reduction was conducted on both raw and segmented representations to visualize spatial organization within a two-dimensional embedding. Figure 6 illustrates that the segmented representation leads to a more structured spatial organization of samples characterized by locally coherent regions. Rather than increasing the global statistical distance between healthy and faulty observations, segmentation appears to reorganize the data points into operationally meaningful clusters, supporting classifiers through improved local coherence instead of global class separation. Overall, these results indicate that the performance advantage conferred by segmentation cannot be simply explained by an increase in global statistical separability. Alternatively, segmentation affords a reconfiguration of the observation space which facilitates model learning by reducing temporal ambiguity and imposing a clearer temporal structure, even in the presence of a lower measured statistical distance between classes. This interpretation aligns well with prior works wherein the Random Forest–XGBoost ensemble reached an AUC-ROC of 0.976 using segmented data, underscoring that improvements arise primarily from enhancements to the representation of operational dynamics rather than from increased geometric distancing of classes.

6 Discussions and Conclusions

The initial hypothesis driving this study assumed that ChangeFinder-based time series segmentation would lead to an increase in statistical separability between normal and faulty operating conditions. This proposition stemmed from previous empirical evidence where segmentation notably enhanced classification performance, elevating the Area Under the ROC Curve (AUC-ROC) from 0.8599 to 0.976. It was thus reasonable to expect that the underlying data representations after segmentation would demonstrate greater separation between class distributions.

To rigorously test this assumption, we employed the Sliced Wasserstein Distance (SWD) as a quantitative, classifier-independent metric for measuring the separability between multivariate empirical distributions. SWD estimations were conducted repeatedly using a stratified bootstrap procedure on both the original process variable space and augmented feature spaces incorporating segmentation.

Contrary to expectations, the results revealed a consistent reduction in SWD when segment identifiers were introduced. The mean SWD dropped from 1.354 in the original feature space to 1.037 in the dynamics-aware segmented representation, a relative decrease of approximately 23%. The bootstrap probability of observing an increase in the statistical distance was zero, underscoring the robustness of this decrease.

When further enriched representations including additional clustering-derived variables were considered, the separability metric decreased even more significantly, reaching about 0.534 — less than 40% of the original separability. These findings suggest that segmentation does not increase geometric separation between healthy and faulty states; rather, it reorganizes the data's temporal structure.

From a theoretical and practical standpoint, this leads to an important reinterpretation. The segmentation operator does not act merely by creating statistically distinct clusters of faults and

healthy samples, but expresses a form of dynamic contextualization. Specifically, segment identifiers provide supplementary information about the temporal context in which observations emerge, reducing the ambiguity inherent in time series where similar instantaneous measurements can belong to different operational conditions.

Additional analysis confirmed that many segments identified by ChangeFinder are very brief, occasionally spanning only single observations. This high temporal resolution encodes dynamic transitions rather than stable regimes, providing classifiers with richer temporal cues to differentiate scenarios that are otherwise indistinguishable on static feature values alone. The evidence aligns with previous investigations indicating that attempts to incorporate more complex or clustering-derived features beyond simple segmentation identifiers do not necessarily enhance classification [5]. Instead, the principal benefit arises from restructuring the learning problem to better capture temporal dependencies and reduce uncertainty.

This insight challenges the somewhat intuitive but overly simplistic assumption that improved predictive performance must coincide with increased class separability. Instead, it highlights that segmentation primarily reduces the effective complexity of the classification task, facilitating the construction of accurate decision boundaries by supervised models.

In conclusion, the effectiveness of segmentation in predictive maintenance frameworks emerges not from increasing statistical distance between classes but from delivering a more informative, temporally coherent data representation. Future research should build on this understanding by designing dynamic-aware descriptors that explicitly characterize segment persistence, evolution, and stability to further enhance model learning.

Furthermore, the generalization of these findings should be investigated through application to diverse datasets and exploration of alternative change point detection methods beyond ChangeFinder. Emphasizing temporal and dynamic representation over geometric separability may open new perspectives in predictive maintenance and time series learning.

Acknowledgements. This work is supported by ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by the European Union – NextGenerationEU. The authors gratefully acknowledge Alfonso Amendola and Emilio Villa for their technical and scientific assistance, and Carlo Acutis for his human support, which has been a source of encouragement during this study.

Funding. *No funding was received for conducting this study.*

References

[1] Malek M, Salim N, Abdallah C, Ghandour J, Al-Saleh Y. An integrated machine learning: Utility theory framework for real-time predictive maintenance in pumping systems. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science. 2021;235(7):1368–1384. https://doi.org/10.1177/0954405420970517.

[2] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013;35(8):1798–1828.

[3] Wang J, Zhao M, Wang Y, Sun Q, Chen H. Multivariate degradation modeling with time-dependent covariates and imperfect maintenance effects. Advanced Modeling and Simulation in Engineering Sciences. 2021;8:20. https://doi.org/10.1002/asmb.2600.

[4] Wei Y, Yin Y, Ma Z, Huang H, Chen G, Wang Z. Temporal dynamic label smoothing against horizon-induced distribution overlap: Improving early fault detection and diagnosis in chemical process. Chemical Engineering Science. 2025;282:116921. https://doi.org/10.1016/j.ces.2025.116921.

[5] Basseville M, Nikiforov IV. Detection of Abrupt Changes: Theory and Application. Englewood Cliffs, NJ: Prentice Hall; 1993.

[6] Anonymous. Omitted for blind review. Under Review / Archive. 2026;Omitted for blind review. Previous work on steam turbine

monitoring systems raising AUC-ROC from 0.8599 to 0.9760.

[7] Yamanishi K, Takeuchi Ji. A Unifying Framework for Detecting Outliers and Change Points from Time Series. ACM, New York, NY, USA: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2002.

[8] Acevedo A, Durán C, Kuo MJ, Ciucci S, Schroeder M, Cannistraci CV. Measuring Group Separability in Geometrical Space for Evaluation of Pattern Recognition and Dimension Reduction Algorithms. IEEE Access. 2022;10:22441–22471. https://doi.org/10.1109/ACCESS.2022.3152789.

[9] Sriperumbudur BK, Fukumizu K, Gretton A, Schölkopf B, Lanckriet GR. On the empirical estimation of integral probability metrics. Electronic Journal of Statistics. 2012;6:1550–1599.

[10] Fukumizu K, Bach FR, Jordan MI. Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. Journal of Machine Learning Research. 2004;5:73–99.

[11] Villani C. Optimal Transport: Old and New. Berlin, Heidelberg: Springer; 2009.

[12] Peyré G, Cuturi M. Computational optimal transport: With applications to data science. Foundations and Trends® in Machine Learning. 2019;11(5-6):355–607.

[13] Bonnotte N. From the Brenier formula to the Wasserstein distance on the-space. SIAM Journal on Mathematical Analysis. 2013;45(1):317–333.

[14] Bonneel N, Rabin J, Peyré G, Pfister H. Sliced and Radon Wasserstein distances for rectangular domains. SIAM Journal on Imaging Sciences. 2015;8(1):83–111.

[15] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. Boca Raton, FL: CRC Press; 1994.

[16] He H, Garcia EA. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering. 2009;21(9):1263–1284.

[17] Vapnik V. The Nature of Statistical Learning Theory. New York, NY: Springer; 1999.

[18] Anonymous.: Omitted for blind review. Omitted for blind review. Proceedings of the International Conference on Control, Decision and Information Processing (CoDIT 2025).

[19] Bishop CM. Pattern Recognition and Machine Learning. New York: Springer; 2006.

[20] DiCiccio TJ, Efron B. Bootstrap confidence intervals. Statistical Science. 1996;11(3):189–228.