POLITECNICO MILANO 1863

Executive Summary of the Thesis

Mitigating Environmental and Operational Variability in Bridge SHM via Regime Partitioning and Adaptive Mahalanobis Novelty Scoring

Civil Engineering for Risk Mitigation

Author: SeyedSoheil Kashfi

Advisor: Prof. Alberto Corigliano

Co-advisor: Prof. Alireza Entezami

Academic year: 2024-2025

Keywords: Structural Health Monitoring (SHM); Environmental and Operational Variability (EOV) Mitigation; Unsupervised Learning; Spectral Clustering (SC); Regime Partitioning; Anomaly Detection; kNN; Mahalanobis Squared Distance (MSD); Z24 Bridge; KW51 Bridge.

1. Introduction

Bridge Structural Health Monitoring (SHM) aims at detecting abnormal structural conditions early enough to support safety and maintenance decisions. A central obstacle in real-world deployments is Environmental and Operational Variability (EOV): changes in temperature, humidity, traffic, and operational settings can induce variations in measured dynamic features that may resemble damage effects. This often leads to either excessive false alarms or missed detections when EOV masks damage signatures. In addition, labeled damage data are usually unavailable, motivating *unsupervised* and reproducible workflows for baseline modeling and novelty detection [1]. This thesis develops and validates a regime-aware novelty detection framework that combines **Tuned Spectral Clustering (TSC)** for baseline partitioning with a **Conditioning-guided Adaptive Mahalanobis Squared Distance (CAMSD)** for robust anomaly scoring. The framework is demonstrated on two benchmark bridge datasets: KW51 (real EOV, pre-intervention baseline) and Z24 (long-term monitoring with progressive damage scenarios) [3].

2. Datasets and experimental set-up

The monitored features considered in this thesis are modal eigenfrequencies for the selected modes. Data are split into training and subsequent evaluation segments following the protocol implemented in the Algorithm 1 for KW51 and Z24 bridges.

2.1. Case Study 1: KW51

The proposed methodology is validated on the KW51 steel bowstring railway bridge, located on the Belgian railway line L36N between Leuven and Brussels, crossing the Leuven–Dijle canal (115 m span, 12.4 m width). Since 2 October 2018, the bridge has been continuously monitored by a permanent SHM system combining vibration and environmental sensors, providing acceleration records and associated environmental measurements. This study focuses on the period before retrofitting from 2 October 2018 to 15 May 2019 and uses only the pre-intervention segment to build and assess a robust baseline. Modal frequencies extracted via an OMA procedure are considered, selecting vertical modes 6, 10, 12, and 13 due to limited missing data, resulting in 2,688 complete samples [2]. The frequency time series exhibits pronounced variability and discontinuities, only weakly explained by air temperature, indicating the presence of multiple EOV sources and motivating the need for normalization and unsupervised learning for structural assessment.

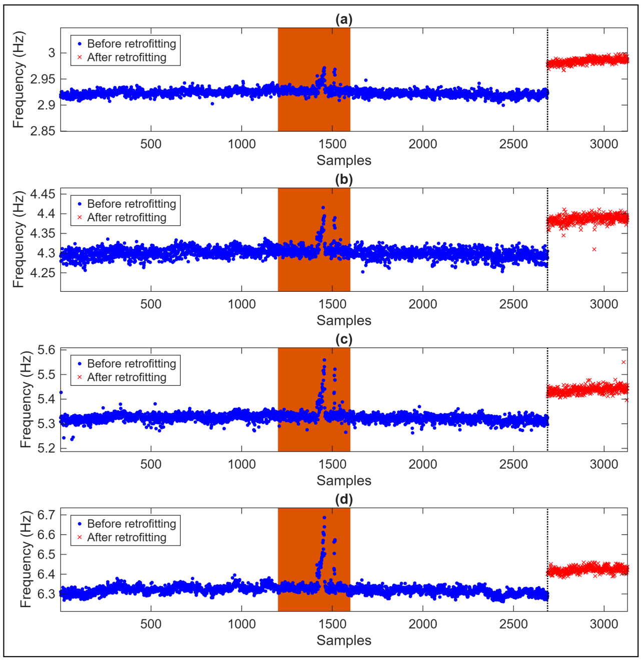

Figure 1: Eigenfrequencies of the KW51 for the selected modes.

2.2. Case Study 2: Z24

The second case study considers the Z24 bridge, a widely used benchmark in vibration-based structural health monitoring. The structure was a simply supported, post-tensioned concrete two-cell box-girder bridge located in the Canton of Bern (Switzerland), carrying a road over the River Emme near Koppigen. Prior to its demolition in the late 1990s, the bridge was subjected to an intensive monitoring campaign combined with a controlled, stepwise introduction of artificial damage scenarios, producing a public dataset that captures both long-term environmental effects and progressive deterioration mechanisms [1, 3]. Although the structural system is statically determinate (hinges and elastomeric bearings on abutments and piers), its dynamic response is influenced by interaction among the girder, supports, and foundation soil, as well as by traffic loading and thermal actions. In particular, temperature gradients and freezing conditions introduce strong variability in the identified eigenfrequencies: during low-temperature periods (air temperature below $0\,^{\circ}$C), the nominally undamaged modal frequencies can exhibit abrupt fluctuations, which may distort baseline learning and lead to false indications of damage if not handled explicitly. In the Z24 benchmark dataset, each sample corresponds to one measurement instance of the monitoring campaign for which modal frequencies were identified and released, enabling systematic evaluation of methods for EOV mitigation and damage/novelty detection [3].

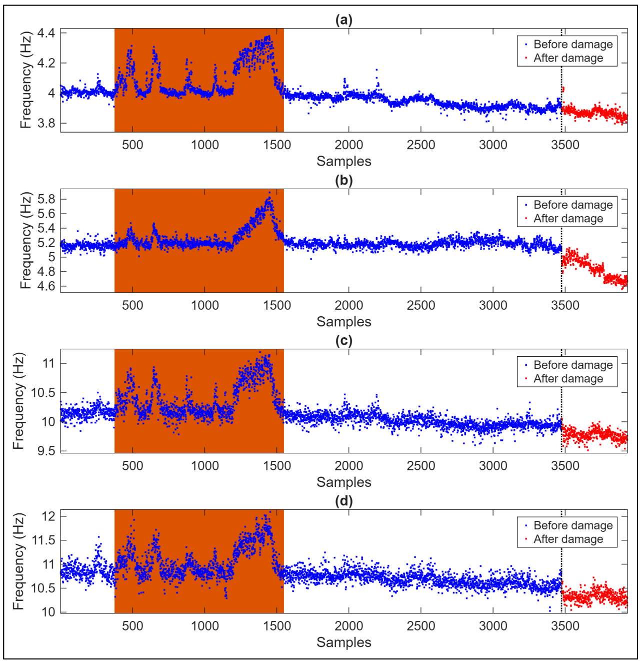

Figure 2: Eigenfrequencies of the Z24 for the selected modes.

2.3. Compact protocol summary for KW51 bridge

The KW51 script sets (i) total number of samples $n_{\mathrm{Fr}} = 3129$, (ii) number of monitored modes $n_{\mathrm{Mode}} = 4$, (iii) training length $n_{\mathrm{Tr}} = \lfloor 0.75\, n_{\mathrm{He}} \rfloor = 2016$, (iv) healthy segment length $n_{\mathrm{He}} = 2688$, and (v) a target false-positive rate of $\alpha = 0.5\%$ for threshold calibration on undamaged data.

Table 1: KW51 split and target FPR.

Quantity	Value
Total samples n_{Fr}	3129
Modes n_{Mode}	4
Training samples n_{Tr}	2016
Healthy samples n_{He}	2688
Validation $n_{\mathrm{Vd}} = n_{\mathrm{He}} - n_{\mathrm{Tr}}$	672
Testing samples $n_{\mathrm{Te}} = n_{\mathrm{Fr}} - n_{\mathrm{Tr}}$	1113
Target FPR α	0.5%

2.4. Compact protocol summary for Z24 bridge

The Z24 script sets (i) total number of samples $n_{\mathrm{Fr}} = 3932$, (ii) number of monitored modes $n_{\mathrm{Mode}} = 4$, (iii) training length $n_{\mathrm{Tr}} = \lfloor 0.8\, n_{\mathrm{Fr}} \rfloor = 3145$, (iv) healthy segment length $n_{\mathrm{He}} = 3475$, and (v) a target false-positive rate of $\alpha = 0.5\%$ for threshold calibration on undamaged data.

Table 2: Z24 split and target FPR.

Quantity	Value
Total samples n_{Fr}	3932
Modes n_{Mode}	4
Training samples n_{Tr}	3145
Healthy samples n_{He}	3475
Validation (healthy tail) $n_{\mathrm{Vd}} = n_{\mathrm{He}} - n_{\mathrm{Tr}}$	330
Testing samples $n_{\mathrm{Te}} = n_{\mathrm{Fr}} - n_{\mathrm{Tr}}$	787
Target FPR α	0.5%

3. Proposed methodology: TSC–CAMSD

The proposed methodology consists of two stages: (I) regime discovery via tuned spectral clustering on training data and (II) novelty scoring via CAMSD within a selected dataset. The intent is to reduce EOV-driven variability by conditioning novelty scoring on more homogeneous training data, while ensuring numerical robustness of local covariance estimation [4].

3.1. Stage I: TSC

Given standardized dataset features $\mathbf{x}_i \in \mathbb{R}^p$, a similarity matrix is built via an RBF kernel,

$$S_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \qquad S_{ii} = 0, \tag{1}$$

and spectral clustering is performed for candidate (k, σ) pairs. In the KW51 and Z24 implementations, the search ranges are $k \in \{2, 3, 4, 5, 6\}$ and $\sigma = \sigma_0\, [0.25, 0.5, 1, 2, 4]$ with $\sigma_0 = \mathrm{median}(\|\mathbf{x}_i - \mathbf{x}_j\|)$ computed from non-zero pairwise distances. The best setting is selected by maximizing mean silhouette score computed on the spectral embedding.

3.2. Stage II: CAMSD

Let $\mathcal{N}_k(\mathbf{x})$ be the k-nearest neighbors of test point $\mathbf{x}$ within the reference baseline set (or a chosen regime subset). CAMSD uses a *locally estimated* covariance structure, but selects k adaptively in a user-defined interval $[k_{\min}, k_{\max}]$ by scanning increasing k and accepting the first value yielding a sufficiently well-conditioned triangular factor R (based on $\mathrm{rcond}(R)$). To stabilize ill-conditioned neighborhoods, CAMSD applies an automatic ridge regularization.

For a chosen neighborhood size k with centered neighbor matrix $C \in \mathbb{R}^{k \times p}$ and local mean $\boldsymbol{\mu}$, CAMSD computes an upper-triangular factor R and defines the squared score

$$d^2(\mathbf{x}) = (k-1)\, \|R^{-T}(\mathbf{x} - \boldsymbol{\mu})\|^2. \tag{2}$$

3.3. Threshold calibration and decision rule

Given novelty scores $d(\mathbf{x}_i)$ on undamaged calibration data, the alarm threshold τ is obtained by an empirical quantile targeting FPR α,

$$\tau = Q_{1-\alpha}\big(\{d(\mathbf{x}_i)\}_{\mathrm{healthy}}\big), \tag{3}$$

and an alarm is issued when $d(\mathbf{x}) > \tau$.

3.4. Workflow Summary

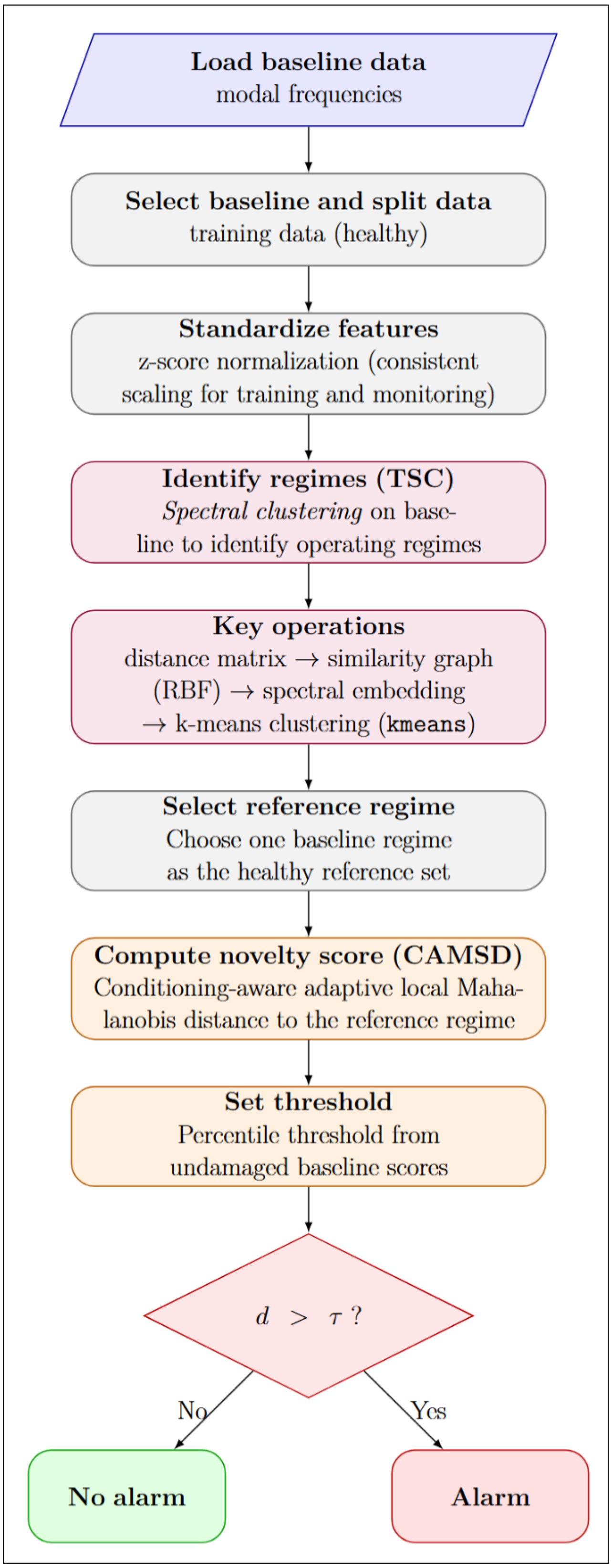

Figure 3: Simplified TSC–CAMSD workflow.

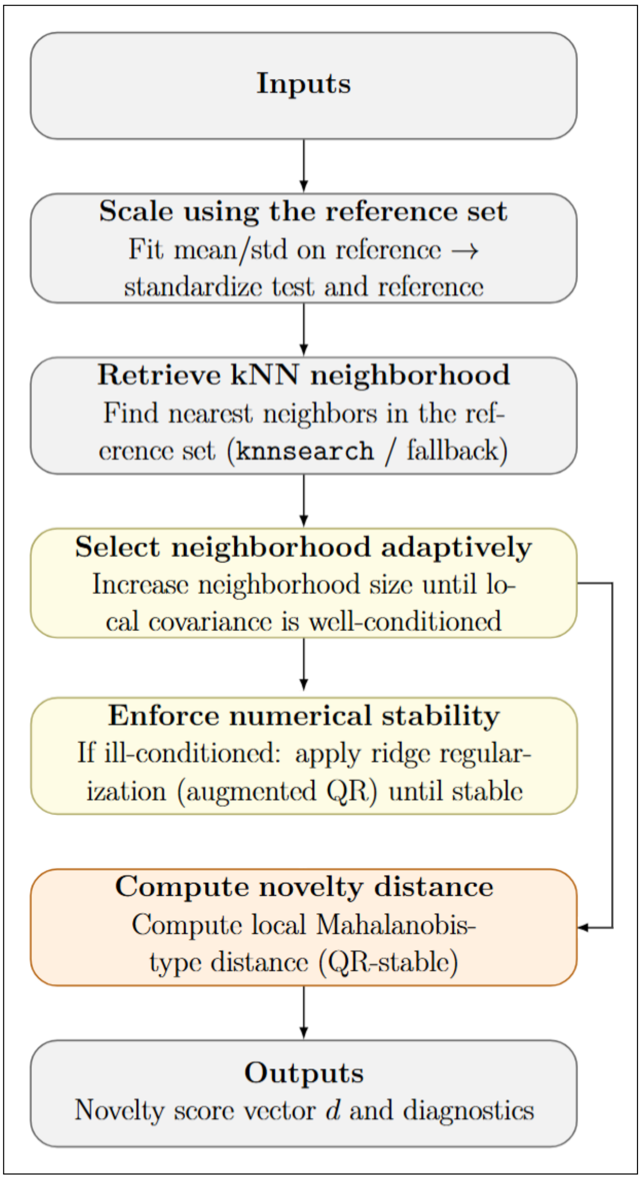

Figure 4: Simplified CAMSD internals.

Algorithm 1 TSC–CAMSD

Input: baseline features $X \in \mathbb{R}^{p \times n_{\mathrm{Tr}}}$; monitoring features $Z \in \mathbb{R}^{p \times n_{\mathrm{Te}}}$ (first n_{Vd} healthy); target FPR α (%).
Output: novelty scores (d_1, d_2, DI), threshold τ, alarms.
(I) TSC on baseline (regime discovery)
Standardize $X^\top$ and compute pairwise distances D^2; set $\sigma_0 \leftarrow \mathrm{median}(\sqrt{D^2_{ij}})$.
Grid-search (k, σ); build affinity $S = \exp(-D^2/(2\sigma^2))$ and run spectral clustering on S.
Select $(k^\star, \sigma^\star)$ by maximizing mean silhouette on the embedding; obtain labels `idx`.
Form clusters $\{$`Cluster`$\{r\}\}$ and set regime reference `refPts` $\leftarrow$ `Cluster`$\{1\}^\top$.
(II) CAMSD training (baseline self-scoring)
Compute $d_1 \leftarrow$ `CAMSD`$(X^\top, X^\top,$ `opts`$)$ (with conditioning-adaptive k and regularized QR).
Set $\tau_{\mathrm{Tr}} \leftarrow Q_{1-\alpha/100}(d_1)$.
(III) CAMSD monitoring (validation-calibrated threshold)
Compute $d_2 \leftarrow$ `CAMSD`$(Z^\top,$ `refPts`, `opts`$)$.
Calibrate $\tau \leftarrow Q_{1-\alpha/100}\big(d_2(1{:}n_{\mathrm{Vd}})\big)$ (healthy-only).
(IV) Fusion and decision
$DI \leftarrow [d_1; d_2]$; $\quad \mathrm{alarm}(t) \leftarrow \mathbb{1}\{DI(t) > \tau\}$.

Algorithm 2 CAMSD

Input: `testPts` $\in \mathbb{R}^{m \times p}$, `refPts` $\in \mathbb{R}^{n \times p}$, options `opts` (incl. $k_{\min}, k_{\max}$, z-score, `minRcond`, auto-ridge).
Output: scores `d2` $\in \mathbb{R}^m$ and diagnostics (`usedK`, `rcondR`, λ).
(1) Standardize fit (μ, σ) on `refPts`; compute scaled `refZ`, `testZ`.
(2) Retrieve candidates for each test point, find $k_{\max}$ nearest neighbors in `refZ` (e.g., `knnsearch`).
(3) Select local k (conditioning-guided) increase k from $k_{\min}$ until $\mathrm{rcond}(R) \geq$ `minRcond` for the (regularized) QR factor R of the centered neighborhood; store `usedK` and λ.
(4) Score with neighbors of size $\hat{k}$:

$$\delta = (\texttt{testZ} - \mu_{\mathrm{loc}})^\top, \quad u = R^\top \backslash \delta, \quad \texttt{d2} = (\hat{k}-1)\, u^\top u.$$

Return `d2` and diagnostics.

4. Main results and discussion

Across the two case studies, the results consistently support the following conclusions:

- **Regime awareness improves robustness to EOV.** TSC reduces baseline heterogeneity induced by environmental and operational variability, producing reference sets that better represent local operating conditions.
- **Conditioning-guided local scoring improves numerical stability.** CAMSD avoids fragile covariance inversions by (i) scanning k and (ii) enforcing a minimum conditioning via $\mathrm{rcond}(R)$, with automatic ridge regularization when needed.
- **FPR-controlled calibration improves interpretability.** Quantile thresholding provides a direct operational control knob for false alarms (see equation (3)).

Overall, these mechanisms translate into lower false-alarm rates under benign variability and a clearer separation between undamaged and damaged regimes, yielding more reliable novelty detection than the classic MSD approach [3, 4]. At the target $\alpha = 0.5\%$, TSC–CAMSD achieves FPR $= 0.00\%$ and FNR $= 0.00\%$ on KW51, and on Z24 it maintains FPR $= 0.49\%$ while reducing FNR to 0.88%. Note that in figures 5 and 6, the novelty index is denoted by $\delta := d^2$ (CAMSD score).

4.1. Anomaly Detection Results for the KW51 Bridge

Figure 5 shows that the TSC–CAMSD novelty index δ stays below the threshold τ throughout the training and undamaged validation segments (no false alarms), then jumps by orders of magnitude at the start of the structural intervention event and remains above τ, indicating near-immediate and clearly separated intervention detection; the vertical dashed lines mark the three data partitions, and τ is set as a training-score quantile to meet the target false-positive rate α.

4.2. Anomaly Detection Results for the Z24 Bridge

Figure 6 illustrates the Z24 novelty index staying mostly below the threshold during training and undamaged validation (with a few false alarms), then rising sharply and remaining well above the threshold in the damaged test segment, with only a small number of missed detections near the onset; the vertical dashed line separates training from validation, and the threshold τ is set as a training score quantile to meet target FPR α.

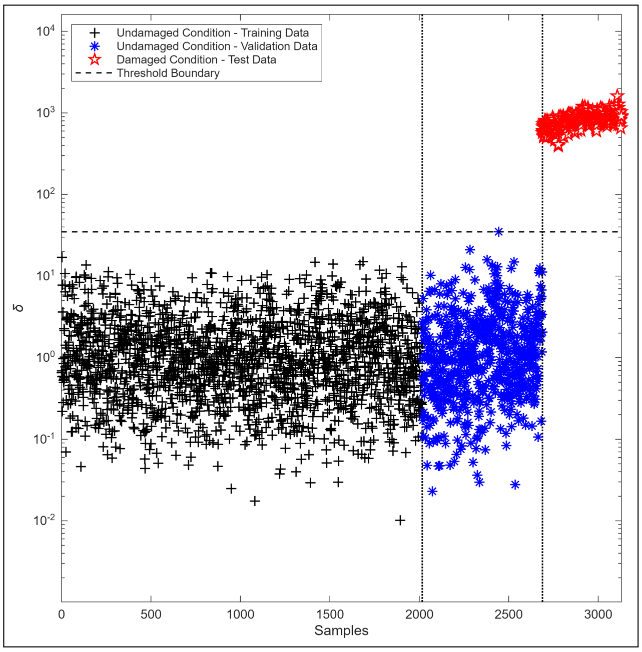

Figure 5: TSC–CAMSD novelty index for KW51.

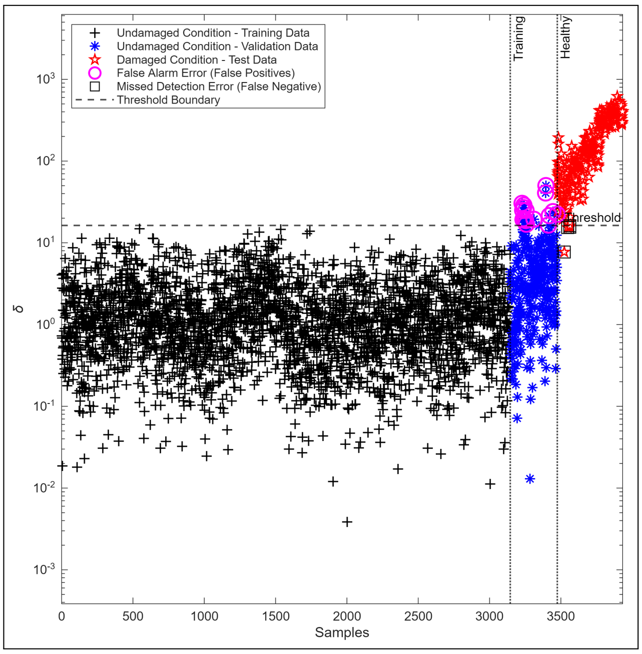

Figure 6: TSC–CAMSD novelty index for Z24.

5. Quantitative Performance Summary

Beyond qualitative plots, performance is quantified by the false-positive rate (FPR) on undamaged data and the false-negative rate (FNR) on damaged/post-event data when labels exist: for KW51 the event aligns with the documented intervention period, while for Z24 the damage timing is known, and Tables 3 and 4 summarize these results.

Table 3: Quantitative summary of novelty-detection performance for KW51.

Method	**FPR (%)**	**FNR (%)**
Raw data *(no clustering)*	0.49	0.00
MSD *(no clustering)*	0.48	99.55
CAMSD *(no clustering)*	0.48	0.00
TSC–CAMSD	0.00	0.00

Table 4: Quantitative summary of novelty-detection performance for Z24.

Method	**FPR (%)**	**FNR (%)**
Raw data *(no clustering)*	0.49	77.24
MSD *(no clustering)*	0.49	98.69
CAMSD *(no clustering)*	0.49	1.31
TSC–CAMSD	0.49	0.88

6. Conclusions

This thesis addresses a key practical bottleneck of bridge SHM: separating damage signatures from EOV-driven changes without relying on labeled damage data. The proposed TSC–CAMSD framework combines (i) tuned spectral clustering to build regime-consistent baseline subsets and (ii) a conditioning-guided adaptive Mahalanobis novelty score to provide stable local anomaly scoring and interpretable, FPR-controlled thresholds. Validation on KW51 and Z24 indicates that regime-aware modeling and numerically robust local scoring are complementary and effective for real monitoring conditions.

References

[1] S. W. Doebling, C. R. Farrar, M. B. Prime, and D. W. Shevitz. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review. Technical report, 1996.

[2] K. Maes and G. Lombaert. Monitoring data for railway bridge kw51 in leuven, belgium, before, during, and after retrofitting (version 1.0), 2020.

[3] B. Peeters and G. De Roeck. One-year monitoring of the z24-bridge: Environmental effects versus damage events. *Earthquake Engineering & Structural Dynamics*, 30(2):149–171, 2001.

[4] H. Sarmadi and A. Karamodin. A novel anomaly detection method based on adaptive mahalanobis-squared distance and one-class knn rule for structural health monitoring under environmental effects. *Mechanical Systems and Signal Processing*, 140:106495, 2020.