Computational Engineering and Technology Innovations

www.ceti.reapress.com

Comput. Eng. Technol. Innov. Vol. 3, No. 4 (2026) 308–322.

Paper Type: Original Article

Calibrated Early Warning of Simulated Bioprocess Transitions: Comparing Scalar, Sequential, and Delay-Embedding Features

Anna L. Marchetti[1,*], Bo Zhang[2], Jan H. Novak[3], Amal Cherif[4]

[1] Department of Computer Science, University of Pisa, Pisa, Italy; a.marchetti@unipi.it.
[2] Department of Computer Science and Technology, Nanjing University, Nanjing, China.
[3] Faculty of Information Technology, Czech Technical University in Prague, Praha, Czech Republic.
[4] Laboratoire d'Informatique et de Modelisation, Ecole Nationale d'Ingenieurs de Tunis, Universite de Tunis El Manar, Tunis, Tunisia.

Citation:

Received: 12 August 2025 *Revised: 24 October 2025* *Accepted: 29 February 2026*	Marchetti, A. L., Zhang, B., Novak, J. H., & Cherif, A. (2026). Calibrated early warning of simulated bioprocess transitions: Comparing scalar, sequential, and delay-embedding features. *Computational Engineering and Technology Innovations, 3*(4), 308-322.

Abstract

Early warning of bioprocess instability is an episode-level decision problem: a detector must signal before a transition while limiting nuisance alarms during stable operation. This simulation study compares classical scalar indicators, a sequential detector, logistic-score fusion, and a delay-embedding graph descriptor in two mechanistically distinct settings—a delayed chemostat approaching an oscillatory transition and a harvesting model approaching a fold. For each setting, 20 nominal-noise trajectories were divided into 12 training and eight test realizations, and five additional trajectories with doubled process noise were reserved for frozen-parameter stress testing. Biomass was evaluated in causal 100-sample windows advancing by five samples. Residual variance, lag-one autoregression, return rate, normalized permutation entropy, and an embedding-derived spanning-tree statistic were compared using window-level discrimination and episode-level detection, lead, and false-alarm rate. On the nominal chemostat test set, residual variance and a training-quantile logistic rule each detected 8/8 transitions with a median lead of 56 strides and reported false-alarm rates of 0.84 and 0.96 per 1,000 negative windows, respectively; the multivariate fusion therefore provided no operational gain over variance. In the harvesting model, feature AUCs remained close to 0.5 and qualifying alarms were sparse, so the tested biomass features did not establish reliable anticipation of the fold. Under doubled noise, ranking changed modestly but false alarms rose by two orders of magnitude, exposing a severe threshold-transfer failure. The analysis also distinguishes window ranking from usable warning, formalizes causal timing and persistence semantics, and shows why a maximum-spanning-tree quantity cannot be interpreted as degree-zero Vietoris–Rips persistence without recomputation. The results support a transition-specific and calibration-centered view: simple indicators can be competitive when their mechanism is aligned with the transition, whereas fusion, topology, and high AUC are not substitutes for held-out episode evaluation and noise-robust operating points.

Keywords: Bioprocess monitoring, Critical transition, Early warning, Delayed chemostat, Fold bifurcation, Alarm calibration, Delay embedding, Persistent homology, Noise robustness.

1 | Introduction

Continuous bioprocesses are expected to remain productive within a narrow operating envelope despite changes in feed composition, cell physiology, sensor quality, and feedback action. Once oscillation, washout,

Corresponding Author: a.marchetti@unipi.it
https://doi.org/10.48314/ceti.v3i4.98
Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

or a state jump becomes visible in routine alarms, the opportunity for a low-cost intervention may already have narrowed. The operationally important question is therefore not simply whether a trajectory is abnormal, but whether an impending loss of resilience can be recognized early enough to support action and with few enough nuisance alarms to remain credible. This distinction separates early warning from retrospective fault detection and makes the temporal definition of an alarm as important as the indicator from which it is constructed.

A large part of early-warning theory is built around critical slowing down. Near certain local bifurcations, the dominant recovery rate approaches zero, perturbations decay more slowly, and observable time series may show rising variance, stronger lag-one autocorrelation, or altered spectral composition [1–3]. These signatures are attractive because they can be computed from a single channel without a fully identified mechanistic model. Yet their generic appearance does not imply generic performance. Finite samples, nonstationary forcing, colored noise, measurement filtering, and transition type can amplify, obscure, or even reverse an apparent trend. Methodological studies consequently emphasize that early-warning skill must be quantified against explicit nulls and operating points rather than inferred from a rising indicator alone [4–8].

The same caution is particularly relevant in bioprocess monitoring. Industrial monitoring has a long tradition of multivariate statistical process control, process-history-based diagnosis, state estimation, and soft sensing [9–15]. These methods are often evaluated by fault classification, estimation error, or post-change detection delay. A pre-transition warning problem is different: onset must be defined independently of the detector; all preprocessing and thresholds must be learned without access to the test trajectory; and a sequence of correlated window scores must be converted into alarm episodes. A method can rank pre-onset windows well and still fail to cross a conservative threshold before onset, or can achieve a long nominal lead only because it alarms repeatedly during stable operation.

Nonlinear and geometric features offer a complementary description of process telemetry. Permutation entropy summarizes ordinal-pattern diversity and is comparatively robust to monotonic transformations [16]. Delay-coordinate embeddings reconstruct local trajectory geometry from a scalar time series, while persistent homology can quantify connected components and loops across scale [17–19]. Such tools have revealed periodic structure in biological time series and transition-related geometry in financial systems [18], [20], [21]. Their promise in bioprocess surveillance is plausible: Oscillatory onset may alter the shape of an embedded trajectory before a large-amplitude cycle is obvious. Nevertheless, a geometric feature must be computed according to a precise filtration, oriented consistently, and evaluated against simpler correlated indicators. Methodological novelty alone does not establish incremental warning value.

This study examines these issues in two simulation settings selected to represent distinct dynamical routes to instability. The first is a delayed Monod chemostat with an asserted delay-induced Hopf onset; the second is a harvested logistic population approaching a saddle-node fold. Five window features are compared with four alarm rules: a selected scalar indicator, a Cumulative Sum (CUSUM) detector, logistic-score fusion at a default threshold, and the same logistic score at a high training-negative quantile. The design separates 12 training from eight test trajectories and includes a five-trajectory doubled-noise stress set with parameters frozen. This structure permits three linked questions: whether features discriminate windows approaching onset, whether that discrimination becomes a timely and low-burden alarm, and whether the calibrated operating point survives a plausible noise shift.

The contribution is intentionally methodological rather than deployment-oriented. First, it integrates critical-transition indicators, sequential detection, and delay-embedding features within a causal evaluation pipeline. Second, it reports ranking and episode metrics side by side so that Area Under the Curve (AUC) cannot stand in for operational warning. Third, it compares transition mechanisms instead of pooling them into a single performance claim. Fourth, it subjects fixed thresholds to an explicit distribution-shift stress test. Finally, it clarifies the topological interpretation of the spanning-tree feature and identifies the exact computational artifacts needed for reproducibility. The resulting picture is more informative than a single best score: variance

is competitive for the oscillatory case, no tested feature resolves the fold case, fusion adds no demonstrated benefit, and noise sensitivity dominates the apparent lead when the operating distribution changes.

2 | Literature Review and Research Positioning

2.1 | Early-Warning Theory, Transition Specificity, and Limits

Critical slowing down links local stability to observable recovery behavior. In a one-dimensional linearization, deviations z from a stable equilibrium approximately satisfy $\frac{dz}{dt} = \lambda z + \text{noise}$, with $\lambda < 0$. As a control parameter approaches a fold or some other stability boundary, λ may approach zero from below. The characteristic recovery time $-\frac{1}{\lambda}$ then increases, and under suitable noise assumptions the stationary variance and short-lag autocorrelation rise [1–3]. This mechanism provides a principled basis for variance and autoregressive indicators, but it is local and conditional. It depends on the eigenmode seen by the sensor, the time scale of forcing, and whether the transition actually involves a vanishing recovery rate.

Simulation protocols have shown that indicator behavior is sensitive to detrending, bandwidth, window size, sampling interval, and the fraction of the pre-transition record used for evaluation [4]. Boettiger and Hastings [5] demonstrated that finite time series can generate false reassurance or false warnings because a trend statistic must distinguish approach to a bifurcation from ordinary stochastic variability. Jäger and Füllsack [6] further showed that standard early-warning workflows can produce systematic false positives under nonstationary but noncritical dynamics. These findings argue against judging an indicator by visual monotonicity or a single Kendall trend coefficient. A warning claim should instead define an onset, an admissible horizon, a decision threshold, and the denominator used for false alarms.

The type of deterioration also matters. Generalized-modeling approaches can estimate changing stability without a complete first-principles model, but they still depend on structural assumptions and informative observations [22]. Recovery-rate experiments in living systems support the idea that resilience can be measured directly under controlled perturbations [23], whereas passive operational streams rarely provide such clean perturbation-response pairs. Thermoacoustic experiments have likewise demonstrated that measurable warning signals can precede a transition, but their transfer across systems requires careful attention to noise and forcing [24]. Spectral indicators and deep-learning classifiers broaden the feature space [25], [26], yet they do not remove the need for prospective labels, threshold calibration, or external validation.

Recent systematic work has moved from asking whether an indicator trends to asking how combinations perform under deteriorating and noisy conditions. Proverbio et al. [7] compared multiple early-warning statistics across models and biological data and proposed optimized composite indicators. Their results reinforce a central point for bioprocess use: A composite score may help when component indicators carry complementary information, but its operating value depends on the deterioration mode and noise distribution. A fusion model trained on correlated transforms of one latent recovery process may add little beyond the strongest scalar feature, while appearing more sophisticated. That possibility is tested directly here by comparing logistic fusion with residual variance at a matched low false-alarm regime.

2.2 | Process Monitoring and Bioprocess Analytics

Data-driven process monitoring typically transforms multivariate measurements into lower-dimensional scores, residuals, latent states, or fault probabilities. Qin's survey organizes this field around representation, detection, diagnosis, and reconstruction [9], while the process-history review of Venkatasubramanian and colleagues emphasizes statistical and machine-learning methods that exploit archived operation [10]. The Tennessee Eastman benchmark illustrates why realistic evaluation requires interacting variables, disturbances, and closed-loop control rather than isolated univariate anomalies [11]. These traditions provide mature tools for fault surveillance, but many benchmarks label a fault after its injection time. Anticipating a nonlinear stability loss before onset requires a different target and stricter causal timing.

Bioprocesses intensify the challenge because key states are often unavailable at high frequency. Biomass, substrate, product concentration, metabolic burden, and viability may be observed through different sensors and at different sampling rates. Reviews of machine learning in bioprocess optimization and control describe growing use of soft sensors, hybrid models, adaptive control, and multivariate prediction [12], [13], [15]. Brunner et al. identify recurring obstacles in soft-sensor development, including limited batches, process drift, scale changes, missing reference measurements, and weak external validation [14]. These obstacles apply with equal force to early-warning systems. A detector derived from simulated biomass alone is therefore a proof of concept, not a substitute for multi-sensor validation under controlled and industrial conditions.

Alarm engineering also imposes requirements that are sometimes absent from feature studies. Operators encounter alarm episodes rather than independent windows, so persistence, suppression, reset logic, and refractory periods affect burden. A nominal probability of 0.001 per negative window is not equivalent to one actionable false alarm per 1,000 independent opportunities when windows overlap by 95%, as they do in the present design. Serial dependence can cluster exceedances, and a three-of-five rule changes both latency and event count. Thresholds must therefore be evaluated empirically on held-out episodes and under distribution shifts, not justified solely by a pointwise quantile.

2.3 | Nonlinear Complexity and Topological Summaries

Permutation entropy represents a time series through the relative ordering of successive values. For embedding order m, the $m!$ possible ordinal patterns form a discrete distribution, and normalized entropy measures how evenly those patterns occur [16]. It is attractive for monitoring because it does not depend on amplitude calibration in the same way as variance. However, short windows and low embedding order limit resolution, and entropy may respond to noise, oscillation, and deterministic complexity in overlapping ways. A declining or rising value therefore requires empirical orientation and cannot be assigned a universal direction of risk.

Delay-coordinate methods map a scalar series into vectors whose coordinates are time-shifted observations. Under appropriate conditions, the resulting point cloud preserves aspects of the underlying attractor. Sliding-window embeddings can reveal periodicity as a loop-like structure, and persistent homology tracks the birth and death of connected components or loops as a scale parameter grows [17], [18]. Degree-zero persistence is especially simple: For a Vietoris–Rips filtration using ordinary pairwise distances, the finite death scales of connected components correspond to edges in a minimum spanning tree. Degree-one persistence, by contrast, is the natural location for a prominent loop score associated with oscillation [19].

Applications to cryptocurrency and financial-crisis data show how topological summaries can change around regime shifts [20], [21]. These studies motivate, but do not validate, transfer to fermentation or cell-culture telemetry. Topological features are sensitive to the embedding dimension, lag, point density, subsampling, normalization, distance metric, and filtration convention. A maximum spanning tree emphasizes long edges and is not interchangeable with the minimum spanning tree that encodes degree-zero Vietoris–Rips merges. In the present experiment, the reported graph quantity is retained as a spanning-tree descriptor so its numerical performance can be compared, but it is not presented as verified persistent homology. A future rerun should compute a documented minimum-spanning-tree H_0 summary and, for oscillatory onset, a genuine H_1 loop-persistence statistic.

2.4 | From Score Discrimination to Calibrated Alarms

Sequential monitoring and classification provide different views of evidence. Page's cumulative-sum scheme accumulates small directional deviations and is designed around a reference shift and decision interval [27]. Logistic regression combines standardized features into a monotone risk score, while the software implementation used in the study follows standard machine-learning practice [28]. Receiver-operating-characteristic area summarizes the probability that a randomly selected positive window receives a higher score than a randomly selected negative window [29]. Because AUC is invariant to any monotone score

transformation, changing only the logistic threshold cannot change AUC; it changes sensitivity, lead, and false-alarm burden.

This distinction is decisive for early warning. A detector with $\mathrm{AUC} = 1$ can still produce zero qualifying alarms if its threshold is too high or if the persistence rule delays completion until after onset. Conversely, a detector with mediocre AUC can report very long lead if its threshold is crossed spuriously near the beginning of a run. Multiple testing procedures such as Benjamini–Hochberg control a defined error rate across a family of p-values [30], but they do not automatically solve sequential alarm dependence or replace explicit event counting. The present study therefore treats AUC as a ranking metric and detection, median lead, and false alarms per 1,000 negative windows as separate operating metrics.

Table 1 positions the study across the main literature strands. The research gap lies at their intersection: Few simulation comparisons simultaneously distinguish transition mechanism, enforce training-only calibration, translate overlapping window scores into persistent alarm episodes, test fixed thresholds under a noise shift, and audit the mathematical identity of an embedding-derived feature. Addressing that intersection is more valuable than adding another uncalibrated indicator to the literature.

Table 1. Literature strands, methodological lessons, and the role of the present study.

Literature Strand	Typical Tools	Unresolved Operational Issue	Response in This Study	Key Sources
Critical slowing down	Variance, autocorrelation, recovery rate	Indicator behavior depends on transition, forcing, sampling, and noise	Compare a Hopf-type and a fold-type case without pooling conclusions	[1–8], [22–24]
Process monitoring	Latent variables, residuals, soft sensors, fault diagnosis	Post-fault detection is not equivalent to pre-onset warning	Define causal windows, onset, persistence, lead, and alarm episodes	[9–15]
Complexity and topology	Permutation entropy, delay embeddings, persistent homology	Parameters and filtration semantics can change the statistic	Audit the spanning-tree descriptor and separate it from verified $\mathrm{H}_0/\mathrm{H}_1$ persistence	[16–21]
Sequential decisions	CUSUM, classifier scores, threshold rules	Ranking quality does not determine a usable operating point	Report AUC beside detection, lead, and false-alarm burden	[27–30]
Robustness	Held-out tests and distribution-shift evaluation	Nominal calibration can fail when variance or dependence changes	Freeze parameters and evaluate doubled-noise trajectories	[5–8], [14]

3 | Materials and Methods

3.1 | Study Architecture and Safeguards

The experiment is a simulation-only comparison with two dynamic cases, one measured channel, and three data partitions. Each case contains 20 nominal-noise trajectories indexed by seeds 0–19. Seeds 0–11 form the training set and seeds 12–19 form the held-out nominal test set. Five additional realizations, seeds 100–104, use twice the nominal process-noise scale and form a stress set. The stress trajectories are scored with the feature transformations, model coefficients, and thresholds selected under nominal training conditions. They are not used for recalibration.

The separation is intended to prevent three common forms of leakage: Using test windows to orient or select a feature, including test trajectories in standardization, and choosing a threshold after viewing test false alarms. All detrending is local to the current causal window, whereas any cross-window mean, scale, model coefficient,

feature direction, or threshold belongs to the training pipeline. *Fig. 1* summarizes this flow from simulated dynamics to held-out episode evaluation.

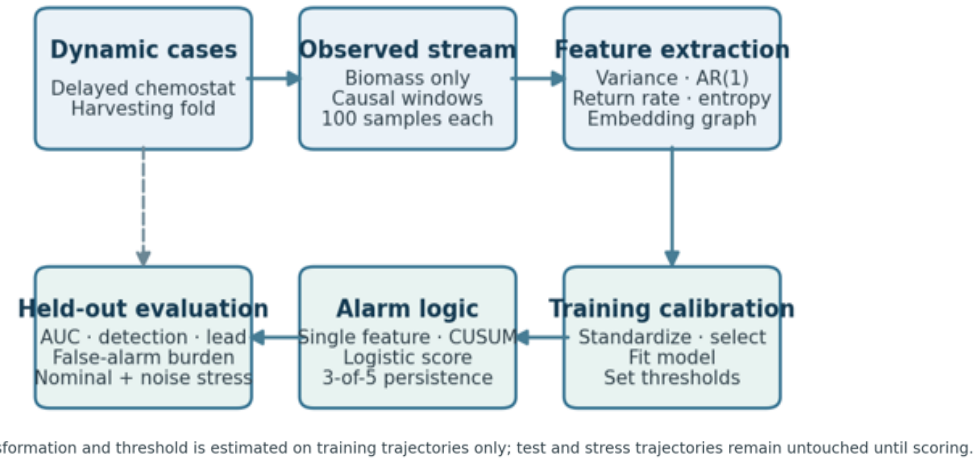


Fig. 1. Causal analysis and validation architecture.

Feature construction acts only on samples available at a window endpoint; transformations and alarm parameters are learned from training trajectories and then frozen for nominal test and doubled-noise stress evaluation.

3.2 | Simulated Transition Mechanisms

The delayed-chemostat case uses a Monod growth relation with maximum specific growth rate $\mu_{max} = 0.5\,h^{-1}$, half-saturation constant $K_s = 0.2\,g\,L^{-1}$, biomass yield $Y = 0.8$, and inlet substrate concentration $S_{in} = 5\,g\,L^{-1}$. The growth law is

$$\mu(S) = \frac{\mu_{max} S}{K_s + S}. \tag{1}$$

A delay $\tau = 0.85\,h$ is associated with growth and uptake, while the dilution rate decreases linearly from 0.45 to 0.26 h^{-1} over $N = 8{,}000$ integration steps of $\Delta t = 0.05\,h$. With n denoting the simulation index, the imposed ramp can be written as

$$D_n = 0.45 + \frac{(0.26 - 0.45)n}{N - 1}, \quad n = 0, \ldots, N - 1. \tag{2}$$

The recorded onset is the first crossing of a computed Hopf boundary at $D^* = 0.3016\,h^{-1}$, corresponding to approximately step 6,247. Because this is a delay-induced oscillatory transition, a complete reproduction requires the exact delayed state equations, history functions, stochastic forcing convention, and stability calculation. These implementation details must be distributed with the executable analysis; the numerical outcomes below are interpreted for this recorded model instance and not as a theorem about all delayed chemostats.

The harvesting case is a normalized logistic-growth model with carrying capacity $K = 10$ and saturating removal. Its stochastic differential representation is

$$dx = \left[x\left(1 - \frac{x}{10}\right) - \frac{h(t)x^2}{x^2 + 1}\right]dt + \sigma\, dW_t. \tag{3}$$

Harvesting effort h rises linearly from 2.0 to 2.8 over the same 8,000-step horizon, and the nominal process-noise scale is $\sigma = 0.05$. For the deterministic drift in *Eq. (3)*, equilibrium and tangency conditions give an

upper-branch fold near $x^* = 4.78$ and $h^* = 2.6044$. The ramp reaches this value near step 6,044, consistent with the recorded onset at step 6,043. This algebraic agreement supports the fold location, although it does not reproduce the stochastic realizations.

Table 2. Summary of the two simulation cases and data partitions.

Design Element	Delayed Chemostat	Harvesting Fold
Transition mechanism	Delayed oscillatory onset (chemostat)	Saddle-node fold (harvesting model)
Control parameter	Dilution D: $0.45 \rightarrow 0.26\ h^{-1}$	Harvesting effort h: $2.0 \rightarrow 2.8$
Recorded boundary	$D^* = 0.3016\ h^{-1}$; step $\approx 6{,}247$	$h^* \approx 2.6044$; step $\approx 6{,}043$
Time discretization	8,000 steps; $\Delta t = 0.05$ h	8,000 steps; $\Delta t = 0.05$ (normalized time)
Observed channel	Biomass	Biomass/state x
Nominal trajectories	12 training + 8 test	12 training + 8 test
Stress trajectories	5 with doubled process noise	5 with doubled process noise

3.3 | Causal Windows and Feature Construction

Only biomass is used for feature extraction. A window contains $L = 100$ consecutive samples and advances by $s = 5$ samples, producing 95% overlap. At $\Delta t = 0.05$ h in the chemostat, one stride corresponds to 0.25 h and one full window spans 5 h. Every feature is assigned to the last sample observed in its window. This end-time convention is essential: Labeling a window by its center or future edge would overstate lead and can leak post-window information into a forecast target.

Within each window $y_1, \ldots, y_L$, a least-squares linear trend $\hat{a} + \hat{b}j$ is removed and the residuals are $e_j = y_j - (\hat{a} + \hat{b}j)$. Residual variance is the sample variance of e_j. The lag-one autoregressive coefficient is estimated by least squares without an additional intercept after detrending,

$$\hat{\varphi} = \frac{\sum_{j=2}^{L} e_j\, e_{j-1}}{\sum_{j=2}^{L} e_{j-1}^2}. \tag{4}$$

The corresponding return-rate feature is $r = -\log|\hat{\varphi}|$. When $|\hat{\varphi}|$ approaches one, r approaches zero, which is consistent with slower local recovery; because r is a deterministic transformation of $\hat{\varphi}$, the two features are strongly dependent and should not be interpreted as independent mechanisms. Numerical handling must also be defined for $|\hat{\varphi}| = 0$ or estimates at and beyond unity.

Permutation entropy uses ordinal patterns of order $m = 3$. If p_π is the observed frequency of pattern π among the $3!$ possibilities, the normalized measure is

$$H_{PE} = -\frac{\sum_{\pi} p_\pi \log(p_\pi)}{\log(3!)}, \tag{5}$$

The geometric descriptor starts from a three-dimensional delay embedding $z_j = (e_j, e_{j-5}, e_{j-10})$ and a 12-point subsample. The recorded implementation describes the largest edge of a maximum spanning tree. That quantity is treated here as an embedding-derived graph descriptor. It is not called degree-zero persistence because H_0 Vietoris–Rips component deaths correspond to edges of a minimum spanning tree under the same metric [19]. Correcting the tree changes the feature and requires rerunning every score, threshold, and table; semantic relabeling after the fact would be inappropriate.

A window is defined as positive when the independently specified transition onset lies within the next 200 window strides and has not yet occurred; earlier windows are negative. At five samples per stride, this horizon corresponds nominally to 1,000 samples or 50 h in the chemostat. Post-onset windows are excluded from

early-warning evaluation. The horizon is an analysis choice rather than a physical guarantee and should be varied in sensitivity analysis.

3.4 | Detectors, Thresholds, and Alarm Persistence

Four detector configurations are evaluated. The single-indicator rule selects a feature and orientation from training performance, then applies a training-derived high threshold. The CUSUM rule accumulates evidence from autoregressive residuals according to Page's sequential formulation [27]. Logistic fusion standardizes the feature vector with training statistics and estimates a logistic score using the training windows [28]. Two decision thresholds are applied to exactly the same fitted logistic score: the conventional probability cutoff 0.5 and a high quantile of scores from training-negative windows chosen to target approximately one exceedance per 1,000 negative windows. Because the score ordering is unchanged, the two logistic configurations must have identical AUC; only their operating metrics can differ.

A raw exceedance becomes an alarm when at least three of the most recent five window scores exceed the detector threshold. The alarm time is the endpoint of the window that completes this condition. This persistence rule suppresses isolated spikes but can add up to four strides of decision delay and creates an episode structure in which adjacent exceedances are not independent. A new false-alarm episode should be counted only after the previous episode has reset; the reset and refractory conventions must be identical in training, test, and stress evaluation.

Fig. 2 formalizes the causal timing. The first completed alarm before τ^* is used to compute lead. An exceedance far before the predefined horizon is a false alarm, even if the same trajectory later transitions. An alarm after τ^* is not an early detection. These rules prevent favorable relabeling of early nuisance events as long-lead warnings.

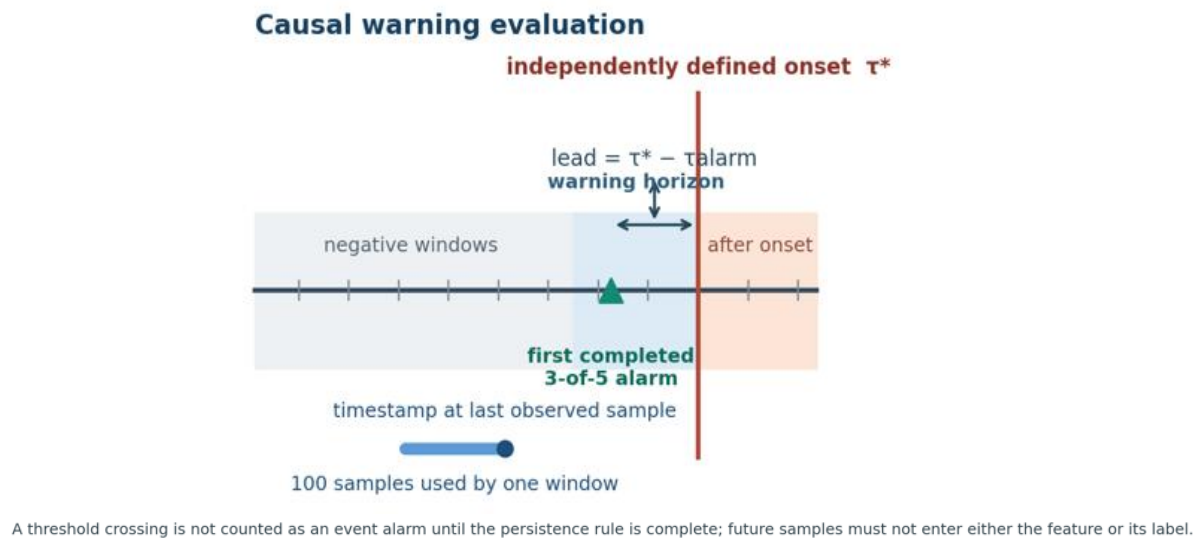


Fig. 2. Timing semantics for causal early-warning evaluation.

Windows are timestamped at their last observed sample; a detector event occurs only when the persistence rule is complete. Lead is measured from the first qualifying pre-onset alarm to the independently defined transition onset.

3.5 | Performance Measures and Statistical Interpretation

Window-level AUC measures score discrimination across labeled windows and is interpreted using its standard rank-probability meaning [29]. It does not incorporate alarm persistence, lead, or event burden. Episode detection is the number of test trajectories with at least one qualifying alarm divided by eight. For

detected trajectories, lead is the number of five-sample strides between alarm completion and onset; median lead is undefined when no trajectory is detected. False-alarm burden is reported as the number of pre-horizon alarm episodes divided by the number of eligible negative windows and multiplied by 1,000.

The small number of independent trajectories limits formal inference. Overlapping windows increase the apparent sample size but not the number of independent simulation realizations. Results are therefore reported descriptively by trajectory count, median lead, and event rate. A paired comparison in which seven of eight lead differences are zero and only one is nonzero contains insufficient information for a meaningful superiority claim; an exact two-sided signed-rank calculation is $p = 1.0$ under that pattern. Similarly, counting two thresholds of the same logistic score as independent algorithms in a rank test would duplicate one ranking. Multiple-testing correction [30] is not applied because the present analysis does not define a valid family of independent p-values.

The doubled-noise set is a transportability check rather than a second tuning set. A robust detector should ideally retain an acceptable event burden when the measurement or process variance changes moderately. If AUC remains similar but false alarms rise sharply, the failure lies primarily in calibration and threshold transfer, not necessarily in score ordering. This distinction guides both diagnosis and redesign.

4 | Results

4.1 | Nominal-Noise Alarm Performance

Table 3 presents the recorded held-out results for eight trajectories per system. In the delayed-chemostat case, residual variance detects all eight transitions with a median lead of 56 strides and 0.84 false-alarm episodes per 1,000 negative windows. At 0.25 h per stride, the nominal median interval is 14 h, although effective intervention time would be shorter once process actuation, assay delay, and persistence latency are included. The training-quantile logistic rule also detects 8/8 trajectories at the same median lead, with a slightly higher false-alarm rate of 0.96. Within the resolution of this experiment, the fusion does not improve the selected variance rule.

The logistic score at threshold 0.5 illustrates why an unconstrained operating point can look attractive and still be unsuitable. It detects all eight chemostat transitions with a much longer median lead of 474 strides, but its false-alarm burden is 13.72 per 1,000 negative windows more than an order of magnitude above the low-alarm rules. Those early crossings cannot be interpreted independently of the increased nuisance burden. CUSUM presents the opposite pattern: its reported AUC is 1.000, but it yields no qualifying pre-onset alarm at the selected threshold. Perfect rank separation therefore does not imply operational detection.

The harvesting fold is a negative case under the tested observation and design. Residual variance detects 2/8 trajectories and the quantile logistic rule detects 3/8. Their median leads are very large because the few threshold crossings occur early, not because a stable and localized pre-fold signal is demonstrated. Logistic threshold 0.5 produces no qualifying alarm. These results do not establish that folds are intrinsically unpredictable; they show that this biomass channel, window design, and feature set do not resolve this simulated fold reliably.

Table 3. Recorded nominal-noise test performance over eight trajectories per system. Lead is measured in five-sample window strides; n/a denotes no qualifying pre-onset detection.

System	Detector	AUC	Median Lead	Detection	False Alarms / 1,000
Hopf	Variance	0.784	56	8/8	0.84
Hopf	CUSUM	1.000	n/a	0/8	0.00
Hopf	Logistic, threshold 0.5	0.786	474	8/8	13.72
Hopf	Training-quantile logistic	0.786	56	8/8	0.96
Fold	Variance	0.504	1,042	2/8	1.38
Fold	CUSUM	0.432	1,194	2/8	2.00
Fold	Logistic, threshold 0.5	0.501	n/a	0/8	0.00
Fold	Training-quantile logistic	0.501	951	3/8	2.38

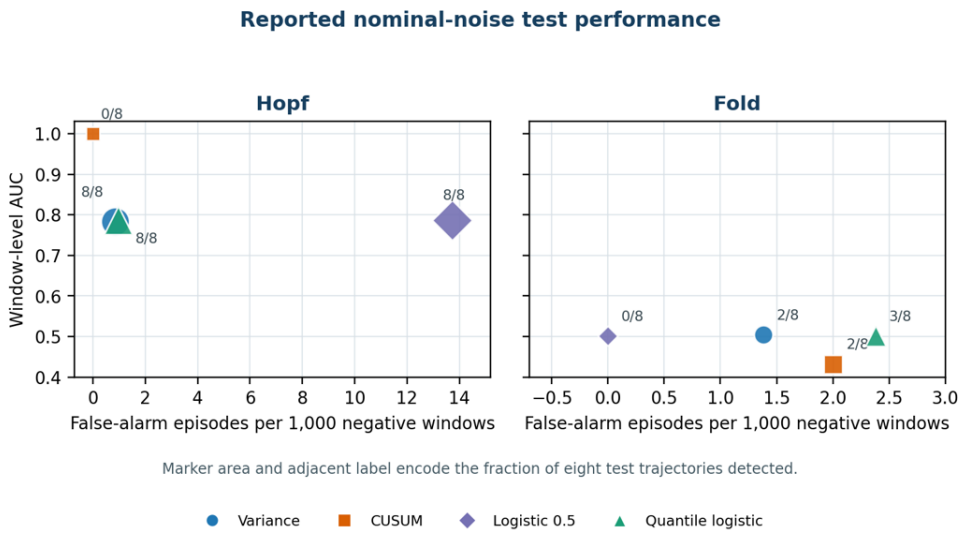


Fig. 3. Recorded nominal-noise trade-off between window-level discrimination and episode-level false alarms. Marker area and labels represent the detection fraction.

The chemostat CUSUM point demonstrates that high AUC can coexist with zero alarm detection, while the logistic 0.5 rule exchanges longer apparent lead for a substantially higher false-alarm burden.

4.2 | Feature Discrimination

Feature-level AUCs in *Table 4* explain the detector outcomes. Residual variance is the strongest recorded scalar indicator for the chemostat test trajectories (0.785), followed by lag-one autoregression (0.755). Return rate has the opposite orientation because it is a decreasing transform of $|\hat{\varphi}|$; its AUC of 0.245 corresponds to 0.755 after a direction reversal fixed in advance. This equality reinforces that the two quantities do not contribute independent ranking information. Permutation entropy is below chance in the reported orientation, and the spanning-tree descriptor has an AUC of 0.222.

Reversing the graph score after observing test results would produce 0.778, but this would be post-selection and cannot be credited as independent performance. Its train AUC of 0.257 suggests that orientation could have been chosen from training data, yet the meaning of the feature remains unresolved because the computation is described as a maximum spanning tree. A corrected topological rerun should predefine direction from training data and report both H_0 and H_1 summaries with sensitivity to lag and subsampling.

All fold feature AUCs lie between 0.480 and 0.520 on the test set. The narrow range around chance is consistent with the weak episode detection in *Table 3*. Importantly, train and test AUCs are also similar, so the principal issue is not an obvious train–test collapse; rather, the selected biomass-derived features contain

little discriminative signal for the chosen pre-fold horizon. Additional state channels, a shorter horizon, a perturbation-derived recovery measure, or mechanistic residuals may be required.

Table 4. Recorded feature AUCs. The spanning-tree descriptor is not interpreted as verified persistent homology pending reconstruction of the graph computation.

Feature	Hopf Train	Hopf Test	Fold Train	Fold Test
Residual variance	0.750	0.785	0.548	0.505
Lag-one autoregression	0.729	0.755	0.542	0.520
Return rate	0.271	0.245	0.458	0.480
Permutation entropy	0.389	0.347	0.482	0.510
Embedding spanning-tree descriptor	0.257	0.222	0.454	0.494

4.3 | Doubled-Noise Stress Test

Table 5 reports the five-trajectory stress evaluation with all nominally learned parameters frozen. In the chemostat, variance and the quantile logistic rule retain AUCs near 0.76 and detect all five transitions. However, false-alarm rates rise from below one to 165.35 and 147.06 per 1,000 negative windows. Their apparent median leads exceed 1,100 strides because threshold crossings occur much earlier under the noisier distribution. Such lead is not actionable evidence; it is a symptom of calibration failure.

The fold stress set behaves similarly but more severely. Both reported AUCs remain modest, yet every trajectory is flagged and false-alarm rates reach 279.96 and 307.41 per 1,000 negative windows. The nominal training quantile therefore does not transport to doubled noise. Because residual variance scales directly with noise amplitude and correlated windows propagate an exceedance across several scores, this failure is expected unless the detector explicitly normalizes or models the current noise regime.

The stress test changes the interpretation of the nominal results. The chemostat variance rule is promising within the nominal simulation distribution, but it is not yet a robust alarm. Online recalibration cannot simply be assumed as a remedy because adapting thresholds on a deteriorating stream may absorb the very signal of interest. A defensible redesign would estimate a slowly varying noise baseline from verified stable periods, normalize features by that baseline, impose guardrails on adaptation, and validate the entire policy on untouched trajectories spanning noise, ramp speed, sampling, and sensor-filter conditions.

Table 5. Recorded doubled-noise stress performance over five trajectories per system with nominal thresholds and model parameters frozen.

System	Detector	AUC	Median Lead	Detection	False Alarms / 1,000
Hopf	Variance	0.765	1,178	5/5	165.35
Hopf	Training-quantile logistic	0.763	1,160	5/5	147.06
Fold	Variance	0.568	1,156	5/5	279.96
Fold	Training-quantile logistic	0.566	1,169	5/5	307.41

5 | Discussion

5.1 | Principal Findings and Mechanistic Interpretation

The strongest result is not that a complex detector wins, but that an aligned simple statistic can match it. In the delayed-chemostat simulation, residual variance and training-quantile logistic fusion have the same detection count and median lead at nearly the same false-alarm burden. This pattern is consistent with a

dominant change in fluctuation amplitude as the system approaches oscillatory instability. Because autoregression, return rate, and variance all respond to related aspects of slowing and because the logistic model is linear in standardized features, fusion has limited complementary information to exploit.

The fold case provides an important counterweight. Critical slowing down can occur near a saddle-node bifurcation, but its observability depends on where the trajectory lies, how quickly the parameter is ramped, which state is measured, and how process noise excites the critical mode. A 200-window label horizon may include many windows too distant from the fold to be distinguishable; conversely, shortening the horizon reduces positive samples. The near-chance AUCs show that the present design does not resolve this trade-off. It would be scientifically incorrect to generalize the negative result to all folds or to claim that the same detector transfers across transition mechanisms.

The comparison between AUC and alarms is equally consequential. CUSUM's reported chemostat AUC of 1.000 with 0/8 detections is not contradictory. AUC evaluates ranking over all labeled windows, while the selected threshold and persistence rule determine whether a qualifying alarm occurs. The default logistic threshold shows the reverse problem: all transitions are detected early, but nuisance events increase sharply. Evaluating only one metric would produce opposite and incomplete conclusions. A credible early-warning paper should always pair ranking with event-based sensitivity, lead, and burden.

5.2 | What Fusion and Topology Can—and Cannot—Add

Multivariate fusion is useful when components observe different latent processes or fail under different conditions. Here, lag-one autoregression and return rate are transformations of one coefficient, and several indicators share the same detrended residuals. Multicollinearity can make logistic coefficients unstable without harming prediction, but it weakens mechanistic interpretation. Future models should compare fusion against a parsimonious baseline under nested cross-validation and quantify incremental value using held-out event metrics, not coefficient magnitude. Nonlinear fusion or deep learning [26] would require substantially more independent trajectories and a carefully designed simulation ensemble to avoid learning idiosyncratic ramps.

Topology should be held to the same standard. Delay embeddings are well motivated for oscillatory signals, and a genuine H_1 loop feature may become informative as a limit cycle emerges. But the current maximum-spanning-tree descriptor is neither a verified H_0 persistence statistic nor an H_1 loop measure. Its poor recorded orientation does not refute topological early warning; it evaluates one incompletely specified graph summary. A rigorous extension should define preprocessing, distance metric, embedding lag and dimension, point subsampling, filtration, persistence summary, and orientation before viewing test results. It should then compare the topological feature with variance at matched false-alarm rates and report computational cost.

The broader lesson is that sophistication should be earned by incremental evidence. A feature deserves operational use if it improves detection or lead at a tolerable alarm burden, remains stable under plausible shifts, and provides value beyond simpler observables. Visual novelty, a complex model, or an AUC gain of a few thousandths does not satisfy that criterion. In the nominal chemostat experiment, variance is therefore the preferred baseline.

5.3 | Implications for Bioprocess Deployment

A deployable early-warning system would operate inside a larger monitoring architecture. First, the transition of concern must be linked to a controllable intervention, such as dilution adjustment, feed modification, gas transfer, or harvest-rate reduction. Second, onset and response deadlines must be expressed in physical time after accounting for sensor filtering, assay latency, window accumulation, persistence logic, communication, and actuation. Third, the alarm policy should include reset, suppression, and escalation rules so that a burst of overlapping exceedances does not become dozens of operator notifications.

Multiple channels are likely necessary. Biomass variance may respond to sensor noise, agitation changes, sampling artifacts, or genuine process instability. Joint evidence from substrate, dissolved oxygen, off-gas, pH, redox state, and soft-sensor residuals can help distinguish those causes. Hybrid models could use

mechanistic residuals to remove expected ramp behavior before applying generic early-warning features. Importantly, any adaptive normalization must be slower than the warning dynamics and anchored to confirmed stable operation; otherwise, adaptation can erase a gradual warning trend.

Validation should progress in stages: simulation ensembles with varied parameters and noise spectra; retrospective data from controlled transition experiments; prospective shadow-mode deployment without operator action; and only then an intervention study. At each stage, acceptance criteria should be specified before evaluation. These may include minimum episode sensitivity, a maximum weekly nuisance-alarm count, a clinically or operationally meaningful lead, and bounded degradation under predefined distribution shifts. Such criteria translate abstract indicator trends into an auditable decision system.

5.4 | Limitations and Reproducibility Requirements

The study is limited to simulated dynamics, one observed channel, two model instances, 20 nominal realizations per system, and five stress realizations. Trajectories generated from different random seeds are more independent than overlapping windows, but they still share equations and parameter ramps. Uncertainty intervals from eight test trajectories would be wide, and the results cannot establish industrial generality. The fold and Hopf cases are not biological replicates; they probe different mechanisms and should remain separate.

Several implementation details are required for exact reproduction. The delayed chemostat needs its complete state equations, delayed-variable placement, history function, stochastic-integrator convention, and Hopf calculation. The analysis package must define random-number generation, boundary handling, detrending, autoregression edge cases, permutation-pattern ties, embedding subsampling, threshold quantiles, alarm reset, and false-event counting. It must also reconcile the maximum/minimum spanning-tree discrepancy and regenerate all dependent results if the graph feature changes.

The computational release should include the executable script identified for the experiment, a machine-readable result file, dependency versions, fixed seeds, per-trajectory state series, per-window features and labels, fitted training parameters, alarm timestamps, and code that rebuilds every table and figure. Source-generated numerical arrays are preferable to screenshots. A reproducibility manifest should record checksums and the command required to run the complete pipeline. Until those materials accompany the article, the numerical tables should be interpreted as recorded simulation outputs rather than independently reproduced estimates.

Finally, doubled white process noise is only one stress condition. Real sensors can exhibit colored noise, drift, calibration jumps, missing values, delayed assays, saturation, and controller-induced dependence. Model parameters and ramp speed may vary between runs, and the true onset may be uncertain rather than exactly known. A future validation grid should vary these factors independently and in combination, with a final untouched test ensemble reserved for the full alarm policy.

6 | Conclusion

This study reframes simulated bioprocess early warning as a calibrated sequential decision rather than a search for a visually rising indicator. In the delayed-chemostat case, residual variance provides the best parsimonious operating point: It detects all eight nominal test transitions with the same median lead as logistic fusion and a slightly lower false-alarm rate. The harvesting-fold case remains unresolved by the tested biomass features, demonstrating that warning skill is transition- and observation-specific. CUSUM and default logistic results show why AUC, lead, and nuisance burden must be interpreted jointly.

The doubled-noise experiment reveals the principal obstacle to deployment. Scores retain much of their ranking, yet nominal thresholds generate hundreds of times more false alarms. Robust early warning therefore requires explicit treatment of noise regime, dependence, persistence, and threshold transport. Delay-embedding topology remains a promising extension for oscillatory onset, but it must be recomputed with a

mathematically valid filtration and evaluated for incremental event-level value. With a complete computational release and broader validation ensemble, the present framework can serve as a transparent benchmark for comparing simple and advanced warning methods before experimental bioprocess application.

Data Availability

The study uses simulated data. A complete release should include the executable simulation and analysis code, machine-readable results, versioned dependencies, random seeds, per-trajectory outputs, per-window features, and alarm-event records needed to reproduce the tables and figures. These computational materials were not embedded in the manuscript file and should accompany the final publication package.

Funding

The authors declare that no specific grant supported this work.

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

H. M.: Conceptualization, methodology, investigation, and original drafting.

B. J.: Software, validation, statistical analysis, and review and editing.

Both authors approved the final manuscript.

Generative AI Disclosure

Generative AI tools were used under author supervision to assist language refinement and document preparation. The authors reviewed the complete manuscript and retain responsibility for scientific accuracy, citations, interpretations, and conclusions.

References

[1] Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, *461*(7260), 53–59. https://doi.org/10.1038/nature08227

[2] Dakos, V., Scheffer, M., Van Nes, E. H., Brovkin, V., Petoukhov, V., & Held, H. (2008). Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, *105*(38), 14308–14312. https://doi.org/10.1073/pnas.0802430105

[3] Kuehn, C. (2011). A mathematical framework for critical transitions: Bifurcations, fast-slow systems and stochastic dynamics. *Physica D: Nonlinear Phenomena*, *240*(12), 1020–1035. https://doi.org/10.1016/j.physd.2011.02.012

[4] Dakos, V., Carpenter, S. R., Brock, W. A., Ellison, A. M., Guttal, V., & Ives, A. R. (2012). Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PloS One*, *7*(7), e41010. https://doi.org/10.1371/journal.pone.0041010

[5] Boettiger, C., & Hastings, A. (2012). Quantifying limits to detection of early warning for critical transitions. *Journal of the Royal Society Interface*, *9*(75), 2527. https://doi.org/10.1098/rsif.2012.0125

[6] Jäger, G., & Füllsack, M. (2019). Systematically false positives in early warning signal analysis. *PLoS One*, *14*(2), e0211072. https://doi.org/10.1371/journal.pone.0211072

[7] Proverbio, D., Skupin, A., & Gonçalves, J. (2023). Systematic analysis and optimization of early warning signals for critical transitions using distribution data. *IScience*, *26*(7). https://doi.org/10.1016/j.isci.2023.107156

[8] George, S. V, Kachhara, S., & Ambika, G. (2023). Early warning signals for critical transitions in complex systems. *Physica Scripta*, *98*(7), 72002. https://doi.org/10.1088/1402-4896/acde20

[9] Qin, S. J. (2012). Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, *36*(2), 220–234. https://doi.org/10.1016/j.arcontrol.2012.09.004

[10] Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., & Yin, K. (2003). A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering*, *27*(3), 327–346. https://doi.org/10.1016/S0098-1354(02)00162-X.

[11] Downs, J. J., & Vogel, E. F. (1993). A plant-wide industrial process control problem. *Computers & Chemical Engineering*, *17*(3), 245–255. https://doi.org/10.1016/0098-1354(93)80018-I

[12] Mondal, P. P., Galodha, A., Verma, V. K., Singh, V., Show, P. L., Awasthi, M. K., & Jain, R. (2023). Review on machine learning-based bioprocess optimization, monitoring, and control systems. *Bioresource Technology*, *370*, 128523. https://doi.org/10.1016/j.biortech.2022.128523

[13] Lim, S. J., Son, M., Ki, S. J., Suh, S. I., & Chung, J. (2023). Opportunities and challenges of machine learning in bioprocesses: categorization from different perspectives and future direction. *Bioresource Technology*, *370*, 128518. https://doi.org/10.1016/j.biortech.2022.128518

[14] Brunner, V., Siegl, M., Geier, D., & Becker, T. (2021). Challenges in the development of soft sensors for bioprocesses: A critical review. *Frontiers in Bioengineering and Biotechnology*, *9*, 722202. https://doi.org/10.3389/fbioe.2021.722202

[15] Mowbray, M., Savage, T., Wu, C., Song, Z., Cho, B. A., Del Rio Chanona, E. A., & Zhang, D. (2021). Machine learning for biochemical engineering: A review. *Biochemical Engineering Journal*, *172*, 108054. https://doi.org/10.1016/j.bej.2021.108054

[16] Bandt, C., & Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, *88*(17), 174102. https://doi.org/10.1103/PhysRevLett.88.174102

[17] Perea, J. A., & Harer, J. (2015). Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, *15*(3), 799–838. https://doi.org/10.1007/s10208-014-9206-z

[18] Perea, J. A., Deckard, A., Haase, S. B., & Harer, J. (2015). SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *BMC Bioinformatics*, *16*(1), 257. https://doi.org/10.1186/s12859-015-0645-6

[19] Edelsbrunner, H., & Harer, J. (2010). *Computational topology: An introduction*. American Mathematical Soc. https://www.researchgate.net/profile/John-Harer/publication/220692408

[20] Gidea, M., Goldsmith, D., Katz, Y., Roldan, P., & Shmalo, Y. (2020). Topological recognition of critical transitions in time series of cryptocurrencies. *Physica A: Statistical Mechanics and Its Applications*, *548*, 123843. https://doi.org/10.1016/j.physa.2019.123843

[21] Ismail, M. S., Noorani, M. S. M., Ismail, M., Razak, F. A., & Alias, M. A. (2022). Early warning signals of financial crises using persistent homology. *Physica A: Statistical Mechanics and Its Applications*, *586*, 126459. https://doi.org/10.1016/j.physa.2021.126459

[22] Lade, S. J., & Gross, T. (2012). Early warning signals for critical transitions: A generalized modeling approach. *PLoS Computational Biology*, *8*(2), e1002360. https://doi.org/10.1371/journal.pcbi.1002360

[23] Veraart, A. J., Faassen, E. J., Dakos, V., Van Nes, E. H., Lürling, M., & Scheffer, M. (2012). Recovery rates reflect distance to a tipping point in a living system. *Nature*, *481*(7381), 357–359. https://doi.org/10.1038/nature10723

[24] Gopalakrishnan, E. A., Sharma, Y., John, T., Dutta, P. S., & Sujith, R. I. (2016). Early warning signals for critical transitions in a thermoacoustic system. *Scientific reports*, *6*(1), 35310. https://doi.org/10.1038/srep35310

[25] Bury, T. M., Bauch, C. T., & Anand, M. (2020). Detecting and distinguishing tipping points using spectral early warning signals. *Journal of the Royal Society Interface*, *17*(170), 20200482. https://doi.org/10.1098/rsif.2020.0482

[26] Bury, T. M., Sujith, R. I., Pavithran, I., Scheffer, M., Lenton, T. M., Anand, M., & Bauch, C. T. (2021). Deep learning for early warning signals of tipping points. *Proceedings of the National Academy of Sciences*, *118*(39), e2106140118. https://doi.org/10.1073/pnas.2106140118

[27] Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*(1/2), 100–115. https://doi.org/10.2307/2333009

[28] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

[29] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

[30] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x