

Anomaly detection in nonlinear sensor dynamics for intelligent monitoring

Denys Symonov^{1,*†}, Oleksandr Palagin^{1,†}, Bohdan Zaika^{1,†} and Yehor Symonov^{1,†}

¹ V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences (NAS) of Ukraine, Akademician Glushkov Avenue 40, 03187, Kyiv, Ukraine

Abstract

The paper proposes a methodological framework for anomaly detection and diagnostic interpretation in nonlinear multivariate sensor dynamics. The main contribution is the integration of three complementary levels of analysis: local entropy-based detection, episode-level nonlinear dynamical characterization, and attention-based predictive validation. The Kozachenko–Leonenko local entropy score with adaptive moving-window thresholds is used to construct an alert set of candidate anomalous observations without imposing a parametric distributional model. These observations are then transformed into temporally extended anomalous episodes and analyzed in a reconstructed phase space. Lyapunov-based sensitivity with multiple surrogate realizations is introduced to assess evidence consistent with local trajectory instability, while correlation-dimension estimation is used only as an auxiliary geometric diagnostic. Multi-head attention is incorporated not as a causal explanation mechanism, but as a predictive consistency check for reconstructed trajectory windows. A case study on one sensor-array subset illustrates how the proposed framework separates pointwise alert generation, episode-level dynamical qualification, and predictive consistency assessment. The study evaluates the feasibility of the staged diagnostic procedure but does not, by itself, establish superiority over baseline detectors or generalization across gases, concentrations, and sensor arrays.

Keywords

nonlinear time series analysis, local entropy estimation, phase space reconstruction, Lyapunov analysis, anomalous episode detection, attention based validation, multivariate sensor data

1. Introduction

Anomalies in multivariate sensor data may originate from noise, sensor malfunction, external disturbances, or genuine changes in system dynamics. Their interpretation is complicated by temporal dependencies and interactions among variables, which often determine whether an observation is truly atypical [1]. In modern sensor systems, abnormal behavior may also be distributed across several channels [2]. Reliable distinction between these cases is particularly important in industrial fault diagnosis [3] and environmental monitoring [4].

This study proposes a staged framework that separates anomaly detection from dynamical interpretation. Candidate observations are first identified using a local estimate of differential Shannon entropy based on the Kozachenko–Leonenko nearest-neighbor principle [5]. Temporally adjacent detections are then grouped into episodes. These intervals are examined in reconstructed state space, where delay-coordinate representations provide access to trajectory structure [6] and support nonlinear time-series analysis [7]. The maximum Lyapunov exponent is used to characterize

ADT-ISMDCC 2026: Advanced Digital Technologies: Intelligent Systems, Modeling, and Data-Driven Control, September 22–24, 2026, Odesa, Ukraine

^{1*} Corresponding author.

[†] These authors contributed equally.

✉ denys.symonov@gmail.com (D. Symonov); palagin_a@ukr.net (O. Palagin); zaikabohdan5@gmail.com (B. Zaika); e.symonov@gmail.com (Y. Symonov)

ORCID 0000-0002-6648-4736 (D. Symonov); 0000-0003-3223-1391 (O. Palagin); 0009-0001-9567-8361 (B. Zaika); 0009-0008-2581-2001 (Y. Symonov)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

local instability [8], whereas the correlation dimension describes the effective geometrical complexity of the reconstructed dynamics [9]. A predictive stage is added to determine whether the detected temporal structure remains informative outside the fitted data. Learning-based models have shown their ability to represent nonlinear dynamical patterns in observed trajectories [10]. Here, multi-head attention is used to capture temporal and inter-variable dependencies relevant to prediction [11], while attention weights are not interpreted as direct causal explanations because they may not faithfully represent input importance [12]. The resulting framework combines local information-based detection, episode formation, nonlinear trajectory analysis, and predictive verification. Experiments on multichannel sensor-array data with nonlinear temporal responses indicate that these components provide complementary evidence for distinguishing transient irregularities from structured changes in system behavior.

2. Problem statement

The main problem is to distinguish short-term irregularities in multivariate sensor data from changes that reflect modifications of the underlying system dynamics. Point-wise anomaly detection alone cannot provide such interpretation, especially for nonlinear and nonstationary processes. A suitable diagnostic procedure should therefore combine local anomaly identification with analysis of temporally coherent episodes. These episodes must be examined in terms of reconstructed trajectory structure, dynamical stability, and predictive relevance. The problem is thus formulated as linking local statistical deviations with trajectory-level and predictive evidence within a unified analysis framework.

3. Goal and objectives of the work

The aim of this study is to develop a multilevel framework for detecting and interpreting anomalous episodes in nonlinear multivariate sensor time series. Therefore, the objectives of the study are to apply the Kozachenko–Leonenko entropy estimator to identify locally atypical observations and group them into temporal episodes; to characterize these episodes in reconstructed state space using the maximum Lyapunov exponent and correlation dimension; to use an attention-based model to assess temporal and inter-variable dependencies; and to evaluate the complete procedure on multichannel sensor data, with particular attention to its practical applicability and limitations.

4. Detection of anomalies

Entropy-based methods provide a useful class of tools for anomaly detection in multivariate time series, since they make it possible to assess changes in the local structure of empirical distributions without imposing a strict parametric model. Recent studies have used entropy-related measures in combination with graph neural networks, transfer entropy, and graph attention mechanisms for anomaly detection in uncertain and high-dimensional environments [13, 14]. However, for nonlinear sensor dynamics, an important methodological requirement is to obtain a local, distribution-sensitive indicator that can respond to changes in empirical density and can subsequently be linked to phase-space and predictive analyses. The Kozachenko–Leonenko estimator is a nonparametric k -nearest-neighbor estimator of differential entropy that uses distances between neighboring observations to approximate the local structure of the underlying probability distribution [15–17].

Let $x_t \in \mathbb{R}^d$, $t = \overline{1, N}$, denote a multivariate sensor observation at time t , where d is the number of measured features. In the entropy estimation step, the ordered time series is treated as a finite sample $\{x_t\}_{t=1}^N$ from an underlying d -dimensional distribution, while its temporal ordering is used later for phase-space reconstruction. The Kozachenko–Leonenko estimator of differential entropy is then computed from the distances between neighboring observations as follows:

$$H_{KLE} = \psi(N) - \psi(k) + \log V_d + \frac{d}{N} \sum_{t=1}^N \ln \rho_k(x_t) + \gamma, \quad (1)$$

where $\psi(\cdot)$ is the digamma function; k is the number of nearest neighbors; V_d is the volume of a unit ball in a d -dimensional space, $V_d = \frac{\pi^{d/2}}{\Gamma(d/2+1)}$; $\rho_k(x_t)$ is the distance from x_t to its k -th nearest neighbor; γ is the Euler–Mascheroni constant.

Local entropy can provide an indication of atypical behavior in a time series. When individual observations deviate substantially from the statistical structure of the surrounding data, the corresponding local density estimate is distorted, which is subsequently reflected in noticeably different KLE values. For each observation x_t , a local entropy-based anomaly score is computed using the distance $\rho_k(x_t)$ from x_t to its k -th nearest neighbor:

$$H_t^{local} = \psi(N) - \psi(k) + \log V_d + d \cdot \ln \rho_k(x_t) + \gamma. \quad (2)$$

The additive term $\psi(N) - \psi(k) + \log V_d + \gamma$ is constant for fixed N , k , and d ; therefore, the ranking of observations by the local entropy-based score is determined by the neighbor-distance term $d \cdot \ln \rho_k(x_t)$.

For a fixed significance level $\alpha \in (0, 1/2)$, global reference thresholds may first be defined from the complete set of local scores as

$$H^- = Q_\alpha(H^{local}), H^+ = Q_{1-\alpha}(H^{local}). \quad (3)$$

where Q_α denotes the empirical quantile of order α . An observation x_t is classified as anomalous if

$$x_t \in A \Leftrightarrow H_t^{local} < H^- \vee H_t^{local} > H^+. \quad (4)$$

Thus, both unusually low and unusually high local entropy scores are treated as indicators of possible anomalous regimes. These global thresholds provide a stationary reference formulation. For nonstationary monitoring, they are replaced by thresholds estimated from a preceding moving window.

For nonstationary sensor time series, let

$$W_t = \{H_s^{local} : \max(1, t - M) \leq s < t\}, \quad (5)$$

denote the set of local scores available in the preceding window of length M . The adaptive thresholds are then defined as

$$H_t^- = Q_\alpha(W_t), H_t^+ = Q_{1-\alpha}(W_t), \quad (6)$$

where Q_α denotes the empirical quantile of order α and the anomaly decision is made by comparing H_t^{local} with H_t^- and H_t^+ . This formulation allows the detection procedure to account for gradual changes in the statistical properties of the signal.

By using neighbor distances in the feature space, the proposed entropy-based score is sensitive to local density changes and can therefore support the separation of structurally meaningful deviations from small random fluctuations. However, the final interpretation of detected points requires subsequent dynamical and predictive analyses. Thus, the use of KLE estimation provides a local density-sensitive basis for detecting candidate anomalous regimes and for their subsequent interpretation. Since anomalies in nonlinear sensor dynamics may be expressed not only as isolated deviations in feature space but also as changes in temporal evolution, the entropy-based detection step is complemented by phase-space reconstruction. This makes it possible to analyze anomalous observations in terms of local trajectory geometry.

Let m denote the embedding dimension and τ the time delay. For each admissible starting index p , the reconstructed phase vector is defined as

$$z_p = (x(t_p), x(t_p + \tau), \dots, x(t_p + (m-1)\tau)) \in \mathbb{R}^{md}, \quad (7)$$

where $p = 1, \dots, N - (m-1)\tau$.

If $z_{q_k(p)}$ denotes the k -th nearest neighbor of z_p in the reconstructed phase space, then the corresponding distance is

$$\rho_k(z_p) = \|z_p - z_{q_k(p)}\|_2 = \left(\sum_{r=0}^{m-1} \|x(t_p + r\tau) - x(t_{q_k(p)} + r\tau)\|_2^2 \right)^{\frac{1}{2}}, \quad (8)$$

where $\|\cdot\|_2$ denotes the Euclidean norm. Thus, $\rho_k(z_p)$ is the Euclidean distance between two-dimensional reconstructed phase vectors.

Embedding parameters are commonly selected using false-nearest-neighbor and mutual-information analyses. In the present case study, however, τ and m were not estimated using these criteria but were fixed before the final evaluation, as described in Section 7.2. This phase-space representation provides additional information about local trajectory geometry and supports the detection of anomalous regimes whose structure may not be visible in the original feature space alone. For reproducibility, all parameters used in the detection procedure must be fixed before validation. These include the number of nearest neighbors k , the window length M for adaptive thresholding, the quantile level α , the embedding dimension m , the delay parameter τ , and the distance metric used in both the original feature space and the reconstructed phase space. In the context of intelligent monitoring, the detected set of anomalous observations A should be interpreted as an alert set for further diagnostic analysis rather than as direct evidence of a specific fault or causal mechanism. The role of the entropy-based stage is therefore to localize candidate regime deviations, while their dynamical nature is assessed in the subsequent nonlinear and predictive analysis.

5. Nonlinear-dynamical diagnostics of anomalous episodes

After the entropy-based detection stage has localized the set of candidate anomalous observations A , the next step is to examine whether the detected deviations exhibit structurally meaningful dynamical properties. Since nonlinear dynamical indicators cannot be reliably interpreted for isolated points, the analysis is performed on anomalous episodes, that is, temporally localized groups of anomalous observations together with their neighboring trajectory context. The analysis of chaotic properties relies on methods from nonlinear dynamical systems theory. In the present framework, these methods are used not as independent anomaly detectors, but as diagnostic tools for assessing the dynamical complexity of the detected anomalous episodes. Lyapunov exponent analysis is widely used to characterize the sensitivity of trajectories to initial conditions [18, 19]. The Grassberger–Procaccia correlation dimension method is used as an auxiliary tool for assessing whether reconstructed trajectories exhibit a low-dimensional geometric structure [20, 21].

Let $E_j \subset \{1, \dots, N\}$ denote the index set of the j -th anomalous episode obtained from the detected set A . Using the reconstructed phase vectors z_p defined in the previous section, the phase-space representation of this episode is denoted by $Z_j = \{z_p : p \in E_j\}$.

The embedding dimension m , the delay parameter τ , and the distance metric are the same as those used in the phase-space reconstruction procedure described in the previous section. This ensures that the dynamical indicators estimated for different anomalous episodes are comparable. The

maximum Lyapunov exponent is used to quantify local sensitivity to initial conditions. In theoretical form, for the j -th anomalous episode it can be written as

$$\lambda_{max,j} = \lim_{s \rightarrow \infty} \frac{1}{s} \ln \frac{\|\delta Z_j(s)\|_2}{\|\delta Z_j(0)\|_2}, \quad (9)$$

where $\|\delta Z_j(0)\|_2$ is the initial distance between two close reconstructed trajectories in the phase-space representation of episode E_j , and $\|\delta Z_j(s)\|_2$ is the corresponding distance after s time steps. The distances are computed in the reconstructed phase space using the Euclidean metric introduced in the previous section.

Since the available sensor trajectories are finite, formula (9) is used as a theoretical reference, whereas in empirical validation the maximum Lyapunov exponent is estimated numerically over finite reconstructed trajectory fragments. A positive estimated value of $\lambda_{max,j}$ indicates local divergence of neighboring trajectories and may therefore be interpreted as evidence consistent with increased dynamical instability. However, for finite and noisy sensor data, such a value should not be regarded as a sufficient proof of deterministic chaos. The next step is to assess the geometric structure of anomalous episodes in the reconstructed phase space. For this purpose, the Grassberger–Procaccia correlation sum is computed [20,21]. Let N_j be the number of reconstructed phase-space points in Z_j . The empirical correlation sum for the j -th anomalous episode is defined as

$$C_j(r) = \frac{2}{N_j(N_j - 1)} \sum_{p=1}^{N_j} \sum_{q=p+1}^{N_j} \Theta(r - \|z_p - z_q\|_2), \quad (10)$$

where Θ is the Heaviside step function and $r > 0$ is the correlation radius. This value estimates the proportion of pairs of reconstructed points whose mutual distance does not exceed r .

The correlation dimension is then defined through the small-scale behavior of the correlation sum:

$$D_{2,j} = \lim_{r \rightarrow 0} \frac{\ln C_j(r)}{\ln r}. \quad (11)$$

In practical computations, $D_{2,j}$ is estimated on a finite scaling interval where the dependence between $C_j(r)$ and $\ln r$ is approximately linear. Because the anomalous analysis intervals considered in this study contain only 61–217 time steps, $\widehat{D}_{2,j}$ is not treated as a statistically consistent estimate of an attractor dimension. In particular, intervals containing approximately 60–70 reconstructed observations are generally insufficient to establish a stable asymptotic scaling law. The reported values should therefore be interpreted only as exploratory finite-scale slopes describing the geometry of the projected episode. Neither a positive finite-sample Lyapunov estimate nor a low finite-scale slope constitutes strict evidence of deterministic chaos.

The interpretation of the obtained indicators is performed in a diagnostic rather than categorical manner. If the estimated maximum Lyapunov exponent is positive and the correlation-dimension analysis indicates a stable low-dimensional scaling region, the corresponding anomalous episode may be regarded as structurally organized and dynamically unstable. If the Lyapunov estimate is close to zero and the correlation-dimension estimate is unstable or degenerate, there is no reliable evidence supporting deterministic chaotic structure. However, this alone does not prove that the episode is purely random.

The obtained indicators are interpreted jointly and qualitatively. An anomalous episode is regarded as exhibiting evidence consistent with local trajectory instability when its finite-sample Lyapunov estimate exceeds the corresponding reference range. The correlation-dimension estimate

is used only as an auxiliary geometric diagnostic and is considered informative only when a stable scaling interval with an adequate linear fit can be identified. If these conditions are not satisfied, the available data are treated as insufficient to support a low-dimensional deterministic interpretation. These indicators do not identify the physical cause of an anomaly and do not constitute proof of deterministic chaos. Thus, the joint use of Lyapunov-based sensitivity analysis and correlation-dimension estimation provides complementary qualitative information about the local dynamical behavior of anomalous episodes. These indicators supplement entropy-based anomaly detection but, given the finite and short episode lengths, should be interpreted as exploratory diagnostics rather than as evidence of deterministic chaos.

6. Attention-based analysis of anomalous episodes in time series

Attention mechanisms provide a data-driven tool for analyzing temporal dependencies in multivariate time series. In the proposed framework, they are used not as a primary anomaly detector, but as a predictive and interpretive component that complements entropy-based detection and nonlinear dynamical analysis. Attention-based models have been used to represent temporal and cross-sensor dependencies in industrial and multivariate anomaly-detection tasks [22–24].

Let z_p denote the reconstructed phase vectors introduced in the previous sections, where $p = 1, \dots, P$ and $P = N - (m - 1)\tau$. For each admissible index p , a local input window is formed as

$$S_p = (z_{p-w+1}, z_{p-w+2}, \dots, z_p), \quad (12)$$

where w is the window length.

The target value is defined as the next multivariate observation $y_p = x_{p+1}$, or, depending on the prediction setting, the next reconstructed state z_{p+1} . Thus, the attention model is trained to approximate the local evolution of the reconstructed trajectory.

For an input window S_p , multi-head attention is defined as

$$\text{MHA}(S_p) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O, \quad (13)$$

where

$$\text{head}_r = \text{softmax}\left(\frac{S_p W_r^Q (S_p W_r^K)^T}{\sqrt{d_k}}\right) S_p W_r^V, r = 1, \dots, h. \quad (14)$$

Here, W_r^Q, W_r^K, W_r^V are trainable projection matrices for the query, key, and value representations in the r -th attention head; W^O is the output projection matrix; h is the number of attention heads; and d_k is the key dimension. The softmax matrix in each head represents the relative predictive relevance of elements within the input window. To evaluate whether anomalous episodes are associated with increased predictive complexity, the model is trained in a one-step prediction setting. Let $f_\theta(S_p)$ denote the prediction produced by the attention model with parameters θ . The loss function is defined as

$$L(\theta) = \frac{\sum_{p \in I} \omega_p \|y_p - f_\theta(S_p)\|_2^2}{\sum_{p \in I} \omega_p}, \quad (15)$$

where I is the set of admissible training indices.

Normalization by the total weight ensures that changing β modifies the relative contribution of anomalous samples without mechanically changing the overall scale of the loss. The coefficient ω_p controls the contribution of observations associated with detected anomalous episodes:

$$\omega_p = 1 + (\beta - 1)\alpha_p, \beta \geq 1, \quad (16)$$

where $\alpha_p = 1$ if the target point belongs to the detected anomalous set A or to one of the anomalous episodes E_j , and $\alpha_p = 0$ otherwise. If the model is used in a purely unsupervised predictive-validation setting, $\beta = 1$, and all observations contribute equally to the loss. If the aim is to emphasize the predictive behavior of detected anomalous episodes, $\beta > 1$ increases their contribution.

The attention coefficients are analyzed as indicators of temporal predictive relevance. High attention weights between an anomalous episode and preceding trajectory fragments may indicate that the episode is linked to specific temporal contexts or transition patterns. However, these coefficients should not be interpreted as evidence of causal relations. They identify associative and predictive dependencies within the reconstructed time-series representation. In the present study, the attention model is used only as an exploratory predictive consistency check. It tests whether detector-defined anomalous intervals exhibit larger out-of-sample one-step prediction errors than the remaining trajectory. Attention coefficients are not treated as causal evidence, and no improvement in anomaly detection is attributed to the attention component unless a reproducible predictive separation between anomalous and non-anomalous intervals is observed. In the context of intelligent monitoring, such information can be used to distinguish isolated statistical deviations from anomalous regimes associated with changes in system behavior, increased predictive error, or possible degradation of the monitored process.

7. Model validation for anomaly detection and intelligent monitoring

7.1. Validation dataset and preprocessing

Experimental validation was performed using the Gas Sensor Array Low-Concentration dataset [25]. The dataset contains multivariate time series recorded from a gas sensor array consisting of ten sensors exposed to different volatile organic compounds under controlled laboratory conditions. Each experimental trial includes approximately 9000 time samples recorded at 1 Hz and represents a complete sensor-response cycle, including baseline stabilization, gas exposure, and recovery phases.

In this study, the validation procedure was conducted on the subset corresponding to ethanol exposure at a concentration of 100 ppb. This subset was selected because it provides controlled experimental conditions while preserving nonlinear temporal behavior, nonstationarity, and sufficiently complex multivariate sensor dynamics. The source file contains 90 recorded variables arranged as rows and approximately 9000 consecutive time samples arranged as columns. No additional sensor-derived features were constructed in this study. After transposing the source matrix, the data were represented as a multivariate time series ($X \in \mathbb{R}^{T \times d}$), where $T = 9002$ and $d = 90$. The 90 variables therefore correspond to the measurement channels provided in the analyzed data file rather than to features generated during preprocessing. The dataset was used to evaluate the proposed validation pipeline in three consecutive stages: entropy-based localization of candidate anomalous observations, dynamical analysis of the corresponding anomalous episodes, and attention-based predictive validation. Prior to analysis, the data were transformed into a numerical matrix, checked for missing or non-finite values, and normalized using a robust scaling procedure. This preprocessing was applied to reduce the influence of scale heterogeneity between sensor features while preserving local deviations relevant for anomaly detection.

7.2. Experimental protocol and parameter settings

The validation protocol was organized according to the three main stages of the proposed framework: entropy-based detection of candidate anomalous observations, dynamical analysis of anomalous episodes, and attention-based predictive validation. All parameters were fixed before the

final evaluation. The Kozachenko–Leonenko local entropy score H_t^{local} was computed with $k = 10$ nearest neighbors. Adaptive thresholds were estimated in a moving window of length $M = 500$ using the lower and upper empirical quantiles $q_{low} = 0.005$ and $q_{high} = 0.995$.

Detected anomalous points were further aggregated into anomalous episodes using a temporal merging rule. Points separated by fewer than $\text{min_gap} = 20$ time steps were assigned to the same episode, and each episode was extended by a symmetric context window of 30 time steps to preserve transition dynamics. For this exploratory case study, the delay and embedding dimension were fixed before the final evaluation at $\tau = 1$ and $m = 5$, respectively, rather than selected using an explicit mutual-information or false-nearest-neighbor criterion. The choice $\tau = 1$ retains the native one-sample temporal resolution, whereas $m = 5$ limits further dimensional expansion of the reconstructed vectors. These values are not claimed to be data-driven optima; consequently, the absence of a sensitivity analysis over alternative (τ, m) combinations is a limitation of the present validation. Before the finite-sample Lyapunov and correlation-sum calculations, the reconstructed vectors were projected onto the first 10 principal components to reduce numerical ill-conditioning and distance concentration in the high-dimensional representation. Ten components were retained because they explained 0.999998 of the total variance in the reconstructed data. This variance criterion does not imply preservation of attractor topology: as a linear projection, PCA may alter local neighborhoods and may therefore bias both trajectory-divergence and finite-scale dimension estimates. Consequently, the reported nonlinear-dynamical indicators characterize the projected representation and are interpreted only as exploratory diagnostics. The attention-based predictive validation used input windows of length $w = 20$, a model dimension $d_{model} = 64$, $h = 4$ attention heads, a batch size of 64, 15 training epochs, the Adam optimizer with a learning rate 10^{-3} , and an anomaly weighting coefficient $\beta = 5$. Such setting of parameters provides a unified and reproducible validation protocol for entropy-based detection, episode-level dynamical analysis, and attention-based predictive assessment of anomalous regimes in nonlinear sensor dynamics. A sensitivity analysis over several retained dimensions would be required to establish that the qualitative conclusions are stable with respect to the PCA projection. Such an analysis was not performed in the present study and remains a limitation.

7.3. Entropy-based localization of candidate anomalous observations

Entropy-based detection was applied to the test segment using the preprocessing procedure and parameter settings described in Sections 7.1–7.2. For each time point t , the local Kozachenko–Leonenko score H_t^{local} was computed from the distance $\rho_k(x_t)$ to the k -th nearest neighbor. To make the procedure consistent with an intelligent monitoring scenario, the neighbor search was performed only over previously available observations, including the reference training segment and the preceding part of the test segment. Thus, the computation of H_t^{local} did not use future observations from the test interval. A time point was assigned to the candidate anomalous set A if its local score crossed the adaptive threshold band defined in Section 7.2, that is, $H_t^{local} < H_t^- \vee H_t^{local} > H_t^+$. The resulting entropy trajectory, adaptive thresholds, and detected candidate anomalous observations are shown in Figure 1. Figure 1 shows that the local entropy score H_t^{local} crosses the adaptive threshold band at a limited number of time points. In total, 38 candidate anomalous observations were detected in the test segment containing 2701 time points, which corresponds to 1.41% of the segment. The detected observations are not uniformly distributed over time; instead, they appear as a small number of isolated threshold crossings and compact local groups. The exact temporal localization of these points is shown in Figure 1, while their aggregation into anomalous episodes is reported in the next subsection.

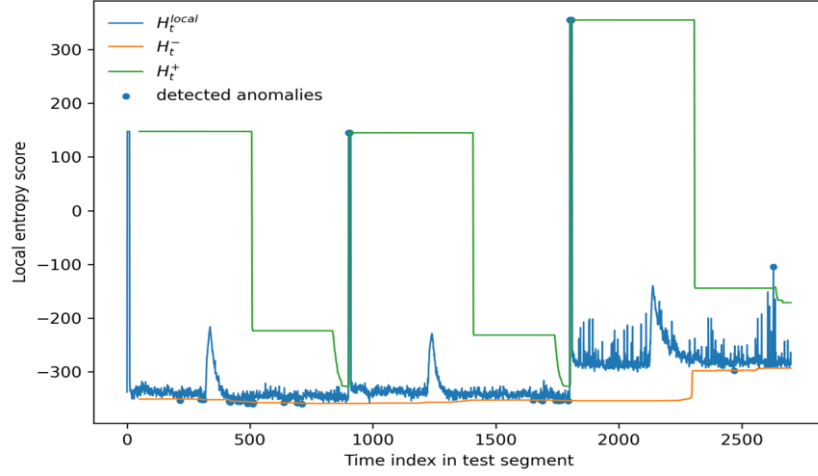


Figure 1: Entropy-based detection with adaptive moving-window thresholds

The step-like behavior of the upper threshold H_t^+ follows from its definition as an empirical rolling quantile rather than from an assumed discrete physical switching mechanism. With $M = 500$ and $q_{high} = 0.995$, the threshold is determined by only a few of the largest scores in the current window. A single extreme score may therefore keep the upper threshold at an elevated level while it remains inside the preceding window and may produce an abrupt change when it leaves that window. The plateaus and jumps in Figure 1 are thus finite-window order-statistic effects.

The entropy-based stage is therefore interpreted as an alert-generation procedure rather than as a final diagnostic classification. The detected set A localizes candidate deviations in the sensor dynamics, whereas their dynamical interpretation requires subsequent episode formation, Lyapunov-based sensitivity analysis, correlation-dimension estimation, and attention-based predictive validation.

7.4. Comparison with baseline anomaly detectors

Isolation Forest and Local Outlier Factor were evaluated using the same preprocessing, chronological split, and alarm budget as the proposed KLE-based detector (see Table 1).

Table 1

Comparison of anomaly detectors under an equal alarm budget

Method	Detected points	Overlap with KLE	Jaccard	Episodes	Matched KLE episodes
KLE	38	38	1.000	8	8/8
Isolation Forest	38	15	0.246	12	5/8
LOF	38	19	0.333	10	6/8

Each method selected 38 observations, corresponding to 1.41% of the test segment. Isolation Forest and LOF showed only moderate agreement with the proposed detector. More importantly, their detections were distributed across 12 and 10 episodes, respectively, whereas the KLE-based method formed eight more temporally concentrated episodes. The two largest KLE episodes contained 65.8% of all detected observations, compared with 34.2% for Isolation Forest and 44.7% for LOF. After nonlinear-dynamical validation, three of the eight KLE episodes exceeded the 95th percentile of the permutation-reference distribution, compared with one episode for each baseline. Two KLE episodes also had stable correlation-sum fits with $R^2 = 0.9$, whereas no Isolation Forest episode and only one

LOF episode satisfied both diagnostic conditions. Thus, under the same alarm budget, the proposed detector produced less fragmented and more diagnostically informative anomalous episodes. Because independent fault labels are unavailable, these results demonstrate an operational advantage within the proposed monitoring pipeline rather than higher accuracy.

7.5. Formation of anomalous episodes and analysis intervals

The candidate anomalous observations obtained at the entropy-based stage were aggregated into episode-level analysis intervals. The resulting intervals, constructed after temporal aggregation and contextual padding, are presented in Table 2.

Table 2

Anomalous analysis intervals obtained after temporal aggregation and contextual padding

Episode	Time interval	Number of time steps	Candidate anomalous observations
ep0	[186-246]	61	1
ep1	[270-341]	72	2
ep2	[386-545]	160	10
ep3	[608-742]	135	3
ep4	[871-936]	66	5
ep5	[1622-1838]	217	15
ep6	[2439-2499]	61	1
ep7	[2599-2659]	61	1

Table 2 shows that 38 candidate anomalous observations were grouped into eight anomalous analysis intervals. The distribution of anomalous observations across these intervals is strongly nonuniform. Two intervals, ep2 [386–545] and ep5 [1622–1838], contain 10 and 15 anomalous observations, respectively. Together, they account for 25 of the 38 detected candidates, that is, approximately 65.8% of all entropy-based detections. These intervals therefore represent the most pronounced candidates for subsequent dynamical interpretation. At the same time, several intervals contain only one or two anomalous observations. In particular, ep0, ep6, and ep7 each contain a single candidate anomaly, while their duration is 61 time steps due to the contextual padding procedure. These intervals should therefore be interpreted cautiously: they may indicate isolated local deviations, but they do not by themselves provide strong evidence of an extended anomalous regime. Thus, Table 2 provides an intermediate transition from pointwise entropy-based detection to episode-level validation. The most informative intervals for further Lyapunov-based and correlation-dimension analysis are those with both sufficient temporal length and a nontrivial concentration of detected candidates, primarily ep2 and ep5, while shorter or sparsely populated intervals serve as auxiliary cases for comparison.

7.6. Lyapunov-based analysis and surrogate comparison

The anomalous analysis intervals obtained in Section 7.4 were further examined using Lyapunov-based sensitivity analysis (see Figure 2). The purpose of this stage was not to prove deterministic chaos, but to evaluate whether particular anomalous intervals exhibit evidence consistent with local

trajectory instability. For each interval (E_j), the finite-sample estimate ($\hat{\lambda}_j^{real}$) was computed from the divergence of neighboring reconstructed trajectories.

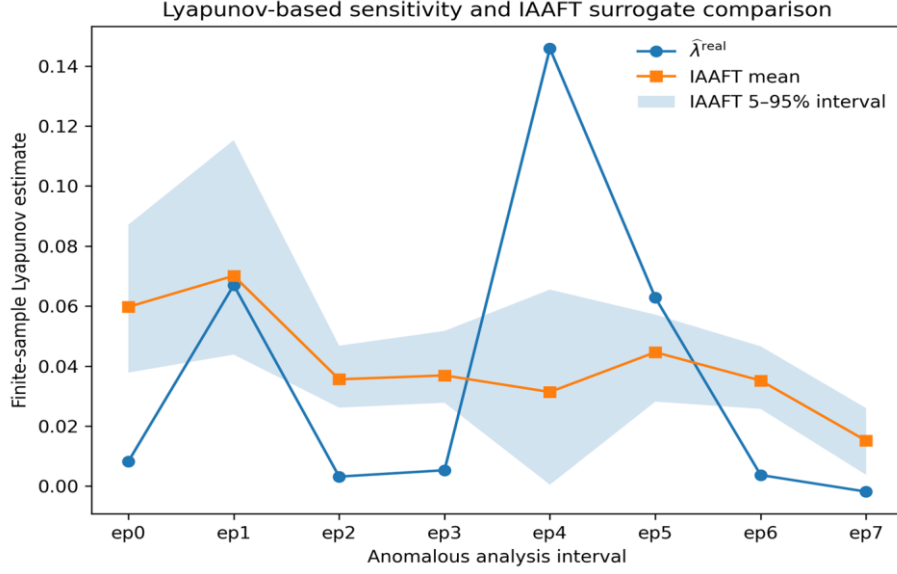


Figure 2: Lyapunov-based sensitivity and IAAFT surrogate comparison

To provide a reference level for comparison, a surrogate analysis was performed. For each anomalous interval, 50 feature-wise iterative amplitude-adjusted Fourier transform (IAAFT) surrogate realizations were generated. For each sensor-derived feature, the IAAFT procedure preserves the empirical marginal distribution and approximately preserves the Fourier amplitude spectrum, and therefore the corresponding linear autocorrelation structure, while disrupting the original phase organization. The resulting surrogate distribution provides a more restrictive reference model than random temporal permutation. However, because the procedure was applied separately to individual features, it does not explicitly preserve cross-feature phase and cross-spectral dependencies. The Lyapunov estimate was then recomputed for each surrogate realization, producing an empirical surrogate distribution for each (E_j). The comparison between the real estimates, surrogate means, and the 5th-95th percentile surrogate intervals is shown in Figure 2.

Figure 2 shows that only ep4 and ep5 exceed the upper 95% IAAFT reference percentile. For ep4, the real finite-sample Lyapunov estimate is 0.1458, compared with an IAAFT mean of approximately 0.031 and an upper percentile of approximately 0.065. For ep5, the corresponding values are 0.0628, approximately 0.045, and approximately 0.056. The estimate for ep1 remains within the IAAFT interval, whereas the remaining real estimates are below their corresponding reference ranges. Thus, ep4 and ep5 exhibit trajectory divergence not reproduced by the feature-wise IAAFT surrogates; however, this does not constitute evidence of deterministic chaos.

7.7. Correlation-dimension analysis

Correlation-dimension analysis was used as an auxiliary geometric verification stage for the anomalous intervals. Before estimation, the reconstructed phase vectors were projected onto the first 10 principal components, which preserved 0.999998 of the total variance. Given the interval lengths of 61–217 time steps, the values of $\hat{D}_{2,j}$ are treated as finite-scale characteristics of the projected trajectories rather than rigorous estimates of invariant attractor dimensions. As shown in Table 3, reliable linear scaling fits were obtained for seven of the eight intervals, with R^2 values ranging from 0.9421 to 0.9850. For ep4, the estimate $\hat{D}_{2,4} = 0.1325$ was accompanied by a poor fit, $R^2 = 0.3789$. This result was therefore classified as unstable and was not used as evidence of a low-dimensional geometric structure.

Table 3
Correlation-dimension estimates for anomalous analysis intervals

N_0	$\widehat{D}_{2,j}$	R^2	IAAFT mean	IAAFT 5–95% interval	Status
ep0	1.7650	0.9583	3.8921	[3.5600, 4.1778]	finite scaling estimate
ep1	1.3870	0.9421	1.6388	[1.2974, 2.0958]	finite scaling estimate
ep2	1.4552	0.9651	3.6915	[3.5281, 3.8537]	finite scaling estimate
ep3	1.6596	0.9578	3.7819	[3.6562, 3.8934]	finite scaling estimate
ep4	0.1325	0.3789	2.5047	[1.2235, 3.0203]	unstable scaling
ep5	1.3651	0.9440	0.0716	[0.0681, 0.0745]	finite scaling estimate
ep6	4.0165	0.9843	4.9178	[4.4226, 5.4426]	finite scaling estimate
ep7	3.1247	0.9850	3.5961	[3.0108, 4.2312]	finite scaling estimate

Comparison with the IAAFT (Table 3) surrogates revealed substantial heterogeneity among the anomalous intervals. For ep1 and ep7, the real estimates of 1.3870 and 3.1247 lie within the corresponding IAAFT intervals of [1.2974,2.0958] and [3.0108,4.2312]. For ep0, ep2, ep3, and ep6, the real estimates lie below their surrogate intervals, indicating a more compact finite-scale geometry than that generated by the adopted linear stochastic null model. These differences are diagnostically informative but do not, by themselves, justify a conclusion of deterministic chaos. The most consistent combined result was obtained for ep5. Its estimate, $\widehat{D}_{2,5} = 1.3651$ with $R^2 = 0.9440$, lies outside the IAAFT interval of [0.0681,0.0745], while its Lyapunov estimate also exceeds the upper IAAFT reference percentile. Thus, two complementary diagnostics indicate local temporal organization that is not reproduced by the adopted surrogate model. In contrast, although ep4 shows the largest Lyapunov-based deviation, its unstable correlation-dimension fit does not support a stronger geometric interpretation. Accordingly, the correlation-dimension stage performs its intended role within the proposed multilevel procedure. It does not duplicate primary anomaly detection; instead, it differentiates the detected episodes according to the reliability and character of their geometric structure. The method separates episodes consistent with the IAAFT model, episodes exhibiting finite-scale geometric departures, and cases for which no stable scaling relationship can be established. This demonstrates the diagnostic selectivity of the proposed approach, while the results are not interpreted as rigorous evidence of deterministic chaos.

7.8. Attention-based predictive validation and monitoring interpretation

Attention-based predictive validation was used as the final stage of the proposed procedure. Its purpose was not to construct the candidate anomaly set A , but to determine whether the detected intervals were associated with increased one-step prediction error. The sequences were split chronologically, with 70% used for training and 30% for validation. The validation set contained 803 sequences, including 122 from anomalous intervals and 681 from the remaining trajectory.

For $\beta = 1$, the overall RMSE was 50.2988, while the anomalous and non-anomalous RMSE values were 49.7867 and 50.3900. For $\beta = 5$, the corresponding values were 51.6714, 51.1522, and 51.7639.

Thus, anomalous intervals did not exhibit higher prediction error, and anomaly-based weighting did not improve predictive accuracy.

These results show that local density and dynamical changes do not necessarily reduce one-step predictability. The attention component therefore serves as an independent consistency check rather than a standalone anomaly criterion. In the current configuration, it did not confirm increased predictive complexity, but it helped distinguish structural anomaly from unpredictability.

8. Conclusion

A multilevel procedure for anomaly analysis in nonlinear multivariate sensor dynamics was proposed, combining entropy-based localization, temporal episode formation, nonlinear-dynamical diagnostics, and predictive validation. In a test segment containing 2701 observations, the local Kozachenko–Leonenko score detected 38 candidate anomalies (1.41%), which were aggregated into eight analysis intervals. Comparison with IAAFT surrogates showed that the Lyapunov estimates for ep4 and ep5 exceeded their corresponding upper 95% reference percentiles. For ep5, this result was supported by the finite-scale correlation-dimension estimate $\hat{D}_{2,5} = 1.3651$ with $R^2 = 0.9440$, whereas the poor fit for ep4, $R^2 = 0.3789$, did not support a stable scaling relationship. The procedure therefore distinguished episodes with consistent dynamical evidence, episodes with partial departures, and cases with unreliable estimates. The attention model did not identify increased prediction error for anomalous intervals. With $\beta = 1$, the RMSE values were 49.7867 for anomalous and 50.3900 for the remaining sequences; with $\beta = 5$, they were 51.1522 and 51.7639, respectively. This indicates that structural anomaly does not necessarily imply reduced one-step predictability. Overall, the proposed approach functions as a selective diagnostic procedure, while the reported results are not interpreted as rigorous evidence of deterministic chaos.

9. Directions for further research

Future work may develop the proposed framework in several complementary directions. One priority is to improve its suitability for complex sensor environments by incorporating models capable of capturing long range temporal relationships and interactions between multiple sensing channels, including recurrent architectures and graph based neural networks. Another important task is to reduce the dependence on manually specified settings by developing data driven procedures for choosing phase space reconstruction parameters and adapting entropy based decision thresholds. Further research should also examine methods for making attention based validation more transparent and easier to interpret in diagnostic applications. The framework should additionally be evaluated under more challenging operating conditions, including nonlinear feedback, gradual regime changes, and low amplitude anomaly patterns. Its extension to streaming scenarios with sequential model adaptation would support real time monitoring. Finally, broader experiments across different datasets, sensing technologies, and application domains are needed to determine the robustness and general applicability of the proposed approach.

Acknowledgements

The work was supported by the state budget research project “To develop evolutionary modeling methods for complex systems aimed at analyzing the dynamics of selforganized structures under conditions of uncertainty” (state registration number 0126U000358) of the V.M. Glushkov Institute of Cybernetics of the National Academy of Sciences (NAS) of Ukraine.

Declaration on Generative AI

During the preparation of this work, the authors used DeepL in order to translate research notes and results from Ukrainian to English. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

- [1] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. C. Aggarwal, M. Salehi, Deep learning for time series anomaly detection: A survey, *ACM Computing Surveys* 57 (1) (2024) 1–42. doi:10.1145/3691338.
- [2] Z. Tian, M. Zhuo, L. Liu, J. Chen, S. Zhou, Anomaly detection using spatial and temporal information in multivariate time series, *Scientific Reports* 13 (2023) 4400. doi:10.1038/s41598-023-31193-8.
- [3] A. R. Sahu, S. K. Palei, A. Mishra, Data-driven fault diagnosis approaches for industrial equipment: A review, *Expert Systems* 41 (2) (2024) e13360. doi:10.1111/exsy.13360.E.
- [4] Santos-Fernandez, J. M. Ver Hoef, E. E. Peterson, J. McGree, C. A. Villa, C. Leigh, R. Turner, C. Roberts, K. Mengersen, Unsupervised anomaly detection in spatio-temporal stream network sensor data, *Water Resources Research* 60 (11) (2024) e2023WR035707. doi:10.1029/2023WR035707.
- [5] L. Devroye, L. Györfi, On the consistency of the Kozachenko–Leonenko entropy estimate, *IEEE Transactions on Information Theory* 68 (2) (2022) 1178–1185. doi:10.1109/TIT.2021.3127938.
- [6] A. Krakovská, Š. Pócoš, K. Mojžišová, I. Bečková, J. X. Gubáš, State space reconstruction techniques and the accuracy of prediction, *Communications in Nonlinear Science and Numerical Simulation* 111 (2022) 106422. doi:10.1016/j.cnsns.2022.106422.
- [7] I. Weber, C. R. Oehr, NoLiTiA: An open-source toolbox for non-linear time series analysis, *Frontiers in Neuroinformatics* 16 (2022) 876012. doi:10.3389/fninf.2022.876012.
- [8] L. Escot, J. E. Sandubete, Estimating Lyapunov exponents on a noisy environment by global and local Jacobian indirect algorithms, *Applied Mathematics and Computation* 436 (2023) 127498. doi:10.1016/j.amc.2022.127498.
- [9] S. Zhou, X. Wang, W. Zhou, C. Zhang, Recognition of the scale-free interval for calculating the correlation dimension using machine learning from chaotic time series, *Physica A: Statistical Mechanics and its Applications* 588 (2022) 126563. doi:10.1016/j.physa.2021.126563.
- [10] W. Gilpin, Generative learning for nonlinear dynamics, *Nature Reviews Physics* 6 (2024) 194–206. doi:10.1038/s42254-024-00688-2.
- [11] H. Kang, P. Kang, Transformer-based multivariate time series anomaly detection using inter-variable attention mechanism, *Knowledge-Based Systems* 290 (2024) 111507. doi:10.1016/j.knosys.2024.111507.
- [12] Y. Liu, H. Li, Y. Guo, C. Kong, J. Li, S. Wang, Rethinking attention-model explainability through faithfulness violation test, in: *Proceedings of the 39th International Conference on Machine Learning*, vol. 162, 2022, pp. 13807–13824.

- [13] G. Li, J. J. Jung, Dynamic relationship identification for abnormality detection on financial time series, *Pattern Recognition Letters* 145 (2021) 194–199. doi:10.1016/j.patrec.2021.02.004.
- [14] C. Wang, G. Liu, From anomaly detection to classification with graph attention and transformer for multivariate time series, *Advanced Engineering Informatics* 60 (2024) 102357. doi:10.1016/j.aei.2024.102357.
- [15] Q. Yang, M. T. J. Spaan, CEM: Constrained entropy maximization for task-agnostic safe exploration, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37 (9), 2023, pp. 10798–10806. doi:10.1609/aaai.v37i9.26281.
- [16] Z. Ao, J. Li, Entropy estimation via normalizing flow, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36 (9), 2022, pp. 9990–9998. doi:10.1609/aaai.v36i9.21237.
- [17] W. Yu, Q. Huang, A deep encoder-decoder network for anomaly detection in driving trajectory behavior under spatio-temporal context, *International Journal of Applied Earth Observation and Geoinformation* 115 (2022) 103115. doi:10.1016/j.jag.2022.103115.
- [18] J.-Y. Lee, Condition monitoring of impact vibrations using observed signals and estimated principal Lyapunov exponents, *Journal of Vibration and Control* 31 (11–12) (2025) 2439–2449. doi:10.1177/10775463241259131.
- [19] Z. R. Shu, P. W. Chan, Q. S. Li, X. H. He, C. Z. Cai, Characterising the variability in ocean data using fractal and correlation dimension analysis, *Applied Ocean Research* 138 (2023) 103590. doi:10.1016/j.apor.2023.103590.
- [20] A. Krakovská, M. Chvosteková, Simple correlation dimension estimator and its use to detect causality, *Chaos, Solitons & Fractals* 175 (2023) 113975. doi:10.1016/j.chaos.2023.113975.
- [21] L. S. Yousuf, Nonlinear dynamics investigation of contact force in a cam–follower system using the Lyapunov exponent parameter, power spectrum analysis, and Poincaré maps, *International Journal of Mechanical System Dynamics* 2 (2) (2022) 214–230. doi:10.1002/msd2.12025.
- [22] M. Kim, S. Park, Unsupervised multi-head attention autoencoder for multivariate time-series anomaly detection, in: *2024 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2024, pp. 1–7. doi:10.1109/BigComp60711.2024.00011.
- [23] D. Symonov, O. Palagin, Y. Symonov, B. Zaika, Optimisation of training samples with KLE and mutual information, in: *Proceedings of the Eighth International Workshop on Computer Modeling and Intelligent Systems (CMIS 2025)*, vol. 3988, CEUR-WS.org, 2025, pp. 296–306. URL: <https://ceur-ws.org/Vol-3988/paper23.pdf>.
- [24] X. Huang, N. Chen, Z. Deng, S. Huang, Multivariate time series anomaly detection via dynamic graph attention network and Informer, *Applied Intelligence* 54 (17–18) (2024) 7636–7658. doi:10.1007/s10489-024-05575-y.
- [25] F. Tian, L. Zhao, S. Deng, Gas sensor array low-concentration, UCI Machine Learning Repository, 2023. doi:10.24432/C5CK6F.