ANOMALY DETECTION IN ELI-NP FRONT-END LASER ENERGY DATA USING AN OPTIMIZED MOVING AVERAGE METHOD

T. IMRAN

Extreme Light Infrastructure-Nuclear Physics (ELI-NP), "Horia Hulubei" National R&D Institute for Physics and Nuclear Engineering (IFIN-HH), Magurele, Jud. Ilfov, Romania
Corresponding author: tayyab.imran@eli-np.ro

Received February 17, 2025

Abstract. Anomaly detection in time-series data is critical for ensuring stability in high-power laser systems, where deviations can indicate potential failures. This study optimizes a moving average-based methodology for anomaly detection accuracy by evaluating window sizes (3, 6, 9, 12, and 15) and threshold multipliers (1.0, 1.5, and 2.0). The analysis integrates Mean Squared Error (MSE), correlation analysis, and graphical evaluations, including anomaly distribution, moving average trends, and parameter sensitivity plots. Results indicate that smaller window sizes effectively detect short-term fluctuations but are more susceptible to noise, while larger windows smooth trends but may overlook minor anomalies. Threshold multipliers significantly impact detection, with lower values capturing more anomalies, potentially increasing false positives, and higher values reducing sensitivity but minimizing false alarms. MSE trends suggest a trade-off between sensitivity and robustness, where smaller windows better fit the data but risk overfitting noise, while larger windows reduce responsiveness but enhance stability. Correlation analysis scatterplots reveal a strong dependency between window size and MSE, while anomaly counts exhibit a nonlinear relationship with threshold multipliers. Anomaly detection plots and MSE *vs.* window size comparisons highlight detection efficiency. The study bridges statistical anomaly detection techniques with real-world laser monitoring, ensuring computational efficiency, robustness, and enhanced fault detection. These findings lay the groundwork for adaptive parameter tuning and machine learning integration in real-time anomaly detection for high-power laser systems.

Key words: anomaly detection, moving average, time-series analysis, machine learning, ELI-NP laser system.

DOI: https://doi.org/10.59277/RomJPhys.2025.70.902

1. INTRODUCTION

High-power laser systems, such as those used in scientific research facilities, operate under extreme conditions where even small deviations in parameters like pulse energy and beam alignment can cause significant issues [1]. The ability to detect such anomalies promptly is essential for maintaining system reliability, optimizing performance, and minimizing downtime. Traditional anomaly detection techniques include statistical models and clustering-based methods. Conventional machine learning [2] models often require large amounts of labeled data and significant computational power. Statistical methods in machine learning, like moving averages [3], are computationally efficient and interpretable, making them ideal for real-time applications. This study focuses on developing and optimizing a moving average-based anomaly detection method tailored to laser energy data.

In the context of related work, statistical methods such as z-score [9] and standard deviation-based [4] thresholding have been widely used for anomaly detection due to their simplicity. However, these methods often lack adaptability, as fixed thresholds may not account for dynamic changes in the data. Clustering-based techniques like k-means and density-based methods [10] offer better adaptability but require significant computational resources and are sensitive to parameter selection. Machine learning models, including Isolation Forests [9] and deep learning approaches like Long Short-Term Memory (LSTM) networks, have shown promise in recent studies for anomaly detection in complex datasets. While these methods offer high accuracy and adaptability, they often require large, labeled datasets for training, substantial computational resources, and careful tuning of hyperparameters.

Anomaly detection in time-series data [5, 6] is crucial, particularly in high-power laser systems where even minor deviations in operational parameters can cause significant performance issues or equipment damage. Maintaining stable operation is essential for optimal performance in scientific research and industrial applications of laser systems. Data-driven methods that automatically detect real-time anomalies can prevent costly downtime, improve system reliability, and enhance overall safety. This study presents an optimized moving average method for anomaly detection in laser energy data. The data used in this study is acquired from the front-end of a high-power ELI-NP laser facility [7, 8]. The front-end is equipped with a data storage and acquisition system, acquiring pulse energy data over time. Unlike traditional anomaly detection methods, which rely on fixed statistical thresholds or computationally intensive machine learning models [9–11], the proposed method innovatively bridges simplicity and accuracy by optimizing key parameters such as the window size and threshold multiplier [12]. The approach proposed in this work combines the interpretability and computational efficiency of statistical methods with the adaptability of machine learning by optimizing key parameters to achieve a balance between sensitivity and robustness. The novelty lies in the systematic optimization of the window size and threshold multiplier, often chosen arbitrarily in conventional moving average-based methods. By minimizing the Mean Squared Error (MSE) [10] and fine-tuning the threshold multiplier based on anomaly count metrics, this study may achieve improved performance without the need for complex models or extensive training data.

2. METHODOLOGY, DATA ANALYSIS, AND OBSERVATIONS

The methodology involved several key steps illustrated in the flowchart Fig. 1, starting with data preprocessing if needed. The raw dataset consisted of timestamped numerical values representing critical laser parameters. Preprocessing included converting timestamps to DateTime format, ensuring consistent intervals between data points, and interpolating missing values to maintain data continuity. Outliers are retained intentionally, as they are the primary focus of anomaly detection.

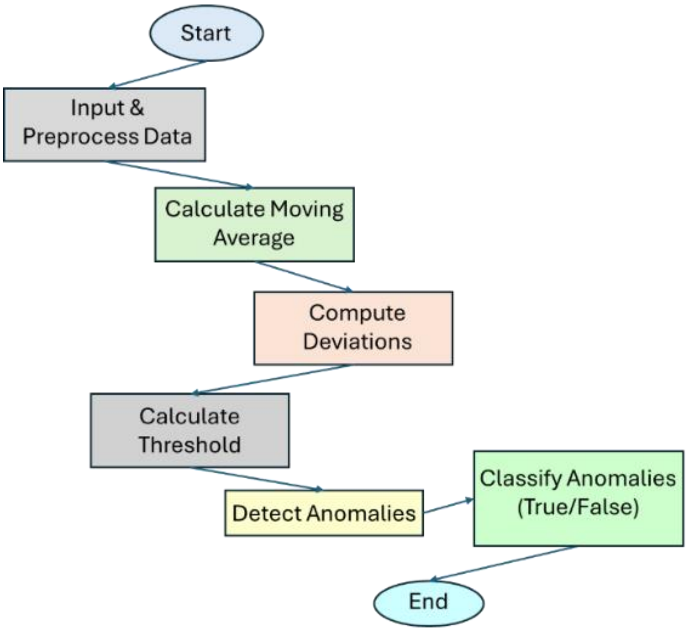

Fig. 1 – Flowchart of the moving average anomaly detection method, key steps: data input and preprocessing, moving average calculation, deviation analysis, threshold determination, anomaly detection, classification, and conclusion.

A sample of the preprocessed data is shown in Table 1, which lists five consecutive data points, highlighting the uniformity in time intervals and the corresponding values of the laser parameter being monitored. Preprocessed data is plotted in Fig. 2, which comprehensively analyzes energy values over time, combining statistical insights and temporal trends [11]. The main plot on the left (Fig. 2a) shows the energy values (in joules) as a function of time in hours. The blue line represents the measured energy data, while the red dashed line indicates the average energy value of approximately 0.008943 J. Fluctuations in energy around the mean are visible, with periods of stability interrupted by deviations both above and below the average. On the right (Fig. 2b) presents a histogram of energy values, illustrating the frequency distribution of these measurements [12]. The histogram reveals a roughly normal distribution, with most energy values concentrated near 0.0089 J, reflecting a strong central tendency and minimal dispersion. Together, these plots provide a detailed statistical and temporal depiction of the energy data, showcasing its stability, variability, and overall distribution.

Table 1

A sample of preprocessed data, timestamp (H) and corresponding energy (J) measurements

Timestamp (H)	Energy (J)
00:00:00	0.009065
00:01:00	0.009114
00:02:00	0.009030
00:03:00	0.009156
00:04:00	0.009002
:	:

Once preprocessing is completed, the moving average is applied to smooth the data and establish a baseline for detecting deviations. The moving average for each time t is calculated using equation (1):

$$MA_t = \frac{1}{N}\sum_{i=0}^{N-1} X_{t-i} \tag{1}$$

N represents the window size, and X_{t-i} denotes the data values within the window preceding time t. Initially, a window size of 3 is chosen, providing a reasonable smoothing level by averaging over three consecutive points. Following calculating the moving average, the next step involved computing the difference between the original data and the moving average to quantify deviations from the trend. This difference is calculated using the formula in equation (2):

$$\text{Difference}_t = X_t - \text{MA}_t \tag{2}$$

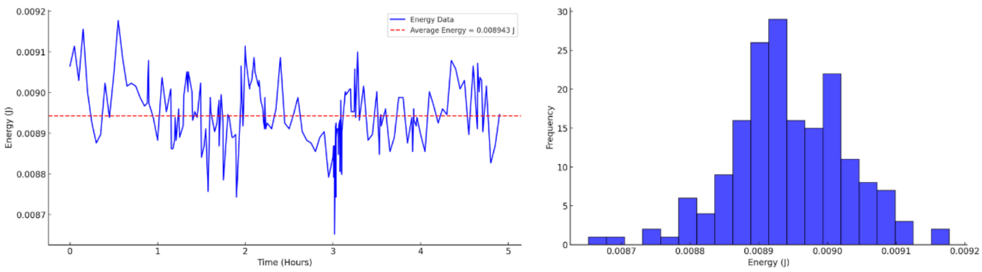

Fig. 2 – a) Energy *vs.* time plot showing fluctuations around the average energy (red dashed line); b) histogram depicting a near-normal distribution of energy values centered around the mean.

The equation (2) represents how much each data point deviates from the expected behavior. For example, at the 3rd data point (N = 3), the difference is computed as Difference_3 = 0.009030 – 0.009070 = − 0.000040. These differences served as the basis for identifying potential anomalies, with larger deviations indicating a higher likelihood of being an anomaly. A similar procedure is followed for other window sizes, *i.e.*, 6, 9, 12, and 15.

Table 2

Window size, moving average, and differences

Window Size (N)	Current (X_t)	Moving Average (MA_t)	Difference ($X_t - \mathrm{MA}_t$)
3	0.00911	0.0091333	-2.333×10^{-5}
6	0.00911	0.0091366	-2.666×10^{-5}
9	0.00911	0.0091433	-3.333×10^{-5}
12	0.00911	0.0091373	-2.733×10^{-5}
15	0.00911	0.0091238	-1.379×10^{-5}

A statistical threshold is applied to determine whether a deviation is significant enough to be classified as an anomaly. The threshold is defined as Threshold = $k \times \sigma$, where k is the threshold multiplier, and σ is the standard deviation of the differences. The standard deviation is computed (equation (3)):

$$\sigma = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\mathrm{Difference}_1 - \overline{\mathrm{Difference}})^2}, \qquad (3)$$

where $\overline{\mathrm{Difference}}$ represents the mean of the differences. Any data point with an absolute difference exceeding the threshold is flagged as an anomaly. To identify anomalies, $T_{\mathrm{upper}} = \mathrm{MA}_t + k\sigma$, $T_{\mathrm{lower}} = \mathrm{MA}_t - k\sigma$, and data is flagged as an anomaly if $X_t > T_{\mathrm{upper}}$ or $X_t < T_{\mathrm{lower}}$.

2.1. THRESHOLD MULTIPLIERS 1, 1.5, AND 2

The detailed analysis of anomaly counts across different combinations of window sizes (N) and threshold multipliers (Th) in Fig. 3 provides a clear understanding of how these parameters influence anomaly detection in the dataset. The observed anomaly counts highlight specific patterns and relationships that can be interpreted to optimize the detection process. For $N = 3$, the number of anomalies detected at Th = 1.0 is 20, which drops drastically to 1 at Th = 1.5 and further to 0 at Th = 2.0. This significant reduction reflects how the algorithm, with a small window size, is highly sensitive to small deviations when the threshold multiplier is low but becomes extremely conservative at a higher threshold like Th = 2.0, essentially flagging no deviations as anomalies. This behavior highlights the algorithm's increased tolerance to noise and minor fluctuations as the threshold multiplier rises. For $N = 6$, the anomaly count is 24 at Th = 1.0, reducing to 6 at Th = 1.5 and dropping further to 2 at Th = 2.0. Compared to $N = 3$, the moving average for $N = 6$ smooths out short-term fluctuations more effectively, resulting in fewer anomalies even at lower thresholds. At Th = 2.0, only the most significant deviations are flagged, demonstrating the algorithm's ability to focus solely on substantial anomalies. Similarly, for N = 9, the anomaly count is 22 at Th = 1.0, decreasing to 4 at Th = 1.5 and dropping to 1 at Th = 2.0. This trend further reinforces the observation that larger window sizes combined with higher thresholds emphasize longer-term deviations while ignoring minor, short-term fluctuations. As we move to larger window sizes, such as N = 12 and N = 15, the impact of smoothing becomes even more pronounced. For N = 12, the number of anomalies detected at Th = 1.0 is 19, which reduces to 12 at Th = 1.5 and drops to 8 at Th = 2.0. The smoothing effect of the larger window size ensures that only substantial and sustained deviations are flagged as anomalies, making it suitable for datasets with less noise or when capturing significant changes, which is the primary objective. Similarly, for N = 15, the anomaly counts are even lower, with 15 anomalies detected at Th = 1.0, 10 at Th = 1.5, and only 6 at Th = 2.0. This further reduction highlights the algorithm's robustness at higher window sizes, as it becomes less sensitive to transient fluctuations and better at identifying major deviations.

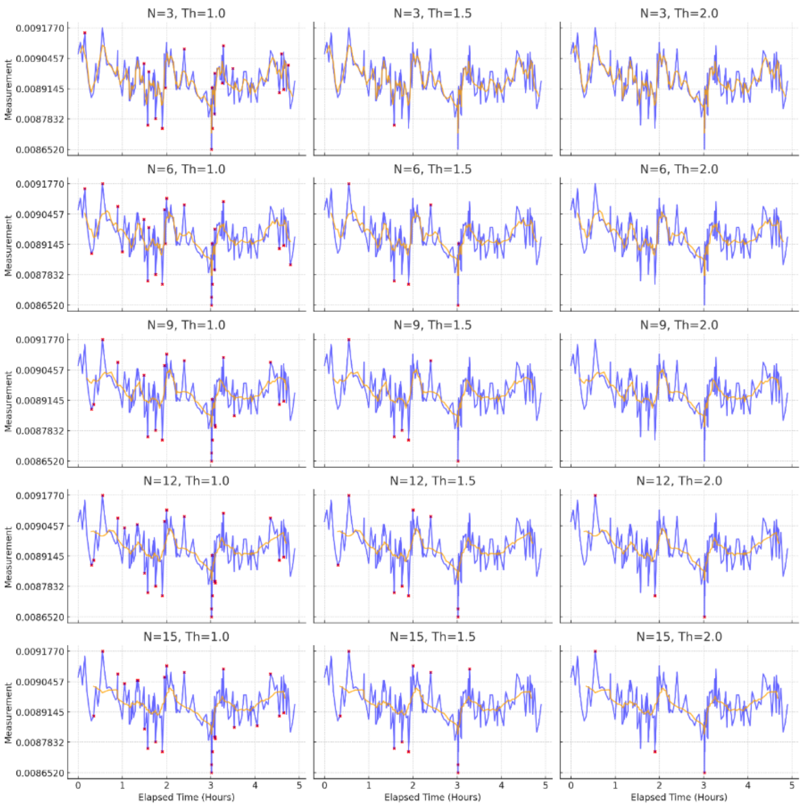

Fig. 3 – Anomaly detection plots using varying window sizes (N) and thresholds, x-axis represents elapsed time in hours, orange line shows the averaging and red points indicate detected anomalies.

The sensitivity to window size plays a significant role in these observations. Smaller window sizes allow the moving average to react quickly to data changes, making the algorithm more noise-sensitive. This explains the higher anomaly counts at smaller window sizes, particularly when combined with a lower threshold multiplier. Conversely, larger window sizes provide a smoother average, filtering out minor fluctuations and reducing anomalies. Similarly, the threshold multiplier determines the tolerance level for deviations from the moving average. Lower thresholds lead to higher sensitivity, capturing even minor deviations, while higher thresholds reduce sensitivity, focusing only on major anomalies. The interaction between window size and threshold multiplier highlights the trade-off between detecting subtle anomalies and avoiding false positives. Smaller window size and threshold multiplier combinations are ideal for environments where capturing every potential deviation is essential, but they may produce more false positives. Larger window size and threshold multiplier combinations, on the other hand, are suitable for scenarios where stability and robustness are prioritized.

The interplay between window size and threshold multiplier determines the effectiveness of the anomaly detection process. Smaller window sizes and thresholds are suited for sensitive environments where capturing every deviation is critical, whereas larger values are ideal for applications prioritizing stability and noise tolerance.

2.2. MEAN SQUARE ERROR (MSE)

As the window size is varied (3, 6, 9, 12, 15), the corresponding MSE (equation (4)) is computed for each size to evaluate how well the moving average fits the original data [10]

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(X_t - MA_t\right)^2 \tag{4}$$

A window size of 3 yielded the lowest MSE of 2.406×10^{-9}, indicating that it provided the best balance between capturing the underlying trend and minimizing noise. Similarly, different values of the threshold multipliers (1.0, 1.5, 2.0) are tested, and the multiplier that minimizes false positives while maintaining sensitivity to true anomalies is considered. Table 3 highlights the impact of window size and threshold multiplier on anomaly detection and MSE. Smaller windows (N = 3) detect more anomalies (*e.g.*, 20 at Th = 1.0) due to their sensitivity to fluctuations, resulting in a lower MSE (2.47×10^{-9}), as the moving average closely follows the data. As window size increases (N = 6, 9, 12, 15), anomalies decrease due to smoothing, but MSE slightly rises (*e.g.*, 3.22×10^{-9} for W = 6) as responsiveness declines. Lower thresholds (Th = 1.0) detect more anomalies and reduce MSE, while higher thresholds (Th = 2.0) filter minor deviations, reducing anomalies but increasing MSE, optimizing window size and threshold multiplier balances sensitivity and robustness, ensuring accurate and reliable real-time anomaly detection in laser systems.

Table 3

Impact of window size and threshold multiplier on MSE and anomaly count

Window Size (N)	Threshold Multiplier (Th)	Anomaly Counts	Mean Square Error (MSE)
3	1	20	2.47×10^{-9}
3	1.5	1	2.47×10^{-9}
3	2	0	2.47×10^{-9}
6	1	24	3.23×10^{-9}
6	1.5	6	3.23×10^{-9}
6	2	0	3.23×10^{-9}
9	1	22	3.43×10^{-9}
9	1.5	6	3.43×10^{-9}
9	2	0	3.43×10^{-9}
12	1	24	3.80×10^{-9}
12	1.5	9	3.80×10^{-9}
12	2	3	3.80×10^{-9}
15	1	25	4.02×10^{-9}
15	1.5	10	4.02×10^{-9}
15	2	3	4.02×10^{-9}

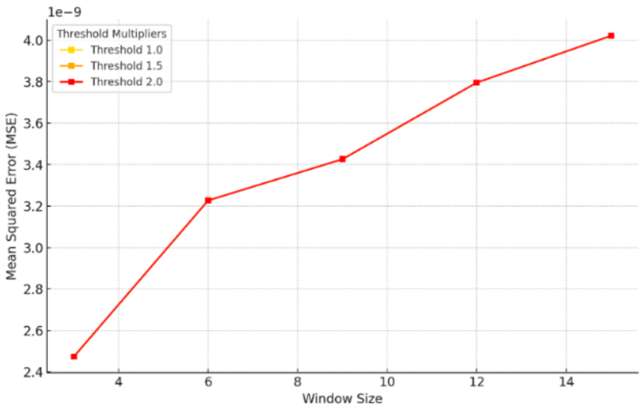

Fig. 4 – Mean Squared Error (MSE) *vs.* window size for thresholds 1.0 (yellow), 1.5 (orange), and 2.0 (red), showing how MSE increases with larger window sizes and smoother data.

The plot in Fig. 4 illustrates the relationship between MSE and window size (N) for three threshold multipliers (Th = 1.0, 1.5, 2.0). The MSE, which measures the average squared differences between the original data and the moving average, increases as the window size grows. For Th = 1.0, 1.5, and 2.0, the MSE starts at

2.47×10^{-9} for $N = 3$ and increases to 3.23×10^{-9} at $N = 6$, 3.89×10^{-9} at $N = 9$, 4.26×10^{-9} at $N = 12$, and 4.58×10^{-9} at $N = 15$. The identical MSE values across all thresholds result in overlapping lines, with only the red line (Th = 2.0) visible. This overlap highlights that MSE depends solely on window size, as the threshold multiplier does not influence the deviation between the moving average and the original data.

2.3. NUMBER OF ANOMALIES *VS.* THRESHOLD

Figure 5 shows the relationship between the number of anomalies and window size (N) for three threshold multipliers (Th = 1.0, 1.5, 2.0). For Th = 1.0 (blue line), the algorithm detects 20 anomalies at window size $N = 3$, increasing to 25 at $N = 6$, fluctuating slightly at $N = 9$ with 22 anomalies, and returning to 25 at $N = 12$, 15. The lower threshold captures even minor deviations, leading to higher sensitivity. For Th = 1.5 (green line), anomalies increase steadily from 1 at $N = 3$ to 12 at $N = 15$, reflecting a focus on significant deviations. In the case of Th = 2.0 (red line), the algorithm detects no anomalies at $N = 3$, 6, then consistently flags one anomaly from $N = 9$ onwards, emphasizing robustness by filtering minor fluctuations.

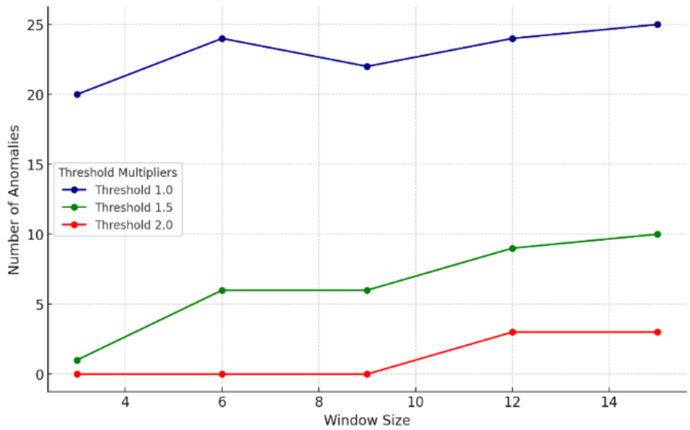

Fig. 5 – Number of anomalies detected *vs.* window size for thresholds 1.0 (blue), 1.5 (green), and 2.0 (red), showing sensitivity variations with parameter changes.

These plots (Fig. 5) highlight how thresholds affect sensitivity, with lower thresholds detecting more anomalies while higher thresholds reduce false positives by focusing on major deviations. Increasing window size generally smooths short-term variations, decreasing anomaly counts for higher thresholds while emphasizing long-term trends. Lower thresholds and smaller windows are suitable for highly sensitive applications like laser diagnostics, while higher thresholds and larger windows prioritize stability and robustness, ensuring efficient and reliable real-time monitoring in laser systems.

2.4. CORRELATION

The scatterplot matrix in Fig. 6 visualizes the pairwise relationships among the variables, including window size, threshold, anomalies, and MSE.

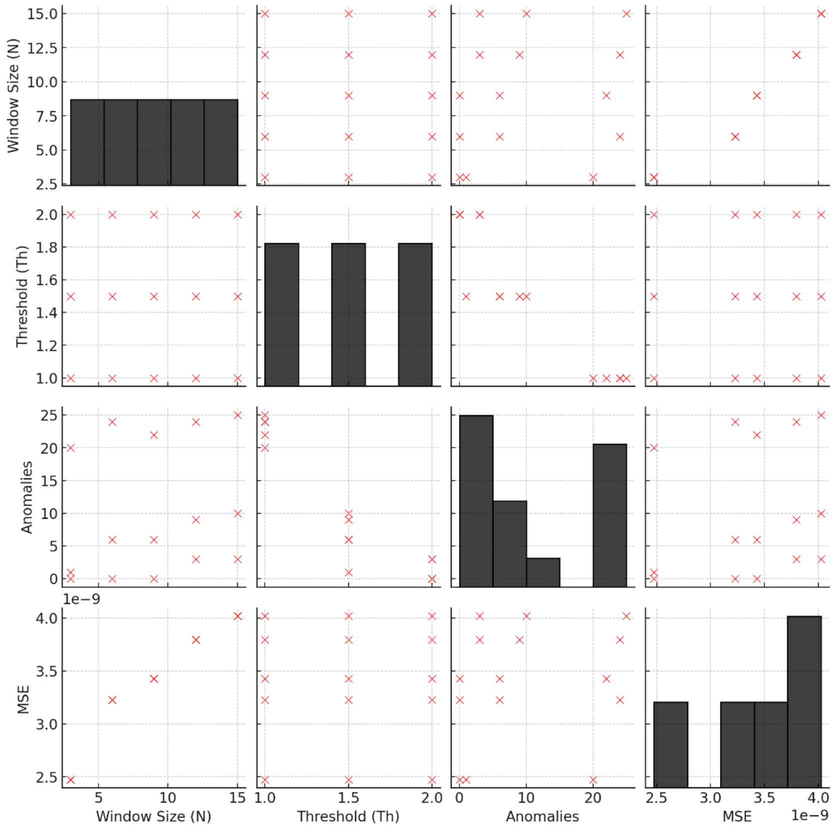

Fig. 6 – Scatterplot matrix showing pairwise relationships among Window Size (N), Threshold (Th), Anomalies, and MSE, with histograms on the diagonal and red crosses indicating scatter points.

The diagonal plots show histograms of individual variables, revealing their distributions. For example, window size displays a uniform distribution, while anomalies and MSE exhibit skewed distributions, indicating variability in their behavior. The off-diagonal scatterplots highlight relationships between variable pairs. A strong positive relationship is evident between window size and MSE, as seen in the upward trend of the scatterplot, indicating that larger window sizes are associated with higher mean squared errors. Conversely, a negative relationship is observed between thresholds and anomalies, where higher thresholds result in fewer detected anomalies. Some variable pairs, such as threshold and window size, show no significant relationship, as indicated by the random scatter of points. This visualization provides a comprehensive overview of the dataset's trends, correlations, and patterns.

3. DISCUSSION AND INSIGHTS

This study presents an optimized moving average-based anomaly detection method for front-end laser energy data, balancing sensitivity, robustness, and computational efficiency. Smaller window sizes effectively capture short-term fluctuations but increase false positives, while larger windows improve stability by filtering noise but risk missing transient anomalies. Choosing an optimal window size requires balancing detection sensitivity with noise suppression. Threshold multipliers significantly impact detection accuracy. Lower values (1.0) capture more anomalies but introduce false positives, whereas higher values (1.5, 2.0) reduce false alarms but may miss subtle deviations. The relationship between threshold multipliers and anomaly counts is nonlinear, reinforcing the need for careful parameter tuning. Correlation analysis confirms that smaller window sizes lead to lower MSE but may overfit noise, while larger windows increase MSE due to reduced responsiveness to short-term variations. The study highlights that tuning detection parameters is crucial for achieving a balance between precision and robustness. The findings suggest smaller window sizes work well for rapid deviation detection but require threshold adjustments to limit false positives. Conversely, larger window sizes reduce noise sensitivity but may overlook critical anomalies, making them suitable for applications prioritizing long-term stability. The proposed method bridges statistical anomaly detection with practical laser diagnostics, offering a scalable and efficient real-time monitoring solution. It ensures computational efficiency while enhancing fault detection in high-power laser systems. Future work can explore automated parameter selection and adaptive learning approaches to improve system reliability, optimize anomaly detection, and adapt to dynamic laser environments.

Acknowledgements. Work was funded by Romanian Ministry for Research, Innovation and Digitalization through ELI-NP IOSIN program, and through Nucleu program PN 23 21 01 05, through the Extreme Light Infrastructure – Nuclear Physics (ELI-NP) Phase II, a project co-financed by the Romanian Government and the European Union through the European Regional Development Fund and the Competitiveness Operational Programme (1/07.07.2016, COP, ID 1334).

REFERENCES

1. J. Yoon *et al.*, Optica **8**, 630 (2021).
2. G. Carleo *et al.*, Rev. Mod. Phys. **91**, 045002 (2019).
3. M. Zubair *et al.*, Annals of Data Science **11**, 1525 (2024).
4. K. Przystupa *et al.*, Metrol. Meas. Syst. **30**, 17 (2023).
5. H. Xu *et al.*, IEEE Trans. Knowl. Data Eng. **35**, 12591 (2023).
6. A. Ikotun *et al.*, J, Inf. Sci. **622**, 178 (2023).
7. K. A. Tanaka, K. M. Spohr, D. L. Balabanski, Matter Radiat. Extrem. **5**, 024402 (2020).
8. C. Radier, O. Chalus, M. Charbonneau, High Power Laser Sci. Eng. **10**, e21 (2022).
9. I. Altman *et al.*, J. Int. Financ. Manag. Account. **28**, 131 (2017).
10. O. Hodson, Geosci. Model Dev. Discuss. **15**, 5481 (2022).
11. F. Taneski, T. Abbas R. Henderson, Light. Technol. **40**, 5884 (2022).
12. X. Yang, Y. Ni, Communications in Statistics-Theory and Methods **50**, 4134 (2021).