COMPREHENSIVE ANALYSIS OF LASER POWER STABILITY USING STATISTICAL AND MACHINE LEARNING MODELS

TAYYAB IMRAN[1], MUDDASIR NAEEM[2]

[1]Extreme Light Infrastructure – Nuclear Physics, "Horia Hulubei" National Institute for R&D in Physics and Nuclear Engineering, 30 Reactorului Street, 077125 Magurele, Romania
E-mail: tayyabmrn@gmail.com
[2]Department of Physics, Syed Babar Ali School of Science and Engineering, Lahore University of Management Sciences (LUMS), Lahore, Pakistan
E-mail: muddasirnaeem98@gmail.com

Received April 10, 2025

Abstract. This study explores a comprehensive approach to analyzing laser power stability by combining statistical evaluation with machine learning-based predictive modeling and anomaly detection. Power data from an Erbium-doped femtosecond fiber laser operating at 775 nm are analyzed to assess variability, trends, and potential instabilities. Statistical analysis revealed moderate fluctuations in power output. Advanced anomaly detection techniques, including Isolation Forest and K-means clustering, identified distinct deviations in the data, with K-means achieving a Silhouette Score of 0.73. Predictive modeling using linear regression and ARIMA demonstrated robust forecasting capabilities. The ARIMA model effectively captured both short-term fluctuations and long-term trends, projecting stabilization of laser power over a 300-minute extension, indicative of equilibrium behavior. This study highlights the integration of statistical and machine learning tools as a valuable framework for enhancing precision and stability in high-performance laser applications.

Key words: Laser power, Laser power stability, anomaly, K-means clustering, Isolation Forest, linear regression, ARIMA model, predictive modeling, statistical analysis, power stability index (PSI), coefficient of variation (CV), machine learning in laser systems, time series forecasting.

DOI: https://doi.org/10.59277/RomJPhys.2025.70.909

1. INTRODUCTION

Laser power stability is critical in high-precision applications, including micromachining [1, 2], medical imaging [4, 5], and scientific experiments [3]. Even minor fluctuations in power can lead to significant errors in processes requiring consistent performance, especially in fields where sub-micrometer accuracy and reproducibility are important. Instability in laser systems often arises due to various factors such as environmental changes, electrical noise, thermal drift, mechanical misalignment, or, in some cases, optical misalignment and malfunctioning optical components, all of which can degrade the quality of the output and reduce overall system efficiency [6, 7]. Therefore, ensuring stable laser power is vital for maintaining system reliability and optimizing precision-based industrial and scientific processes.

Numerous studies have focused on analyzing laser stability using statistical methods to quantify variability and detect anomalies in laser output [8, 9]. While these approaches have contributed to understanding laser stability, many rely on small datasets or simplified models, which may not fully capture the complex dynamics of real-world laser systems. Other research efforts have primarily used retrospective data analysis without considering predictive modeling or advanced techniques for anomaly detection [12]. In contrast, our work extends the scope by employing a combination of statistical analysis and implemented machine learning-based methods, enabling a deeper understanding of power stability trends, variability, and potential anomalies in laser energy output. This approach is more adaptable to real-time, large-scale datasets [10, 11]. The novelty lies in our exploration of predictive modeling and advanced anomaly detection methods, evaluated for their potential to improve laser system stability. Unlike many previous works that rely solely on retrospective data analysis, our approach emphasizes predictive modeling to anticipate potential instabilities and enable proactive corrective measures. Techniques such as Isolation Forest [13, 14] and K-means clustering [15, 16] are applied as machine learning models to detect subtle anomalies overlooked by traditional methods. Additionally, predictive models like linear regression and ARIMA [17] are used in our analysis to forecast future behavior and guide timely interventions to ensure consistent performance.

Although the current analysis is conducted on a relatively small dataset, the methodology can be scaled to accommodate real-time, large-scale laser systems. This scalability makes it possible to implement similar techniques across various laser platforms, enabling proactive identification of potential instabilities before they escalate into significant issues. This study significantly bridges the gap between traditional statistical approaches and emerging machine learning-driven solutions for laser stability. The primary objectives of this study include identifying variability, detecting anomalies, and predicting future power trends. The methodology involves statistical analysis to quantify stability, employing machine learning-based anomaly detection techniques such as Isolation Forest and K-means clustering, and predictive modeling using linear regression and ARIMA. Key evaluation metrics, including the Power Stability Index (PSI), Coefficient of Variation (CV), and Mean Squared Error (MSE) [17], are used to evaluate system performance and model accuracy.

2. DATA ACQUISITION

The experiment used a commercially available Erbium-doped femtosecond fiber laser (EFOA-SH, Avesta, Inc.) [18] capable of generating pulses at over 100 fs. The laser operates at a central wavelength of 1560 nm, with a second harmonic output at 775 nm and a repetition rate of 71 MHz [19]. We selected the 775 nm center wavelength due to its proximity to the typical operational wavelength of conventional femtosecond lasers, ensuring compatibility with standard femtosecond laser systems. When configured at this wavelength, the laser delivered an output with ~ 84 mW average power, peak power of 13 kW, and a pulse energy of 1.2 nJ with a pulse duration of around 125 fs. The emitted pulses were directed to a calibrated power meter (PM100D, Thorlabs, Inc.) [20] to monitor the output power. Data acquisition was carried out continuously over a period of one hour, recording one dataset per minute, as shown in Fig. 1a, to ensure a comprehensive analysis of the acquired power data focused on detecting anomalies and understanding fluctuations in laser output.

3. RESULTS, ANALYSIS, AND DISCUSSION

3.1. STATISTICAL ANALYSIS

The statistical evaluation quantifies the central tendency and variability of the laser's power. The mean power (μ) is calculated as the average of all recorded values, $\mu = (1/N) \sum P_i = 78.32$, $N = 60$ is the total number of observations, and P_i represents the laser power at each time interval. The mean power of 78.32 mW provides a reference point around which the power values fluctuate. The standard deviation (σ), which measures the spread of the data around the mean, is computed $\sigma = \sqrt{[(1/N) \sum (P_\mathrm{i} - \mu)^2]} = 3.28$ mW [21]. The variance (σ^2), representing the degree of dispersion in the data, $\sigma^2 = 10.74$ mW^2. This value represents the squared deviation of the power values from the mean and reinforces the moderate variability observed in the dataset.

The histogram in Fig. 1b provides a detailed representation of the distribution of power values within the dataset, offering critical insights into the behavior and stability of the laser system. It highlights the frequency of power measurements across defined intervals, with the majority clustering near the mean value of approximately 78.32 mW, marked by the red dashed line. The median value, 79.6 mW, shown by the green dashed line, is slightly higher than the mean, suggesting a mild right skew that indicates occasional higher power readings. The histogram reveals that most power values lie within a narrow range of 80 mW to 81 mW, as evidenced by the tallest bin, with fewer measurements outside this interval. This concentration of values underscores the system's consistent performance with minimal variability. Furthermore, the histogram's smooth distribution and lack of sharp peaks or excessive spread reinforce the stability of the laser's power output.

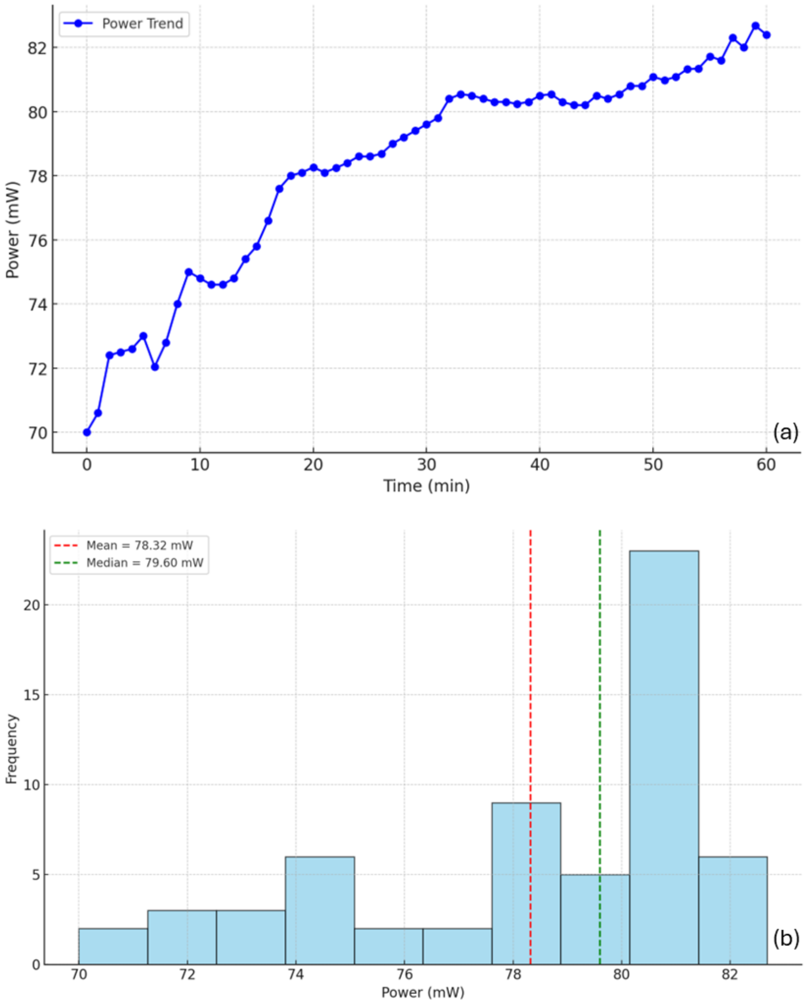

Fig. 1 – a) Laser power trend over time showing a gradual increase in power values; b) histogram of power data distribution, highlighting the mean (78.32 mW) and median (79.6 mW) power levels.

3.2. TREND ANALYSIS

Trend analysis uses a linear regression model [22] for the laser power data over 60 intervals. Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a straight line through the data points. The dependent variable is the laser power P in milliwatts, and the independent variable is time t in seconds. The linear regression model aims to find the best-fit line that minimizes the sum of the squared differences between the actual data points and the predicted values. Using this approach, the best-fit line for the given dataset $P(t) = 0.172\, t + 73.15\, P(t)$. The slope m = 0.172 mW indicates that the laser power increases by approximately 0.172 mW per second, while the intercept c = 73.15 mW suggests that the initial power at time zero is approximately 73.15 mW.

We used the least squares method to fit the linear regression model, which minimizes the sum of the squared residuals. Residuals are the differences between the actual observed values and the corresponding predicted values from the regression model. By minimizing these residuals, the model ensures that the fitted line best represents the overall trend of the data. The fitted trend line is shown in Fig. 2, where the blue curve represents the actual laser power measurements, and the red line is the linear regression line.

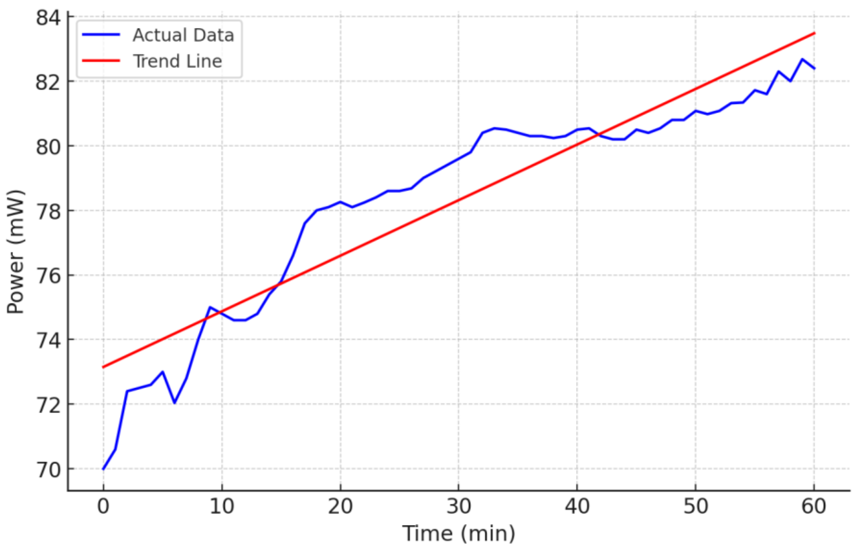

Fig. 2 – Trend analysis of laser power over time; the blue line represents the actual data, and the red line indicates the fitted linear trend, highlighting a gradual upward drift in peak power.

The Mean Absolute Error (MAE) [23] is a metric that quantifies the average magnitude of errors between the predicted values and the actual data points. This analysis uses MAE to assess the accuracy of the linear regression model fitted to the laser power data. The MAE is calculated by MAE = $(1/N) \sum |P_{\text{actual},i} - P_{\text{predicted},i}|$, N is the total number of data points, $P_{\text{actual},i}$ represents the observed (actual) laser power at time i, and $P_{\mathrm{predicted},i}$ is the predicted laser power at time i derived from the regression model. The absolute difference $|P_{\mathrm{actual},i} - P_{\mathrm{predicted},i}|$ represents the error for each data point. Table 1 shows a few data points to illustrate the process.

Table 1

Comparison of actual laser power measurements with predicted values using a linear regression model, including the corresponding absolute errors at different time intervals

Time (min)	Actual Power (mW)	Predicted Power (mW)	Absolute Error (mW)
0	70.6	73.32	2.72
10	75.0	74.87	0.07
20	78.1	76.42	1.66
30	79.6	78.32	1.28
40	80.3	80.04	0.46
50	80.8	81.76	0.68
60	82.4	83.48	1.08

Summing the absolute errors for all 60 data points and dividing by 60 gives MAE = $(1/N) \sum |P_{\mathrm{actual},i} - P_{\mathrm{predicted},i}|$ = 1.02 mW; this result indicates that, on average, the predicted power values deviate from the actual measurements by around 1.02 mW. The relatively low MAE value demonstrates that the linear regression model approximates the laser power trend with minimal deviation from the observed data points. This metric underscores the accuracy of the model and its potential use in predicting future laser power behavior.

3.3. ANOMALY DETECTION

Anomaly detection in laser power systems ensures consistent performance and prevents critical failures. This analysis employed four methods: Z-score [24], Moving Average [25], Isolation Forest [26], and K-means clustering [27] to detect anomalies in the provided laser power data. The Z-score method involves standardizing the data using the relation $Z = (x_i - \mu)/\sigma$, where x_i is the data point, μ = 78.32 mW is the mean of the dataset, and σ = 3.28 mW is the standard deviation. Data points with a Z-score greater than 3 or less than –3 are considered anomalous; however, no data points fell outside this range, indicating the absence of anomalies using this criterion. Similarly, the moving average method with a window size of 5 is applied, where the moving average at any point i is calculated using $\mathrm{MA}_i = (1/5)\Sigma^{i+2}_{j=i-2}\mathrm{x}_j$. A confidence interval is defined as Upper

Bound = $MA_i + 2\sigma$ and Lower Bound = $MA_i - 2\sigma$. Despite small fluctuations in the data, no data points exceeded these bounds, and hence, no anomalies were detected using this approach.

Moving towards machine learning-based approaches, the flowchart in Fig. 3a illustrates the anomaly detection process using Isolation Forest and K-means clustering. The process begins by inputting and preprocessing the data (removing missing values and normalizing). After preprocessing, the data is processed separately by the Isolation Forest and K-means methods. The anomalies detected by both methods are compared, and the results are output to identify significant anomalies. The Isolation Forest method is employed for anomaly detection, which is based on the principle that rare and distinct from normal data points anomalies can be isolated with fewer partitions when the data is randomly divided. The algorithm constructs an ensemble of binary trees by randomly selecting features and splitting values to create partitions. The depth of the path required to isolate a data point determines its anomaly score, whereas points requiring shorter path lengths are considered anomalies. The anomaly score for a data point x is given by $\mathrm{Score}(x) = 2^{-h(x)/c(n)}$, $h(x)$ represents the average path length of x across all trees, and $c(n)$ is the average path length of an unsuccessful search in a Binary Search Tree. The function $c(n)$ is approximated as $c(n) \approx 2\ln(n) + \gamma$, n is the total number of data points, and $\gamma \approx 0.577$ is the Euler-Mascheroni constant. In our case, the Isolation Forest model is trained on a dataset of 60 observations representing laser power values collected over one hour. Each data point is assigned an anomaly score, and a threshold of 0.7 is used to flag anomalies. The 12 data points are identified as anomalies, distributed roughly over the first 17 minutes, and around 3 (57 – 60 mins) points are flagged as anomalies at the end of the power data plot (82.3 – 82.4 mW) because they deviated significantly from the measurement's general trend in power values (Fig. 4a). To illustrate the process, assume the average path length for a normal point is $h_{\mathrm{normal}} = 8.7$, and for an anomalous point, it is $h_{\mathrm{anomaly}} = 3.0$. Using $n = 60$ the average path length $c(n)$ is computed as $c(60) = 2\ln(60) + 0.577 = 8.766$. The anomaly score for a normal point is then calculated as $\mathrm{Score}_{\mathrm{normal}} = 2^{-8.5/8.765} = 2^{-0.97} \approx 0.511$, and for an anomalous point, it is $\mathrm{Score}_{\mathrm{anomaly}} = 2^{-3.0/8.765} = 2^{-0.342} \approx 0.789$. Since the anomaly score for the anomalous point (0.789) exceeds the threshold of 0.7, it is correctly flagged as an anomaly, whereas the normal point (0.511) remains below the threshold and is not flagged. These anomalies, highlighted in Fig. 4a, are likely attributed to sudden changes in environmental conditions, transient electrical noise, or temporary misalignment of optical components.

On the other hand, K-means clustering is applied to the laser power data by dividing it into $k = 2$ clusters. The objective of K-means is to minimize the within-cluster sum of squares (WCSS), as $\mathrm{WCSS} = \Sigma^{k}_{i=1}\Sigma^{ni}_{j=1}\|x^{(i)}_j - \mu_i\|^2$, $x^{(i)}_j$ represents the j-th data point in cluster i, μ_i is the centroid of cluster i, and n_i is the number of points in cluster i, WCSS ≈ 124.33. In this case, two clusters are formed: the first cluster had a mean value of μ_1 = 73.62 mW, representing the normal power range, and the second cluster had a mean value of μ_2 = 80.13 mW, representing anomalous power values. Data points in the second cluster, farther from the normal centroid, are flagged as anomalies. The Silhouette Score is computed to evaluate the quality of clustering. The Silhouette Score measures how well each data point fits within its assigned cluster compared to the nearest cluster $S = (b–a)/\max(a,b)$, where a is the average intra-cluster distance (*i.e.*, the average distance between a point and other points in the same cluster), and b is the average nearest-cluster distance (*i.e.*, the average distance between a point and points in the nearest neighboring cluster). In our analysis, a = 0.45 and b = 0.82. After substituting, we yield S ≈ 0.451, which indicates that the clusters are well-separated, meaning the clustering result is reliable, with clear distinctions between the normal and anomalous data points. Figure 4b shows that 17 anomalies are detected using K-means clustering. These anomalies are distributed roughly over the first 17 minutes of the measurement. These anomalies are identified as belonging to the cluster with the higher mean power (μ_2 = 80.13 mW).

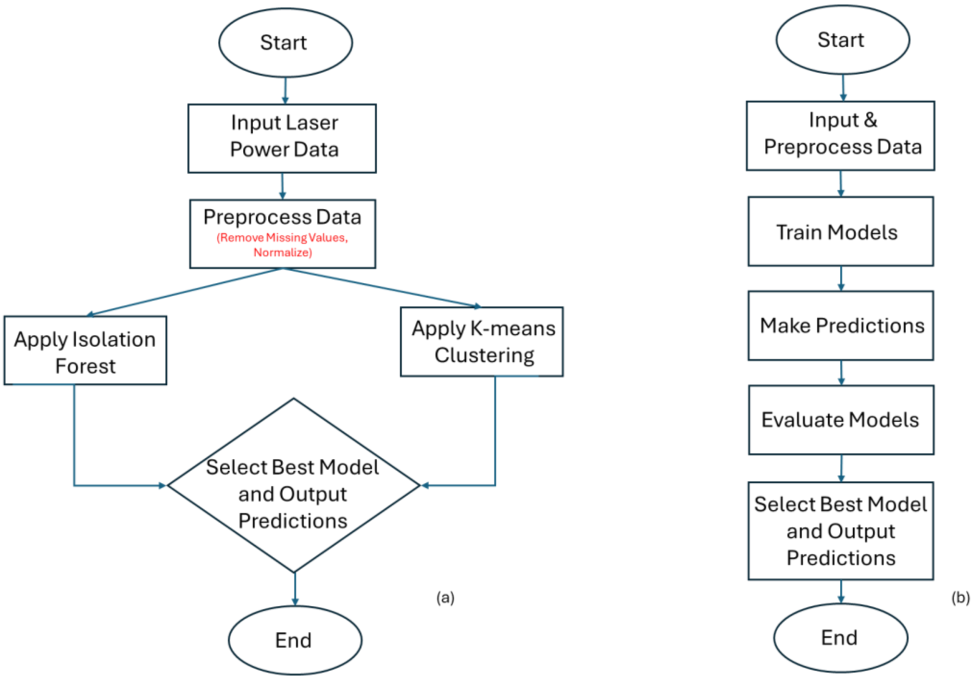

Fig. 3 – Flowcharts illustrating the workflows for: a) anomaly detection, including preprocessing and application of Isolation Forest and K-means clustering; b) predictive modeling, detailing preprocessing, model training (linear regression and ARIMA), and evaluation for forecasting laser power trends.

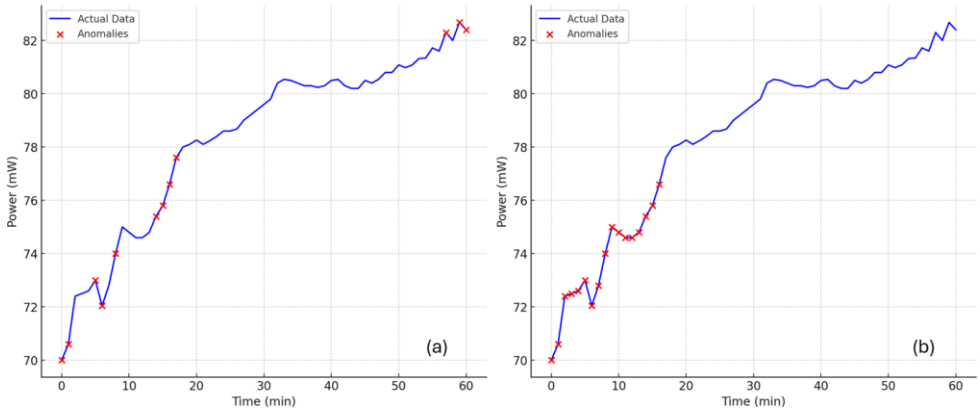

Fig. 4 – a) Anomaly detection in laser power using Isolation Forest; b) anomaly detection in K-means clustering plot, the blue lines represent the actual data, and the red markers indicate detected anomalies at specific time points.

A comparison of the methods reveals that machine learning-based techniques, particularly Isolation Forest and K-means clustering, provided more reliable detection of anomalies in the laser power data, while the *Z*-score and moving average methods failed to detect significant deviations. The Isolation Forest identified thirteen anomalies, and K-means clustering identified seventeen anomalies, each indicating critical deviations that may impact system performance. The detected anomalies highlight the importance of monitoring laser power outputs for sudden fluctuations caused by external disturbances or internal component behavior. Moreover, numerical analysis and clustering metrics, such as the Silhouette Score, ensure that the detected patterns are statistically significant and represent real deviations rather than noise in the system.

3.4. PREDICTIVE MODELING

Predictive modeling is crucial for anticipating future laser power fluctuations and implementing proactive measures to ensure system stability. Two models, Linear Regression [28] and ARIMA [29], are applied to the provided laser power data for forecasting purposes.

The flowchart in Fig. 3b represents the step-by-step process for building a predictive model. It begins with inputting and preprocessing data, including removing missing values, normalizing, and splitting the dataset. The next steps involve training models and making predictions using both models. Finally, the models are evaluated, the best model is selected, and the predicted power values are output. A linear regression model is initially trained on historical power data to forecast future values. The regression equation $P_{\mathrm{pred}(t)} = \beta_0 + \beta_{1t}$, where $P_{\mathrm{pred}(t)}$ represents the predicted power at time t, β_0 is the intercept, and β_1 is the slope of the regression line. The model is evaluated using the Mean Squared Error (MSE) and R-squared (R^2) metrics [30], $MSE_{linear\text{-}aggression} = (1/n)\,\Sigma^{n}_{i=1}(P_i - P_{pred,i})^2 = 1.38$, and $R^2 = 0.87$. A linear regression model with an MSE of 1.38 and an R^2 of 0.87 indicates strong predictive performance. The low MSE reflects minimal prediction errors, while the high value of 0.87 of R^2 explains 87% of the pattern in the data; overall, the model is accurate and closely follows the real data. The model is then used to predict power values for the next 10 and 300 minutes, respectively, as shown by the green dashed line in Figs. 5a,b. The predicted values closely follow the observed upward trend.

An ARIMA(1, 1, 1) model [28] is also applied for time series forecasting. The ARIMA model incorporates autoregressive terms (AR), differencing (I), and moving average terms (MA) to model both short-term fluctuations and long-term trends. The ARIMA model equation $P_t = \phi P_{t-1} + \theta\epsilon_{t-1} + \epsilon_t$, where ϕ is the autoregressive coefficient, θ is the moving average coefficient, and ϵ_t represents white noise. After fitting the model, the resulting MSE is $MSE_{ARIMA} = 0.18$; the ARIMA model has a lower MSE compared to the linear regression model, which shows that ARIMA is more effective in handling the complexities of time-series data, such as fluctuations and patterns, while linear regression is better suited for capturing simpler trends. The ARIMA model proved robust in capturing both short-term fluctuations and long-term trends, as evidenced by the red dotted line in Figs. 5a,b, which shows the predicted values for the next 10 and 300 minutes. The red dotted line in the 10-minute extension closely follows the recent upward trend of the observed data, showing accurate short-term predictions, and in the case of the 300 minute extension, the ARIMA red dotted line demonstrates stability around 93–94 mW, suggests the laser may reach an equilibrium, as ARIMA predicts based on past trends indicating a saturation-like behavior. This reflects ARIMA's tendency to stabilize predictions over extended periods, balancing recent trends while avoiding unrealistic extrapolation into high values. However, this reflects a data-driven assumption and may not account for external factors affecting real-world laser behavior.

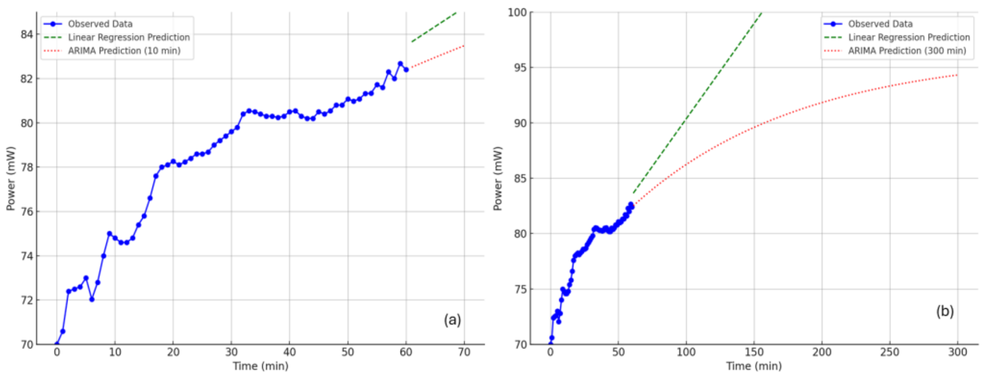

Fig. 5 – Predictive modeling: a) 10 minute extension of observed data using Linear Regression and ARIMA models; b) 300 minute extension of the same observed data using the same models, observed data is represented by the blue line, Linear Regression predictions by the green dashed line, and ARIMA predictions by the red dotted line.

Table 2

A summary of key metrics in predictive modeling

Metrics	Value
Power Stability Index (PSI)	0.1577
Coefficient of Variation (CV)	0.0399
Silhouette Score	0.73
ARIMA Mean Squared Error (MSE)	0.18
Linear Regression MSE	1.38
R-squared (R2)	0.87

Key metrics are computed to evaluate the system's performance and variability further. The Power Stability Index (PSI), which measures the range of fluctuations relative to the mean power, is calculated as PSI = $(P_{max}-P_{min})/\mu = (82.5-70.0)/79.27 \approx 0.1577$, *i.e.*, indicating that the range of fluctuations is approximately 15.77% of the mean power, suggesting moderate instability. Furthermore, the Coefficient of Variation (CV), a dimensionless metric representing the ratio of the standard deviation to the mean, is computed as CV = $\sigma/\mu = 3.16/79.27 = 0.03986$. A CV of 3.99% indicates relatively low variability compared to the mean power, confirming that the system is fairly stable despite the observed anomalies. As mentioned earlier, the K-means model's clustering performance is evaluated using the Silhouette Score of 0.73, reflecting good cluster separation. The linear regression and ARIMA models provided reliable predictions; the ARIMA model's ability to account for fluctuations and trends makes it a robust choice for time series forecasting in laser power systems. These predictive insights and computed metrics (Table 2) provide a comprehensive framework for monitoring laser performance and ensuring system stability over time.

4. CONCLUSION

This study presents a comprehensive framework for analyzing and predicting laser power stability by integrating statistical evaluation with machine learning-based anomaly detection and predictive modeling. Real-world data analysis of an Erbium-doped femtosecond fiber laser operating at 775 nm revealed moderate fluctuations, with a mean power of 78.32 mW, a standard deviation of 3.28 mW, and a Power Stability Index (PSI) of 0.1577, indicating a stable yet moderately variable output suitable for precision applications. Machine learning-based anomaly detection techniques, such as Isolation Forest and K-means clustering, proved effective, identifying 12 and 17 anomalies and highlighting environmental and system-specific instabilities. Predictive ARIMA modeling demonstrated superior forecasting capabilities, projecting long-term stabilization at 93–94 mW over a 300-minute horizon. These results emphasize the critical role of advanced statistical and machine learning methods in ensuring consistent laser performance, providing actionable insights for real-time monitoring, adaptive control, and predictive maintenance. This study establishes a robust foundation for advancing laser technology by integrating data-driven techniques, contributing to more reliable and efficient systems for high-precision applications.

Acknowledgements. The authors would like to greatly acknowledge the support of Muhammad Sabieh Anwar, Professor at Syed Babar Ali School of Science and Engineering, Lahore University of Management Sciences (LUMS), Lahore, Pakistan.

Data availability statement. The data supporting this manuscript's findings is not available in any repository. The data supporting this study's findings are available within the article.

REFERENCES

1. Mishra, S., Yadava, V., *Laser beam micromachining (LBMM)–a review*, Optics and Lasers in Engineering **73**, 89–122 (2015).
2. Liu, X., Du, D., Mourou, G., *Laser ablation and micromachining with ultrashort laser pulses*, IEEE Journal of Quantum Electronics **33**(10), 1706–1716 (1997).
3. Smrž, M., Mužík, J., Štěpánková, D., Turčičová, H., Novák, O., Chyla, M., Mocek, T., *Picosecond thin-disk laser platform PERLA for multi-beam micromachining*, Osa Continuum **4**(3), 940–952 (2021).
4. Kieffer, J. C., Fourmaux, S., Krol, A. (2017, January), *The ultrafast high-peak power lasers in future biomedical and medical X-ray imaging*. In *19th International Conference and School on Quantum Electronics: Laser Physics and Applications*, Vol. 10226, pp. 306–315. SPIE.
5. Leblans, P., Vandenbroucke, D., Willems, P., *Storage phosphors for medical imaging*, Materials **4**(6), 1034–1086 (2011).
6. Oancea, C. D., *A Method of Reducing Errors Due to Sampling in the Measurement of Electric Power*, Applied Sciences **14**(9), 3827 (2024).
7. Carp, J., Compton, R. J., *Alpha power is influenced by performance errors*, Psychophysiology **46**(2), 336–343 (2009).
8. Sheta, S., Afgan, M. S., Hou, Z., Yao, S. C., Zhang, L., Li, Z., Wang, Z., *Coal analysis by laser-induced breakdown spectroscopy: a tutorial review*, Journal of Analytical Atomic Spectrometry **34**(6), 1047–1082 (2019).
9. Yoon, J. W., Kim, Y. G., Choi, I. W., Sung, J. H., Lee, H. W., Lee, S. K., Nam, C. H., *Realization of laser intensity over 10 23 W/cm^2*, Optica **8**(5), 630–635 (2021).
10. Döpp, A., Eberle, C., Howard, S., Irshad, F., Lin, J., Streeter, M., *Data-driven science and machine learning methods in laser–plasma physics*, High Power Laser Science and Engineering **11**, e55 (2023).
11. Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Zdeborová, L., *Machine learning and the physical sciences*, Reviews of Modern Physics **91**(4), 045002 (2019).
12. Bedolla, E., Padierna, L. C., Castaneda-Priego, R., *Machine learning for condensed matter physics*, Journal of Physics: Condensed Matter **33**(5), 053001 (2020).
13. Liu, F. T., Ting, K. M., Zhou, Z. H. (2008, December), *Isolation forest*. In *2008 Eighth IEEE International Conference on Data Mining* (pp. 413–422). IEEE.

14. Xu, H., Pang, G., Wang, Y., Wang, Y., *Deep isolation forest for anomaly detection*, IEEE Transactions on Knowledge and Data Engineering **35**(12), 12591–12604 (2023).
15. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., Heming, J., *K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data*, Information Sciences **622**, 178–210 (2023).
16. Zubair, M., Iqbal, M. A., Shil, A., Chowdhury, M. J. M., Moni, M. A., Sarker, I. H., *An improved K-means clustering algorithm towards an efficient data-driven modeling*, Annals of Data Science **11**(5), 1525–1544 (2024).
17. Hodson, T. O., *Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not*, Geoscientific Model Development Discussions **2022**, 1–10 (2022).
18. https://avesta.ru/en/product/efoa-sh-femtosecond-er-fiber-laser-with-built-in-shg/
19. Imran, T., Naeem, M., Hussain, M., *An experimental study of intensity-phase characterization of femtosecond laser pulses propagated through a polymethyl methacrylate*, Microwave and Optical Technology Letters **66**(6), e34217 (2024).
20. https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=3341
21. Przystupa, K., Kolodiy, Z., Yatsyshyn, S., Majewski, J., Khoma, Y., Petrovska, I., Hut, T., *Standard deviation in the simulation of statistical measurements*, Metrology and Measurement Systems **30**(1), (2023).
22. Maulud, D., Abdulazeez, A. M., *A review on linear regression comprehensive in machine learning*, Journal of Applied Science and Technology Trends **1**(2), 140–147 (2020).
23. Chai, T., Draxler, R. R., *Root mean square error (RMSE) or mean absolute error (MAE)*, Geoscientific model development discussions **7**(1), 1525–1534 (2014).
24. Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., Suvas, A., *Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model*, Journal of International Financial Management & Accounting **28**(2), 131–171 (2017).
25. Yang, X., Ni, Y., *Least-squares estimation for uncertain moving average model*, Communications in Statistics-Theory and Methods **50**(17), 4134–4143 (2021).
26. Cheng, Z., Zou, C., Dong, J., *Outlier detection using isolation forest and local outlier factor*, in *Proceedings of the conference on research in adaptive and convergent systems*, pp. 161–168, 2019, September.
27. Ahmed, M., Seraj, R., Islam, S. M. S., *The k-means algorithm: A comprehensive survey and performance evaluation*, Electronics **9**(8), 1295 (2020).
28. Hope, T. M., *Linear regression*, in *Machine learning*, Academic Press, 2020, pp. 67–81.
29. Shumway, R. H., Stoffer, D. S., Shumway, R. H., Stoffer, D. S., *ARIMA models*, Time series analysis and its applications: with R examples, 75–163 (2017).
30. Gao, J., *R-Squared (R2)–How much variation is explained?*, Research Methods in Medicine & Health Sciences **5**(4), 104–109 (2024).