

FACILITY-WIDE MONITORING AND EARLY DETECTION OF FAILURES AT LANSCE*

N. Yampolsky[†], A. Garcia, En-Ch. Huang, J. Quemuel, M. Rautela, A. Scheinker,
Los Alamos National Laboratory, Los Alamos, United States

Abstract

Accelerator complexes contain tens of thousands of interdependent components, and aging infrastructure amplifies the risk of equipment faults and costly, unscheduled shutdowns. At the Los Alamos Neutron Science Center (LANSCE), we are developing a data-driven framework that flags developing problems early enough to address them during scheduled maintenance, thereby improving reliability and increasing beam availability for users. Our approach analyzes all available signals within a subsystem to learn the facility’s “normal” operating envelope and to detect subtle deviations that precede failures. Unlike the current warning scheme, it captures hidden correlations among parameters and generates interpretable indicators of abnormal behavior. Predictions are validated against historical control-room log records. We report progress on three fronts: (i) extending anomaly prediction from a single beamline to all major LANSCE subsystems; (ii) expanding data archiving capacity by an order of magnitude to support broader coverage and longer look-back windows; and (iii) developing operator-facing algorithms that both warn of emerging anomalies and localize likely problem elements along the beamline. Together, these advances are designed to shift maintenance from emergency response to planned intervention, reducing downtime and enhancing overall facility performance.

INTRODUCTION

The Los Alamos Neutron Science Center (LANSCE) is a multi-user accelerator facility that accelerates both H⁺ and H⁻ beams and delivers them to several experimental areas with different energy and timing requirements [1]. The accelerator must therefore operate in multiple beam modes, and many diagnostics characterize only a subset of those modes. This makes automated assessment of machine status more difficult than monitoring a single-purpose industrial system. At the same time, LANSCE is a mature facility whose reliability is affected by aging equipment, intermittent subsystem degradation, and complex operational history. The need for early identification and localization of equipment problems is therefore directly connected to beam availability.

Traditional alarm systems are based primarily on threshold checks: an alarm is generated when a selected process variable exceeds a predefined range. Such systems are

useful, but they have two limitations for predictive maintenance. First, only a small fraction of the available process variables can be manually curated by operators and subject matter experts. Second, threshold alarms usually treat variables independently and therefore miss correlated deviations in which every individual variable remains within its allowed range. Accelerator operation is highly correlated: magnet setpoints and readbacks, beam current, losses, vacuum, radio-frequency parameters, cooling, and environmental variables all evolve together. A physically abnormal state may therefore appear not as a single out-of-range value, but as an unusual combination of otherwise acceptable readings.

Another difficulty is accelerator redundancy. A failing element may be compensated by retuning other components, allowing acceptable beam delivery to continue while the underlying machine state drifts away from its historical norm [2]. Such compensation is useful operationally, but it hides the early stages of degradation. A predictive maintenance system must therefore separate harmless retuning and environmental drift from signatures associated with equipment degradation.

This work develops and evaluates an unsupervised anomaly-detection framework for accelerator data, extending recent efforts to apply machine-learning anomaly detection to accelerator facilities [3]. It does not require a complete labelled database of failures, which is important because equipment failures are rare and historical records are incomplete. Instead, the method learns the distribution of normal operation from archived process variables, assigns an anomaly score to new states, and identifies the channels most responsible for the deviation. The approach was first tested on Line D, the high-energy beam transport line from the linear accelerator to the Proton Storage Ring (PSR), and was later observed during 2024 front-end operation. The method is intended to complement operator judgment rather than replace it: high anomaly scores and high-impact channels provide a ranked list of machine states requiring attention.

MATHEMATICAL FORMALISM

At each archived time step, the accelerator state is represented by an n -dimensional vector

$$x(t) = \{x_1(t), x_2(t), \dots, x_n(t)\}^T, \quad (1)$$

where each coordinate is a readout from an accelerator diagnostic or controls process variable. These variables include beam currents, magnet currents and voltages, vacuum signals, RF parameters, cooling-system parameters, beam losses, status readbacks, *etc.* During normal operation, $x(t)$ does not remain fixed. It changes because of

*Research presented in this paper was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project number 20240474MFR.

[†]nyampols@lanl.gov

operator tuning, beam-mode changes, environmental effects, measurement noise, and slow equipment evolution. A collection of normal operating points therefore occupies a finite volume in the n -dimensional phase space.

For a training interval selected to represent normal operation, the mean state and covariance matrix are

$$\mu = \langle x \rangle, \quad C = \langle (x - \mu)(x - \mu)^T \rangle. \quad (2)$$

The data are analyzed in the principal-component basis of the covariance matrix,

$$C = V\Lambda V^T, \quad (3)$$

where the columns of V are eigenvectors of the covariance matrix and Λ contains the corresponding eigenvalues. The eigenvectors with the largest eigenvalues represent the dominant correlated patterns in the accelerator data. These are the directions along which the machine normally moves during routine operation. Eigenvectors with small eigenvalues are more likely to represent poorly conditioned directions, weakly informative combinations of variables, missing-data artifacts, or noise-dominated behavior.

In practice, only a retained subset of eigenvectors is used to compute the anomaly score. Let V_K denote the matrix containing the retained eigenvectors corresponding to the largest eigenvalues. This projection acts as a filter: it reduces the amount of information used for alarm generation and keeps only the dominant correlated behavior of the machine. As a result, the method can accept a large set of available accelerator data without requiring extensive manual preselection. Uninformative channels, poorly conditioned combinations of channels, and noise-dominated directions are naturally suppressed by the PCA filter.

For a new machine state, we define

$$\delta(t) = x(t) - \mu, \quad (4)$$

and compute the anomaly score as the Mahalanobis norm in the retained PCA subspace,

$$S(t) = [\delta^T(t) V_K (V_K^T C V_K)^+ V_K^T \delta(t)]^{1/2}, \quad (5)$$

where $+$ denotes the pseudoinverse. Equivalently, in the retained PCA coordinates,

$$S(t) = \left[\sum_{k \in K} \frac{(v_k^T \delta(t))^2}{\lambda_k} \right]^{1/2}. \quad (6)$$

A large value of $S(t)$ indicates that the present accelerator state is unlikely relative to the historical correlated behavior used for training.

The number of retained eigenvectors is an adjustable parameter. Keeping more eigenvectors preserves more information but can also admit noisy or poorly conditioned directions. Keeping fewer eigenvectors produces a stronger filter but may remove useful diagnostic information. The optimal retained fraction can therefore be tuned for a particular machine, beam mode, or subsystem. In this study, we retained 15% of the eigenvectors corresponding to the largest eigenvalues, i.e., the most dominant correlated modes in the training data.

This procedure makes the alarm system sensitive to changes in the learned correlations between accelerator variables. Under normal conditions, many process variables move together: power-supply setpoints and readbacks remain correlated, beam losses are related to upstream

beam conditions, and groups of magnets, RF systems, vacuum signals, and cooling parameters follow historically observed relationships. Equipment degradation can break these correlations before any single process variable crosses a conventional threshold. The anomaly score therefore does not simply measure whether one signal is too high or too low. It measures whether the combined accelerator state is inconsistent with the correlated behavior learned from historical operation.

Routine retuning that moves the machine along previously observed correlated directions is largely absorbed by the PCA model and does not necessarily produce a large anomaly score. However, a retune into a new configuration that is not represented in the training data may also trigger an alarm. Such a signal should be interpreted as a warning that the machine has moved outside the learned normal operating state, not automatically as proof of equipment failure. In practice, alarms are attributed to equipment problems only after considering the machine context and the ranked list of contributing process variables.

Localization is obtained by estimating the contribution of each process variable to the anomaly score. For each channel i , the score is recomputed after replacing that channel by its historical mean value, or equivalently by setting $\delta_i(t) = 0$, while leaving all other channels unchanged. The reduction in anomaly score defines the contribution of that channel,

$$\Delta S_i(t) = S(t) - S_i^{(0)}(t), \quad (7)$$

where $S_i^{(0)}(t)$ is the anomaly score computed with channel i set to its historical mean. Channels that produce the largest reduction in score are ranked as the largest contributors to the anomaly. This procedure provides an interpretable way to identify which process variables are most responsible for the alarm.

The instantaneous ranking of channels can fluctuate because of measurement noise, unaccounted environmental effects, and routine machine adjustments. We therefore analyze time-averaged rankings in addition to instantaneous rankings. If channel i has rank $N_i(t)$, where $N_i = 1$ corresponds to the largest contribution, we define its average rank over a selected time window as

$$\bar{N}_i = \left\langle \frac{1}{N_i(t)} \right\rangle^{-1}. \quad (8)$$

This definition gives more weight to times when a channel appears near the top of the contribution list. Smaller values of \bar{N}_i correspond to more problematic channels. The averaging window can be adjusted from minutes to days. Short windows are useful for identifying rapidly developing failures, while longer windows are used to reveal slowly developing equipment degradation [4].

OPERATIONAL EXPERIENCE

Operational observations showed that the anomaly-score analysis produces two useful classes of signals. The first is a sudden increase in anomaly score, corresponding to an accelerator state that rapidly leaves the learned region of normal operation. These events are interpreted as crisis-type alarms: the score indicates urgency, while the ranked

process-variable contributions provide localization. This is important because failure recovery time is often dominated not only by repair, but also by identifying which subsystem caused the interruption.

The second type of signals is slow degradation. The anomaly score may not show a sharp spike because operators can often compensate for developing hardware problems by retuning other parts of the accelerator. Beam delivery may continue, while the machine state gradually drifts away from historical correlations. Such degradation appears as persistent high contribution from a related group of process variables (PVs) rather than as a single excursion. Time-averaged contribution rankings are used to separate rapidly developing faults from long-term degradation.

Predictability was tested by training on one run cycle and applying the model to the next. A model trained on 2020 data was used to forward-predict 2021 anomaly behavior and compared with a model built directly from 2021 data. The two anomaly-score time series were correlated above 80%, showing that for stable subsystems the learned correlations are not merely retrospective; they remain predictive in later operation. However, individual PV statistics can change between run cycles, so deployed models require periodic retraining or recalibration.

A separate robustness test compared models built from the full available channel list with models built from carefully selected channels. In both cases, the same number of PCA eigenvectors was retained. The resulting anomaly-score time series were correlated above 90%. This indicates that manual preselection of useful signals is not required. The PCA projection acts as an information filter: poorly conditioned directions, uninformative PVs, missing-data artifacts, and noise-dominated combinations are suppressed automatically. The method can therefore ingest broad accelerator data sets, while the retained PCA fraction controls how much information is actually used for alarm generation. This fraction is a tunable parameter for each machine section or operating mode.

Operational performance depends strongly on the subsystem. The method is more predictive in sections with stable operating points and limited retuning, where deviations from the covariance model are more likely to indicate true equipment problems. Downstream transport lines satisfy this condition better than the front end. The front end is frequently retuned and shows larger run-to-run variability; consequently, legitimate tuning changes can resemble abnormal behavior and reduce model transferability.

Even in this challenging case, the 2024 front-end study showed a measurable benefit. Model predictions were compared in detail with Central Control Room logs, and each alarm was reviewed to determine whether it was real or false. The anomaly-detection model identified approximately 25% more true alarms than operators recognized in real time, but with an approximately 50% higher false-alarm rate. Thus, the model is not yet suitable as a fully autonomous alarm system. Its present value is as a complementary operator tool that highlights unusual correlated behavior and points to the PVs most responsible for it.

The 2025 run cycle provided a substantially improved data environment. A new EPICS Archiver Appliance was implemented before the run to support recording of all available PVs [5], and the number of front-end PVs available to the anomaly model more than doubled. Many added PVs were constant, inactive, or otherwise not dynamically informative, but PCA filtering allowed them to be included without manual removal. With the expanded data set, the anomaly score behaved better than in previous years: it was more predictive and produced fewer false alarms. This supports the central assumption of the approach: more complete accelerator data improve anomaly detection when combined with a robust filtering step.

SUMMARY AND OUTLOOK

The operational studies demonstrate that covariance- and PCA-based anomaly detection provides useful predictive information for accelerator operations. The method identifies sudden failures through rapid increases in anomaly score and slow degradation through persistent high contribution from related PVs. Historical tests showed that models trained on one run cycle can forward-predict later behavior in stable subsystems, and that PCA filtering reduces requirements on data quality as inputs. This reduces the need for manual signal selection and makes the approach scalable to large accelerator data sets.

The 2024 front-end study showed both promise and limitations: the model found additional true alarms missed in real time, but the false-alarm rate remained significant. The expanded 2025 archiver improved input-data completeness, and early analysis indicates better anomaly-score behavior with fewer false alarms. These results show that the system works on real accelerator data and that more complete archiving improves performance. The next step is real-time validation, in which anomaly scores and ranked PV contributions are monitored during operation and compared with subsequent operator actions, log entries, and equipment interventions.

REFERENCES

- [1] P. W. Lisowski and K. F. Schoenberg, "The Los Alamos Neutron Science Center", *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 562, no. 2, pp. 910–914, 2006. doi:10.1016/j.nima.2006.02.178
- [2] A. Scheinker, E.-C. Huang, and C. E. Taylor, "Extremum seeking-based control system for particle accelerator beam loss minimization", *IEEE Trans. Control Syst. Technol.*, vol. 30, no. 5, pp. 2261–2268, 2022. doi:10.1109/TCST.2021.3136133
- [3] A. Das, M. Borland, L. Emery, X. Huang, D. F. Ratner, H. Shang, G. Shen, R. Smith, and G. Wang, "Anomaly detection in accelerator facilities using machine learning", in *Proc. IPAC'21*, Campinas, SP, Brazil, May 2021, pp. 304–307. doi:10.18429/JACoW-IPAC2021-MOPAB077
- [4] E. Huang and N. Yampolsky, "Anomaly detection of slow-moving variables at LANSCE for improved beam quality", in *Proc. NAPAC'25*, Sacramento, CA, USA, Aug. 2025, pp. 55–57. doi:10.18429/JACoW-NAPAC2025-MOP007

- [5] J. Quemuel, S.A. Baily, A. Georgiou, E. Huang, E. Westbrook, N. Yampolsky, “EPICS Archiver Appliance at Los Alamos Neutron Science Center”, in *Proc. ICALEPCS’25*, Chicago, IL, USA, Sep. 2025, pp. 609–611.
[doi:10.18429/JACoW-ICALEPCS2025-TUPD038](https://doi.org/10.18429/JACoW-ICALEPCS2025-TUPD038)