Project work: NASA Turbofan Jet Engine A3

Alessia Tani, Iiro Vendelin, Sara Zambetti

1 Introduction

The degradation behavior of turbofan jet engines can be effectively analyzed through multivariate sensor data collected during operation. The NASA C-MAPSS datasets (FD001–FD004) provide such measurements, enabling the development of data-driven monitoring frameworks for predictive maintenance. The present work aims to model healthy operating conditions and detect deviations that precede failure by approximately 10–20 cycles. The methodology integrates preprocessing (RUL labeling, removal of constant variables, z-score normalization), dimensionality reduction via Principal Component Analysis (PCA) and its nonlinear extension (KPCA), and Multivariate Statistical Process Control (MSPC) through Hotelling's T^2 and Q/SPE charts. Models are calibrated on healthy early-life data and validated in MATLAB. Performance is assessed by True Positive Rate, False Alarm Rate, and Detection Lead Time, with sensor contributions analyzed for diagnostic insight.

2 Materials and Methods

2.1 Data Description

The analysis is based on the NASA C-MAPSS turbofan engine degradation datasets (FD001–FD004), which simulate the progressive wear of multiple engines operating under different conditions until failure. Each dataset contains multivariate time series of sensor measurements collected over successive operating cycles for several engine units. Each record corresponds to one cycle of a specific engine and includes 3 operational settings and 21 sensor readings, together with a unit and cycle index. The training sets provide complete trajectories from start-up to failure, whereas the test sets include partial sequences with corresponding Remaining Useful Life (RUL) labels.

Prior to modeling, a pretreatment pipeline was designed to ensure data consistency and comparability. For each engine in the training set, RUL labels were computed as the difference between the maximum and current cycle. Constant sensors were removed to avoid numerical redundancy, and all remaining variables were standardized through z-score normalization:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}, \tag{1}$$

where x_{ij} is the original measurement of sensor j in observation i, and μ_j, σ_j denote the mean and standard deviation of sensor j across all engines. These preprocessing steps preserved the temporal structure of each unit and prepared the data for dimensionality reduction and multivariate monitoring.

2.2 Mathematical Methods

The adopted methodology combines dimensionality reduction with Multivariate Statistical Process Control (MSPC) to model normal engine behavior and detect deviations indicative of degradation or faults. The modeling framework comprises three main stages: (i) feature extraction through Principal Component Analysis (PCA) and its nonlinear extension, Kernel PCA (KPCA); (ii) monitoring through multivariate control charts based on Hotelling's T^2 and Squared Prediction Error (SPE or Q) statistics; and (iii) model calibration, validation, and performance evaluation.

Principal Component Analysis (PCA). PCA is employed to reduce the dimensionality of the sensor data while retaining most of its variance. For a centered and standardized data matrix $X \in \mathbb{R}^{n \times p}$, with n observations and p variables, the sample covariance matrix is

$$S = \frac{1}{n-1} X^\top X. \tag{2}$$

The eigenvalue decomposition of S is given by

$$SP = P\Lambda, \tag{3}$$

where the columns of P are the eigenvectors (loadings) and Λ is a diagonal matrix containing the eigenvalues, which represent the variance explained by each principal component. The projection of the data into the principal component space is obtained as

$$T = XP, \tag{4}$$

where T is the score matrix. The first k principal components—typically those explaining at least 90% of the total variance—are retained to form the reduced feature space, thereby minimizing multicollinearity and noise while preserving the dominant structure of the data.

Kernel Principal Component Analysis (KPCA). While PCA captures linear correlations among sensors, KPCA generalizes this approach to nonlinear relationships through the kernel trick. Given observations $\{x_i\}_{i=1}^n$, KPCA constructs a kernel matrix K whose elements are defined as

$$K_{ij} = \kappa(x_i, x_j), \tag{5}$$

where $\kappa(\cdot, \cdot)$ is a positive semi-definite kernel function representing inner products in a high-dimensional feature space $\mathcal{F}$. The Radial Basis Function (RBF) kernel is particularly effective for nonlinear sensor data and is defined as

$$\kappa(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right), \tag{6}$$

where σ controls the spread of the Gaussian. The kernel matrix is centered as

$$\tilde{K} = K - \mathbf{1}_n K - K\mathbf{1}_n + \mathbf{1}_n K \mathbf{1}_n, \tag{7}$$

where $\mathbf{1}_n = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$. The nonlinear components are obtained from the eigendecomposition

$$\tilde{K}\alpha_k = \lambda_k \alpha_k, \tag{8}$$

with eigenvalues λ_k and normalized eigenvectors α_k. The projection of a new sample x onto the k-th nonlinear component is computed as

$$t_k(x) = \sum_{i=1}^{n} \alpha_{ki}\, \kappa(x_i, x), \tag{9}$$

allowing the representation of complex degradation trajectories that may not be linearly separable in the original space.

Multivariate Statistical Process Control (MSPC). Once PCA or KPCA has been applied, process monitoring is performed using two complementary statistics:

1. **Hotelling's T^2 statistic:**
$$T^2 = t^\top \Lambda^{-1} t, \tag{10}$$
which measures the distance of an observation within the retained component subspace. Large T^2 values indicate significant shifts in the correlation structure of the variables.

2. **Squared Prediction Error (SPE or Q statistic):**
$$Q = \|x - \hat{x}\|^2 = \|x - PP^\top x\|^2, \tag{11}$$
which quantifies the part of the observation not captured by the PCA/KPCA model, representing residual variance.

Control limits for both statistics are determined from healthy data using a 3σ criterion:

$$L_{T^2} = \overline{T^2} + 3s_{T^2}, \qquad L_Q = \overline{Q} + 3s_Q. \tag{12}$$

Observations exceeding either limit are considered out-of-control, suggesting possible anomalies or early degradation.

Model calibration and validation. The MSPC model is calibrated using the healthy early-life data of the training engines, defined as approximately the first 30% of cycles for each unit. This ensures that the model captures normal operating behavior and is not biased by degradation effects. To assess robustness, the training data are split into calibration and validation subsets, holding out 20% of the engines for validation. The PCA/KPCA model and corresponding control limits are computed from the calibration subset and tested on the validation data before application to the test sets.

Performance evaluation. The model is evaluated on the test partitions, which include true Remaining Useful Life (RUL) labels. For each engine, the T^2 and Q statistics are computed over cycles, and an alarm is triggered when

$$T^2 > L_{T^2} \quad \text{or} \quad Q > L_Q. \tag{13}$$

The first alarm is compared to the failure cycle to assess timeliness. Three performance metrics are used:

- **True Positive Rate (TPR):**
$$\text{TPR} = \frac{N_{\text{on-time alarms}}}{N_{\text{total engines}}} \times 100\%, \tag{14}$$
measuring the proportion of engines for which the model correctly anticipates failure within a predefined lead window L.

- **False Alarm Rate (FAR):**
$$\text{FAR} = \frac{N_{\text{early alarms}}}{N_{\text{total engines}}} \times 100\%, \tag{15}$$
quantifying the frequency of alarms raised too early (outside the lead window).

- **Detection Lead Time (DLT):**

$$\mathrm{DLT} = \frac{1}{N_{\mathrm{alarms}}} \sum_{i=1}^{N_{\mathrm{alarms}}} \left(\mathrm{cycle}_{\mathrm{failure},i} - \mathrm{cycle}_{\mathrm{alarm},i}\right), \tag{16}$$

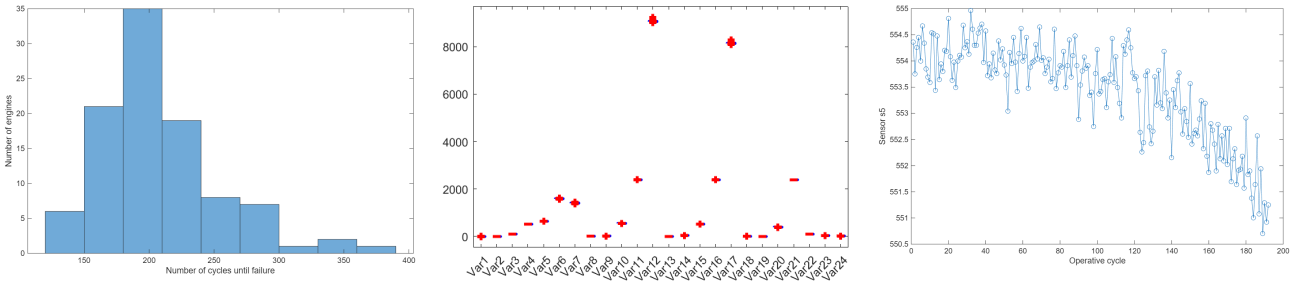
representing the average number of cycles between the first alarm and the actual failure.

In addition, sensor contribution plots are computed to identify the variables most responsible for deviations, improving model interpretability and diagnostic insight.

3 Results and Discussion

3.1 Exploratory Data Analysis and Pretreatment Results

The NASA C-MAPSS datasets (FD001–FD004) were first examined to assess data quality, temporal structure, and sensor behavior prior to modeling. Each dataset contains multiple independent engine units recorded over successive operational cycles. Preliminary inspections confirmed that the length of the degradation trajectories varies across engines, leading to unsynchronized failure times.



(a) Engine lifetimes (FD001).

(b) Sensor boxplots (FD001).

(c) Selected sensor time series (FD001).

Figure 1: Exploratory Data Analysis.

Figure 1 (a) shows the spread of failure cycles in FD001, reflecting heterogeneous operating conditions and degradation rates; the average life per dataset was computed to characterize overall durability. Data quality checks verified the absence of missing values. Figure 1 (b) highlights variance differences and outliers among sensors; constant variables were identified in FD001 and FD003 and removed to avoid numerical redundancy. Figure 1 (c) confirms the sequential nature of measurements and the progression of degradation across cycles. Overall, the datasets comprise 24 predictors (operational settings plus sensors) and many observations per engine, motivating preprocessing steps such as normalization, dimensionality reduction, and RUL computation.

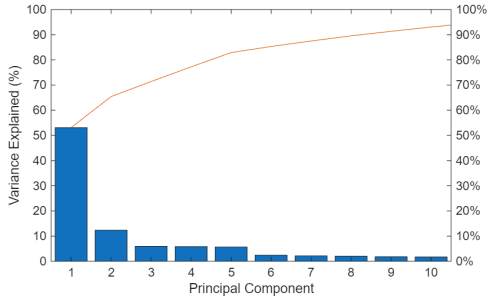
During pretreatment, RUL labels were computed for each training engine as

$$\mathrm{RUL}_{i,t} = \max(\mathrm{cycle}_i) - \mathrm{cycle}_{i,t};$$

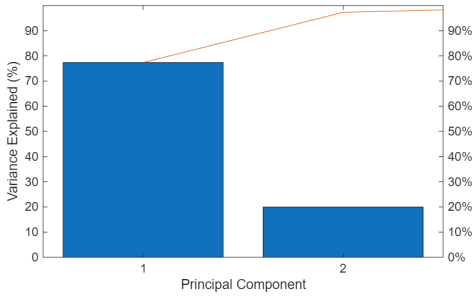
constant variables were removed and all features were standardized via z-score normalization to ensure equal weighting. Unlike many process-monitoring applications, extreme values were retained because they correspond to natural end-of-life behavior and contain essential degradation information. These steps ensured consistency between training and test data and preserved temporal integrity for modeling.

3.2 Principal Component Analysis (PCA) Results

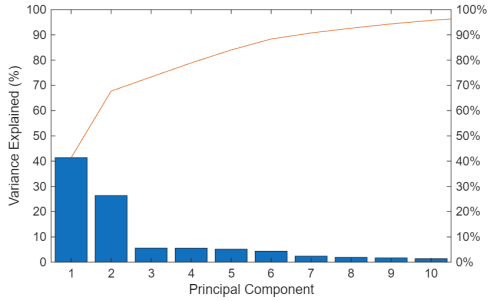
PCA was first applied to explore correlation structures and dimensionality reduction potential across the four datasets. The scree plots in Figure 2 illustrate the cumulative variance explained by the principal components.



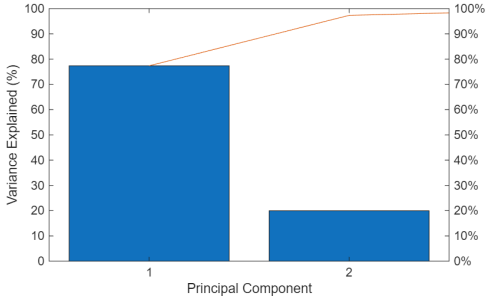
(a) FD001



(b) FD002



(c) FD003

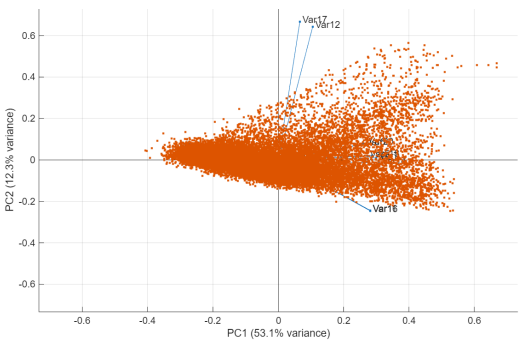


(d) FD004

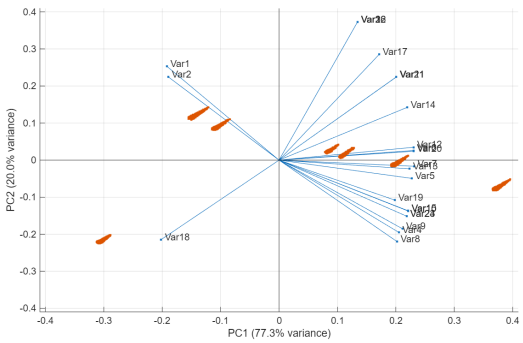
Figure 2: Scree plot for the four datasets.

In FD001, the explained variance declines gradually, with the first component capturing slightly more than half of the total variance. This indicates complex sensor interactions requiring multiple components to preserve information. By contrast, FD002 and FD004 show steep drops in variance after the first two components, implying more constrained operating conditions and higher redundancy among sensors. FD003 displays an intermediate pattern, where the first ten components already capture nearly all variance.

The PCA biplots illustrate the relationships among sensors and the structure of the data in the space of the first two principal components.

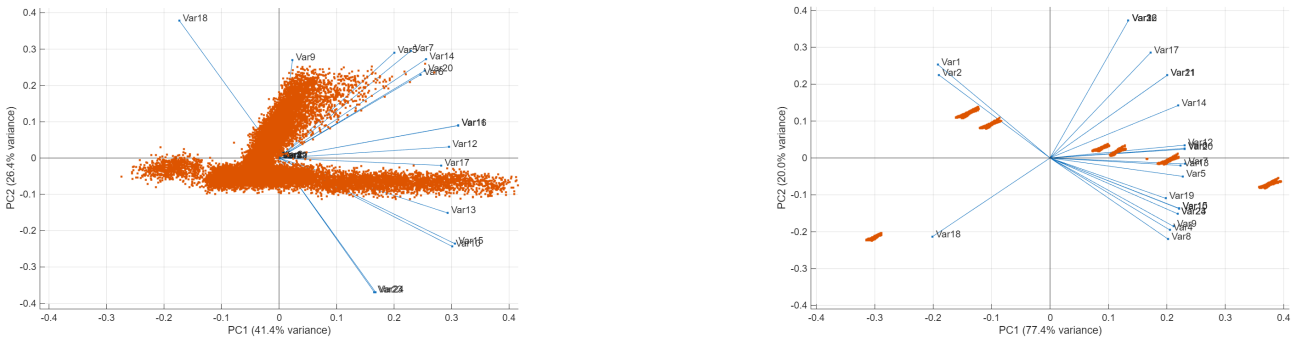


(a) FD001



(b) FD002

Figure 3: PCA biplots for datasets FD001 and FD002.



(a) FD003 (b) FD004

Figure 4: PCA biplots for datasets FD003 and FD004.

Figure 3-4 shows the PCA biplots for the four C-MAPSS subsets, isualizing both the projection of the samples and the loading vectors of the sensor variables on the first two principal components. The arrows represent the direction and strength of sensor correlations: vectors pointing in similar directions indicate positive correlations, opposite directions indicate negative correlations, and orthogonal vectors suggest independence between sensors.

Overall, the biplots reveal marked differences in structure among the datasets. In FD001 (Figure 3a), engine trajectories gradually shift along one dominant direction as the cycle number increases, forming an elongated pattern that reflects a consistent degradation trend. The orientation of the loading arrows indicates which sensors contribute most strongly to this progression, confirming that PCA captures the underlying wear process of the engines. FD003 (Figure 4a) displays a similar trajectory pattern but with a wider spread of points, indicating greater variability in sensor responses across engines.

Conversely, FD002 (Figure 3b) and FD004 (Figure 4b) exhibit multiple dense clusters rather than a single trajectory, suggesting the coexistence of different operating regimes. In these two datasets, sensors show tighter groupings of loading vectors along the first component, implying strong inter-sensor correlations and more homogeneous measurement behavior within each regime. Such patterns demonstrate that degradation and operating conditions are closely intertwined, and a single linear model may not fully capture these effects.

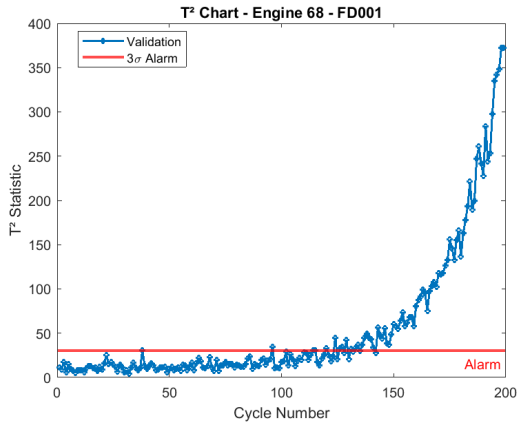
Across all datasets, sensors such as `Var7`, `Var12`, and `Var17` consistently point in similar directions, confirming redundancy among sensors that measure related physical phenomena. These variables could be grouped or partially excluded without significant information loss. This redundancy is precisely what PCA exploits, enabling dimensionality reduction while retaining the essential variance structure of the data.

From a quantitative perspective, retaining between five and ten principal components, depending on the dataset, was found to capture more than 90% of the total variance. This balance provides compact yet informative representations of the sensor space. The PCA visualization also helps identify correlated sensor groups and track the evolution of each engine through its operating cycles, making it a valuable diagnostic tool for understanding degradation dynamics. This analysis, combined with the PCA insights, provided a robust foundation for subsequent MSPC and KPCA modeling.

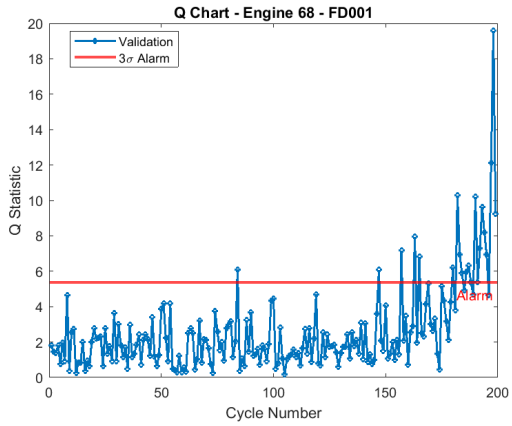
3.3 MSPC Modeling Results

A PCA-based Multivariate Statistical Process Control (MSPC) model was developed for each dataset. Following refinement, the training data were randomly split into calibration and validation subsets, reserving 20% of the engines for validation. Control limits were calibrated exclusively on healthy data (approximately the first 30% of cycles), ensuring unbiased detection of faults.

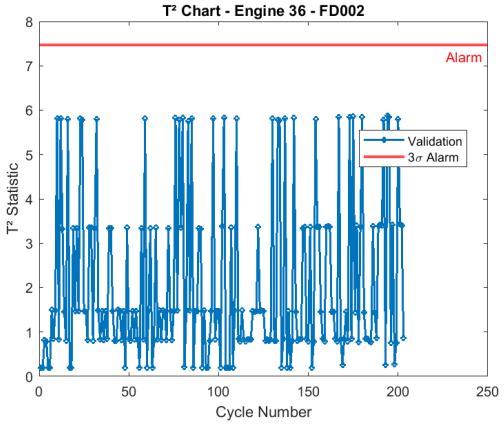
The number of retained principal components differed across datasets: FD001 = 12, FD002 = 2, FD003 = 11, and FD004 = 2. This reflects the dataset-specific correlation structures observed earlier.



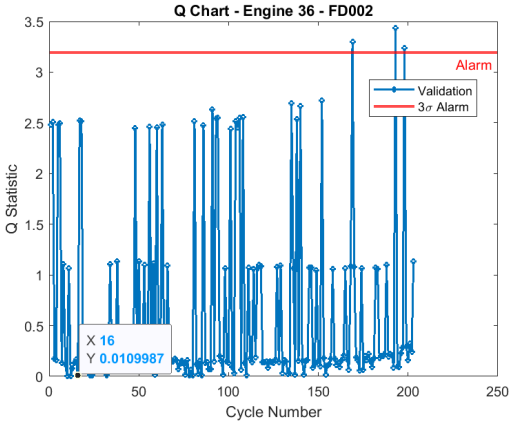
(a) FD001 T^2 chart



(b) FD001 Q chart



(c) FD002 T^2 chart



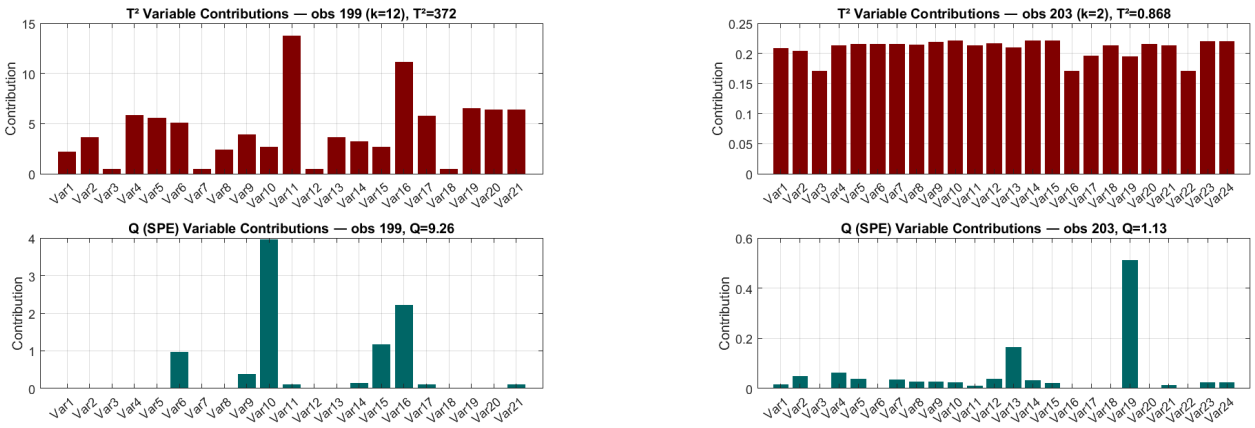
(d) FD002 Q chart

Figure 5: Control charts for FD001 and FD002.

Figure 4 displays representative control charts for FD001 and FD002. In FD001 and FD003, both T^2 and Q statistics remain low during early life and rise significantly near end-of-life, crossing the 3σ control limits and signaling incipient faults. For FD002 and FD004, similar trends are observed but with sharper spikes closer to failure, reflecting abrupt degradation.
Variable-contribution plots reveal which sensors most strongly drive the threshold exceedances. In FD001 and FD003, contributions are concentrated among a few sensors, whereas in FD002 and FD004 they are more dispersed, consistent with multiple operating regimes.

3.4 Validation and Testing Results

Model validation on the test partitions confirmed the consistency of the observed behavior. Using the calibrated PCA models and 3σ limits, both T^2 and Q statistics were computed along each test trajectory. Typically, Q triggered earlier than T^2, while T^2 captured coordinated shifts closer to failure. In several engines, alarms occurred well within the desired 10–20 cycle window, demonstrating the framework's capability for early detection.
For FD002 and FD004, alarms were less frequent. This is attributed to their simpler structure (fewer retained PCs) and abrupt degradation patterns, making threshold crossings rarer. Nev-



(a) FD001 (b) FD002

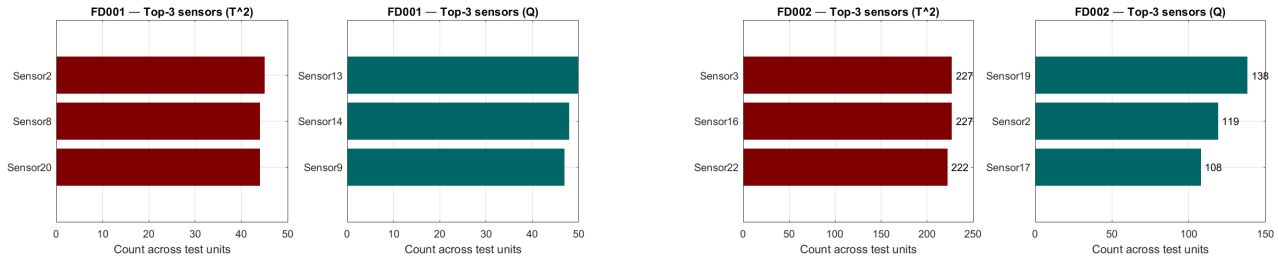
Figure 6: Contribution plots for FD001 and FD002.

ertheless, the complementary use of T^2 and Q provided robust monitoring performance across all datasets.
Table 1 summarizes the detection performance for $\alpha = 0.995$.

Table 1: Detection performance at confidence level $\alpha = 0.995$.

Dataset	**TPR (%)**	**FAR (%)**	**Mean Lead Time (cycles)**
FD001	24.0%	43.0%	78.45
FD002	24.3%	29.3%	78.54
FD003	32.0%	30.0%	88.10
FD004	16.1%	23.8%	101.48

As shown, FD001 and FD003 exhibit moderate detection performance with controlled false alarms, while FD002 and FD004 show limited sensitivity due to higher healthy variability and fewer components. A sensitivity analysis across thresholds ($\alpha \in \{0.95, 0.99, 0.995\}$) confirmed that stricter limits reduce false alarms at the cost of slightly lower true positive rates.
Figure below reports the top three sensors contributing to the first alarm for each dataset. The rankings remain stable across thresholds, typically sharing most entries between confidence levels. Consistently, T^2 emphasizes sensors defining the dominant correlation structure, while Q highlights those associated with residual variance, confirming their diagnostic complementarity.



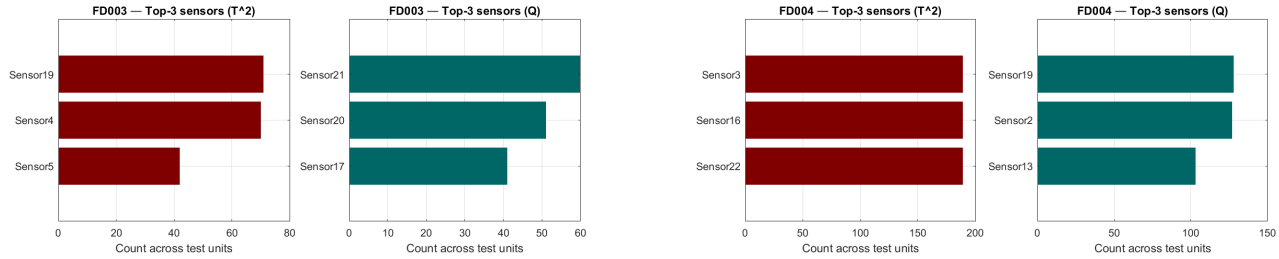


Figure 7: Top 3 sensors for each dataset at level $\alpha = 0.95$.

3.5 Kernel PCA (KPCA) Extension

Kernel Principal Component Analysis (KPCA) was introduced to account for potential nonlinear dependencies and the multimodal structure observed in FD002 and FD004. These subsets exhibit multiple operating regimes in the linear PCA biplots, motivating a nonlinear mapping to a higher–dimensional feature space. KPCA was therefore applied only to FD002 and FD004 using the RBF kernel with standardized inputs and centering in feature space. The monitoring workflow mirrored the PCA pipeline: calibration on healthy early–life data, computation of Hotelling's T^2 and Q/SPE with control limits, and evaluation on test trajectories.
In practice, the kernel mapping did not yield material gains: KPCA control charts closely resembled those of PCA, with comparable alarm timings and variability in both T^2 and Q. The detection metrics at $\alpha = 0.995$ are reported in Table 2.

Table 2: Detection metrics with KPCA (confidence level $\alpha = 0.995$) on FD002 and FD004.

Dataset	**TPR (%)**	**FAR (%)**	**Mean Lead Time (cycles)**
FD002 (KPCA)	25.5%	47.1%	86.81
FD004 (KPCA)	21.4%	36.7%	98.57

Compared with the PCA results at the same confidence level (Table 1), KPCA produced only a slight increase in TPR (FD002: $24.3\% \to 25.5\%$; FD004: $16.1\% \to 21.4\%$) accompanied by a substantial rise in FAR (FD002: $29.3\% \to 47.1\%$; FD004: $23.8\% \to 36.7\%$), while mean lead times remained of the same order. This indicates no systematic improvement in early–fault sensitivity.
The ranking of the three sensors most frequently contributing at the first alarm largely overlaps with the PCA-based analysis, indicating that kernel mapping preserves the same diagnostic structure despite the lack of performance gains.
A plausible interpretation is that (i) most degradation–relevant variance is already captured by a few linear components, (ii) degradation in FD002/FD004 tends to be abrupt and late, limiting benefits from smoother nonlinear embeddings, and (iii) multimodality reflects distinct operating regimes more than curved trajectories within one regime. Under these conditions, the extra flexibility of KPCA increases false alarms without delivering earlier or more stable detections. The PCA–based MSPC model therefore remains the preferred baseline for these datasets, as it provides a simpler and more computationally efficient solution while achieving equivalent detection accuracy and clearer interpretability of sensor contributions.

4 Conclusion

This work presented a data-driven monitoring framework for turbofan engines based on the NASA C-MAPSS datasets (FD001–FD004). The approach combined systematic pretreatment,

dimensionality reduction through PCA (and selectively KPCA), and Multivariate Statistical Process Control (MSPC) using Hotelling's T^2 and Q/SPE charts. PCA analysis revealed consistent degradation trajectories in FD001 and FD003, while FD002 and FD004 showed multiple operating regimes and higher redundancy among sensors.

The PCA-based MSPC model effectively captured normal behavior and signaled degradation in advance, especially for FD001 and FD003, achieving moderate true-positive rates and stable lead times. Contribution plots identified a few sensors as dominant indicators of failure, confirming good interpretability. In contrast, applying KPCA to FD002 and FD004 did not lead to significant performance improvements: detection metrics remained comparable, with slightly higher false-alarm rates.

Overall, results indicate that degradation patterns in the C-MAPSS data are largely governed by linear correlations, making PCA-based MSPC a robust and efficient baseline.

References

[1] H. Hoffmann, "Kernel PCA for novelty detection," *Pattern Recognition*, vol. 40, no. 3, pp. 863–874, 2007.

[2] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer, New York, 2002.

[3] T. Kourti and J. F. MacGregor, "Process analysis, monitoring and diagnosis, using multivariate projection methods," *Chemometrics and Intelligent Laboratory Systems*, vol. 28, no. 1, pp. 3–21, 1995.

[4] S. J. Qin, "Statistical process monitoring: basics and beyond," *Journal of Chemometrics*, vol. 17, no. 8-9, pp. 480–502, 2003.

[5] B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[6] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, Cambridge, 2004.

[7] B. M. Wise and N. B. Gallagher, "The process chemometrics approach to process monitoring and fault detection," *Journal of Process Control*, vol. 6, no. 6, pp. 329–348, 1996.