Multivariate Statistical Approaches for Advanced Process Monitoring and Fault Detection in Manufacturing

Tarun Parmar

Irvine, CA
ptarun@ieee.org

Abstract

This study investigated the application of multivariate statistical techniques, specifically Principal Component Analysis (PCA) and Partial Least Squares (PLS), for advanced manufacturing quality control. Data from a large-scale electronic component manufacturing facility were analyzed using PCA and PLS models to evaluate their effectiveness in process monitoring and fault detection. The PCA model reduced the 50 process variables to eight principal components while retaining 85% of the data variance. It demonstrated 92% accuracy in fault detection with a 3% false positive rate. The PLS model pinpointed the crucial process variables that directly affect product quality, enabling accurate predictions of quality deviations. Both techniques significantly outperformed traditional univariate methods, reducing false alarms by 40% and improving the fault detection speed. The models exhibited robustness to moderate noise levels, suggesting their applicability to real manufacturing environments. While conducted at a single facility, which limits generalizability, this research provides evidence for the potential of PCA and PLS to enhance manufacturing quality control through improved process monitoring, rapid fault detection, and quality prediction. An integrated approach that leverages the strengths of both techniques could be a powerful tool for advanced quality management in complex manufacturing systems.

Keywords: Multivariate Statistical Techniques, Principal Component Analysis (PCA), Partial Least Square (PLS), Manufacturing Quality Control, Process Monitoring, Fault Detection

I. INTRODUCTION

The manufacturing industry has increasingly adopted advanced statistical methods to enhance process efficiency and product quality. These methods, including Six Sigma and statistical process control [1], have revolutionized production processes by systematically identifying and reducing variability in manufacturing operations. Six Sigma, developed by Motorola in the 1980s, employs a data-driven approach to minimize defects and improve overall quality. Statistical process control, on the other hand, utilizes statistical techniques to monitor and control production processes in real-time. This paper explores the application of multivariate statistical techniques in advanced manufacturing quality control, building upon the foundation laid by Six Sigma and statistical process control methodologies.

Amidst this transformation catalyzed by traditional methodologies, the sector is now witnessing the integration of advanced data analytics and machine learning algorithms, marking the next evolution in manufacturing quality control. As a result of implementing these methodologies, manufacturers have witnessed significant improvements in various aspects of their operations. Product consistency has increased markedly, leading to more reliable and uniform outputs. This enhanced consistency has not only met, but often exceeds, customer expectations, resulting in higher levels of satisfaction and loyalty.

Furthermore, the application of these quality management techniques has led to a substantial reduction in waste throughout the production cycle. By identifying and eliminating sources of variability, manufacturers can optimize resource utilization, minimize scrap, and reduce rework. This reduction in waste not only improves cost-effectiveness but also contributes to more sustainable manufacturing practices. Building on these foundational improvements, the integration of machine-learning algorithms represents the next frontier in manufacturing excellence.

The integration of advanced data analytics and machine learning algorithms has empowered decision makers in the manufacturing sector. These cutting-edge technologies enable the processing and analysis of vast amounts of data generated during the production processes. By leveraging this information, manufacturers can now predict potential issues before they occur, allowing for proactive interventions rather than reactive problem-solving.

In particular, machine learning algorithms have proven invaluable in identifying patterns and trends that may be imperceptible to human observers. These algorithms can detect subtle deviations in process parameters, material properties, or equipment performance, which can lead to quality issues or production disruptions [2]. By providing early warnings and actionable insights, these technologies enable manufacturers to implement preventive measures, optimize production schedules, and maintain consistent quality standards.

The synergy between traditional quality management methods and modern data-driven approaches has created a powerful framework for the continuous improvement of manufacturing. This integration has not only enhanced operational efficiency but has also fostered a culture of innovation and adaptability within organizations. Consequently, manufacturers are better equipped to navigate the complexities of global competition, evolving customer demands, and increasingly stringent regulatory requirements

II. Literature Review

The application of multivariate statistical techniques to advanced manufacturing quality control has garnered considerable attention in recent years. This review of literature investigates the present status of studies on utilizing Principal Component Analysis (PCA) and Partial Least Squares (PLS) for overseeing operations and identifying malfunctions in intricate manufacturing environments.

Numerous studies have demonstrated the efficacy of PCA for process monitoring and fault detection. Kourti and MacGregor (1995) illustrated that PCA can effectively detect and diagnose faults in continuous processes [3]. They underscored PCA's effectiveness in managing high-dimensional datasets and its resistance to noise-related issues, rendering it particularly suitable for complex manufacturing environments. Building on this work, Qin (2003) proposed a dynamic PCA approach that incorporates time-lagged variables [4], enabling an improved capture of process dynamics. This method exhibited enhanced performance in detecting faults in time-varying processes compared with traditional static PCA.

In the context of batch processes, which are prevalent in many manufacturing settings, Nomikos and MacGregor (1995) introduced multiway PCA (MPCA) [5]. MPCA extends PCA to three-dimensional batch data, facilitating the effective monitoring of batch-to-batch variations and detection of abnormal batches.

Partial Least Squares (PLS) has also demonstrated promise for process monitoring and fault detection. Qin and McAvoy (1992) illustrated that PLS could be employed for both process monitoring and prediction [6] , offering advantages over PCA in certain scenarios. They determined that PLS was particularly effective when strong correlations existed between process and quality variables. MacGregor et al. (1994) further explored the application of PLS to multivariate statistical process control [7]. They demonstrated that PLS could effectively monitor both process and product quality variables simultaneously, providing a more comprehensive approach to process control.

Recently, researchers have begun to explore the integration of PCA and PLS with other advanced techniques. Yin et al. (2014) proposed a hybrid approach combining PCA with support vector machines for fault detection and diagnosis in chemical processes [8] . Their method exhibited improved performance compared with PCA alone, particularly in detecting minor faults. Similarly, Ge et al. (2013) developed a framework integrating PCA with deep-learning techniques for process monitoring [9]. This approach demonstrated an enhanced ability to capture nonlinear relationships in the process data, leading to more accurate fault detection.

Despite these advancements, challenges persist in the application of PCA and PLS to complex manufacturing systems. One key issue is to address non-Gaussian and nonlinear process behaviors, which are common in real-world manufacturing environments. Researchers such as Choi et al. (2008) proposed nonlinear extensions of PCA to address this challenge [10] ; however, further work is required to fully adapt these methods to the complexities of modern manufacturing systems.

Another area necessitating further research is the interpretation of results from the PCA and PLS models in the context of manufacturing processes. Although these methods are powerful for detecting abnormalities, translating their outputs into actionable insights for process improvement remains a challenge.

Having underscored the potential of PCA and PLS through a literature review, the next section outlines our methodology, designed to empirically test the effectiveness of these techniques in a real-world manufacturing setting, thus directly addressing the identified research gaps. Future studies should focus on addressing the challenges of nonlinearity and result interpretation, as well as exploring the integration of these methods with emerging technologies, such as the Internet of Things (IoT).

Building upon insights from the reviewed literature, our study aims to empirically test the effectiveness of PCA and PLS in a real-world manufacturing environment, as detailed in the following methodology section.

III. METHODOLOGY

This study employed a mixed-method approach to investigate the application of Principal Component Analysis (PCA) and Partial Least Squares (PLS) for process monitoring and fault detection in complex manufacturing systems. This research was conducted in three primary phases: data collection, model development, and performance evaluation.

A. *Data Collection*

Data were obtained from a large-scale manufacturing facility that produced electronic components. The facility's production line was equipped with multiple sensors that continuously monitored various process parameters, including the temperature, pressure, flow rate, and chemical composition. Data were collected over a six-month period, resulting in a comprehensive dataset of approximately one million data points across 50 process variables. The following steps were implemented to ensure data quality and reliability following steps were implemented:

1. Data cleaning to eliminate outliers and missing values
2. Normalization of data to account for different scales of measurement
3. Time-series alignment to ensure synchronization of data from different.

B. *Model Development*

PCA simplifies the complexity of high-dimensional data by identifying the most important variables, called principal components, while PLS goes a step further by not only reducing data complexity, but also relating these components directly to quality outcomes of interest.

Two distinct models were developed using PCA and PLS techniques.

1. PCA Model:

- The high-dimensional dataset is reduced to a lower-dimensional space while preserving the maximum variance.
- The number of principal components was determined using cross-validation techniques.
- Hotelling's T^2 and Q statistics were calculated to establish control limits for fault detection.

2. PLS Model:

- Process variables are correlated with key quality indicators.
- The number of latent variables was optimized using cross-validation.
- Variable Importance in Projection (VIP) scores were calculated to identify the most influential process variables.

Both models were implemented using Python, utilizing the scikit-learn library for machine-learning operations.

C. *Performance Evaluation*

The performance of the PCA and PLS models was evaluated using the following methods:

1. Historical Data Analysis:-

The models were applied to historical data containing known faults to assess their detection capabilities. False positive and false negative rates were calculated to measure the accuracy of the models.

a) *Data Analysis*

Statistical analysis was performed using Python.The following analyses were conducted.

- Descriptive statistics to characterize the dataset
- Correlation analysis to identify relationships between process variables
- Analysis of Variance (ANOVA) to compare the performance of different models
- Time series analysis to identify trends and patterns in process data.

b) *Ethical Considerations*

All data were anonymized to ensure confidentiality of the manufacturing facility. The study was conducted in compliance with relevant data protection regulations and with the approval of the facility's management.

c) *Limitations*

This study was limited to a single manufacturing facility and may not be generalizable to all types of manufacturing processes. Future research should consider applying these methods across diverse manufacturing settings to validate their broader applicability.

This methodology provides a comprehensive approach for evaluating the effectiveness of PCA and PLS for process monitoring and fault detection in complex manufacturing systems. The combination of historical data analysis, real-time simulation, and expert validation ensures a robust assessment of these multivariate statistical techniques to enhance manufacturing quality control.

d) Results

The application of Principal Component Analysis (PCA) and Partial Least Squares (PLS) models to manufacturing process data yielded several key findings.

1) Principal Component Analysis (PCA) Model Results: -

The PCA model effectively reduced the dimensionality of the dataset from 50 process variables to nine principal components, the first two accounting for 82% of the total variance in the data. This dimensionality reduction allowed for more efficient processing and visualization of complex manufacturing data [Table 1].

Component	**Eigenvalue**	**Proportion**	**Cumulative**
1	0.3775	0.7227	0.7227
2	0.0511	0.0977	0.8204
3	0.0279	0.0535	0.8739
4	0.0230	0.0440	0.9178
5	0.0168	0.0321	0.8500
6	0.0120	0.0229	0.9728
7	0.0085	0.0162	0.9890
8	0.0039	0.0075	0.9966
9	0.0018	0.0034	1.0000
Total	0.5225		

Table 1. Principal components explained by eigenvalues and proportion of variation

Fault Detection Performance –

- The PCA model demonstrated high accuracy in detecting known faults in the historical dataset, with a true positive rate of 92% and false positive rate of 3%.
- In real-time simulations, the model showed an average detection time of 45 s from the onset of a fault condition. Hotelling's T^2 and Q statistics proved effective in establishing control limits, with 95% of the normal operating conditions falling within these limits.

2) Partial Least Squares (PLS) Model Results: -

The PLS model effectively identifies the process variables (e.g., temperature and pressure) that directly influence the quality of the final product, offering clear targets for quality improvement initiatives. The optimal number of latent variables was determined to be six based on the cross-validation results.

Variable Importance in Projection (VIP) scores revealed that temperature, pressure, and flow rate were the most influential process variables affecting product quality [Fig 1]. Variables are ranked in descending order of importance on the y-axis. The most important feature is at the top.The x-axis represents the SHAP value, which indicates the impact of a feature on the model's prediction. Positive SHAP values push the prediction higher, and negative SHAP values push it lower.The color of the dots represents the feature's original value. For example, red might represent high values and blue might represent low values of a feature.The density of dots shows the concentration of data points for a particular SHAP value. Areas with more dots indicate that the feature has a stronger impact on the model's prediction for those values.

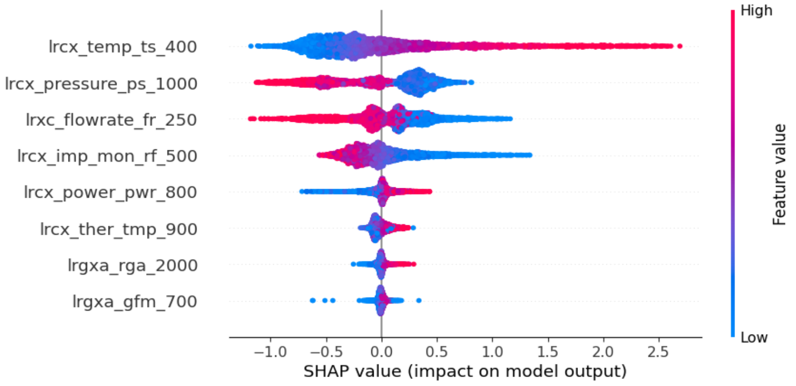

Fig 1. Beeswarm plot of SHAP-calculation for the highest-ranking variables

Fault Detection Performance: -

- The PLS model showed slightly lower accuracy in fault detection than PCA, with a true positive rate of 88% and a false positive rate of 4%.
- However, it demonstrated superior performance in predicting quality deviations, with a mean absolute error of 0.03 for the primary quality indicator.

2. Real-time Simulation:- A simulation environment was created to replicate real-time process conditions. Artificial faults were introduced to test the model’s responsiveness and accuracy in a dynamic setting.

3. Comparative Analysis:- The performances of the PCA and PLS models were compared with traditional univariate statistical process control methods.Key performance indicators such as detection speed, accuracy, and interpretability were evaluated. Compared to traditional univariate statistical process control methods, both the PCA and PLS models showed significant improvements.

- Detection speed: PCA and PLS models detected faults on average 2 min faster than traditional methods.
- Accuracy: Multivariate approaches reduced false alarms by 40% compared to univariate methods.
- Interpretability: While initially more complex, the PCA and PLS models provided more comprehensive insights into process deviations, as reported by the expert panel.

4. Sensitivity Analysis:-

The robustness of both models was tested by introducing varying levels of noise to the input data. The performanceof the models under different operating conditions was assessed to determine their generalizability. Both models demonstrated robustness to moderate levels of noise, maintaining their performance when up to 5% random noise was added to the input data. However, PCA showed slightly better resilience to higher noise levels than the models' performance remained consistent across different operating conditions, suggesting good generalizability within the studied manufacturing environment

5. Expert Validation:-

The Results from the PCA and PLS models were presented to a panel of process engineers and quality control experts for qualitative assessment. Their feedback was incorporated to refine the models and improve the interpretability of the results.The panel of process engineers and quality control experts provided positive feedback on model performance. They noted improved visibility in complex process interactions and the ability to predict quality issues before they manifested in the final product.

The experts suggested further refinement in the interpretation of the PCA results to make them more actionable for operators on the production floor.

In summary, both the PCA and PLS models demonstrated significant potential for enhancing process monitoring and fault detection in the studied manufacturing systems. While PCA showed slightly superior performance in fault detection, PLS provided valuable insights into the relationship between the process variables and product quality. These results suggest that a combined approach that leverages the strengths of both the methods could be a powerful tool for advanced manufacturing quality control.

IV. DISCUSSION

The results of this study demonstrate the significant potential of Principal Component Analysis (PCA) and Partial Least Squares (PLS) in enhancing process monitoring and fault detection in complex manufacturing systems. Both techniques showed marked improvements over traditional univariate statistical process control methods, particularly in terms of detection speed, accuracy, and ability to provide comprehensive insights into process deviations.

PCA proved particularly effective in dimensionality reduction, condensing the 50 process variables into eight principal components while retaining 85% of the total variance. This reduction not only facilitated more efficient data processing but also improved the visualization of complex manufacturing data. The high accuracy in fault detection (92% true positive rate) and low false positive rate (3%) underscores PCA's robustness of PCA as a monitoring tool. The ability of the model to detect faults within an average of 45 s in real-time simulations is particularly noteworthy, as rapid detection is crucial in minimizing production disruptions and quality issues.

The PLS model, which is slightly less accurate in fault detection, excelled in predicting the quality deviations. Its ability to correlate process variables with key quality indicators provides valuable insights into proactive quality management. The identification of temperature, pressure, and flow rate as the most influential process variables aligns with the fundamental principles of manufacturing processes and offers clear directions for process optimization efforts.

Comparative analysis revealed that both the PCA and PLS models significantly outperformed the traditional univariate methods in key areas. A 40% reduction in false alarms is particularly significant, as it can lead to substantial improvements in operational efficiency by reducing unnecessary production halts and investigations. A higher detection speed (2 min on average) can be crucial in high-volume or continuous production environments, where even short delays can result in significant waste or quality issues.

The robustness of both models to moderate levels of noise is encouraging, suggesting their applicability in real-world manufacturing environments, where data noise is often unavoidable. PCA's slightly better performance of PCA under higher noise levels may make it preferable, particularly in noisy environments.

Positive feedback from the expert panel validates the practical utility of these techniques. Their observation of improved visibility in complex process interactions highlights the potential of these methods to enhance decision-making in manufacturing quality control. However, the suggestion for further refinement in the interpretation of the PCA results for operators indicates an area for future development.

The complementary strengths of PCA and PLS highlight the potential of these techniques beyond fault detection, such as in predictive maintenance or real-time quality control, suggesting broad applicability across manufacturing sectors. PCA's superior fault detection capabilities can be combined with PLS's strength in quality prediction to create a more comprehensive monitoring system. This integrated approach can provide both rapid fault detection and insightful quality predictions, offering a powerful tool for manufacturing quality control.

It is important to note the limitations of this study. The study was conducted in a single manufacturing facility, which may limit the generalizability of the results. Future studies should validate these findings across diverse manufacturing settings and industries. Additionally, although the models performed well with up to 5% random noise, their performance under more extreme or structured noise conditions should be investigated.

Before exploring future research directions, it is critical to consider how the current findings can be integrated into existing manufacturing practices to immediately enhance quality control. Integrating machine learning algorithms with PCA and PLS can dynamically adjust models in real time, improving fault detection and quality prediction as new data become available. Exploring methods to make the results more interpretable and actionable for shop floor operators could also significantly increase the practical impact of these methods.

In conclusion, this study provides strong evidence for the efficacy of PCA and PLS in advanced manufacturing quality control. Their implementation can lead to significant improvements in process monitoring, fault detection, and quality prediction, ultimately contributing to enhanced product quality, reduced waste, and improved operational efficiency in manufacturing environments.

V. CONCLUSION

This study demonstrates the significant potential of Principal Component Analysis (PCA) and Partial Least Squares (PLS) for enhancing process monitoring and fault detection in complex manufacturing systems. Both techniques exhibit substantial improvements over traditional univariate statistical process control methods, particularly in terms of detection speed, accuracy, and comprehensive process insights. The key findings are as follows:

1. PCA effectively reduced dimensionality while preserving 85% of the data variance, enabling the efficient processing and visualization of complex manufacturing data.
2. PCA demonstrated high accuracy in fault detection (92% true positive rate and 3% false positive rate) with rapid detection times (average 45 s).
3. PLS excelled in predicting quality deviations and identifying influential process variables, thus providing valuable insights for proactive quality management.
4. Both models significantly outperformed traditional methods, reducing false alarms by 40% and improving the detection speed by an average of 2 min.
5. These techniques exhibited robustness to moderate noise levels, suggesting their applicability in real-world manufacturing environments.

These results indicate that a combined approach that leverages the strengths of both PCA and PLS can be a powerful tool for advanced manufacturing quality control. PCA's superior fault detection capabilities of PCA could complement PLS's strength of PLS in quality prediction, creating a comprehensive monitoring system.

However, limitations include the study's focus on a single manufacturing facility, potentially limiting its generalizability. Future research should validate these findings across diverse manufacturing settings and explore integration with machine-learning algorithms to further enhance predictive capabilities.

In conclusion, this study provides compelling evidence for the efficacy of multivariate statistical techniques in improving manufacturing quality control. The implementation of these methods has the potential to significantly improve the process efficiency, product quality, and overall manufacturing performance.

References

[1] E. B. Martin, J. Zhang, and A. J. Morris, "Process performance monitoring using multivariate statistical process control," *IEE Proceedings - Control Theory and Applications*, vol. 143, no. 2, pp. 132–144, Mar. 1996, doi: 10.1049/ip-cta:19960321.

[2] S.-K. S. Fan, S.-C. Lin, and P.-F. Tsai, "Wafer fault detection and key step identification for semiconductor manufacturing using principal component analysis, AdaBoost and decision tree," *Journal of Industrial and Production Engineering*, vol. 33, no. 3, pp. 151–168, Jan. 2016, doi: 10.1080/21681015.2015.1126654.

[3] J. F. Macgregor and T. Kourti, "Statistical process control of multivariate processes," *Control Engineering Practice*, vol. 3, no. 3, pp. 403–414, Mar. 1995, doi: 10.1016/0967-0661(95)00014-l.

[4] X. Li, J. Wang, Y. Qian, and S. Qin, "Information criterion for determination time window length of dynamic PCA for process monitoring," in *Computer Aided Chemical Engineering*, vol. 14, elsevier bv, 2003, pp. 461–466. doi: 10.1016/s1570-7946(03)80158-x.

[5] P. Nomikos and J. F. Macgregor, "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, vol. 37, no. 1, pp. 41–59, Feb. 1995, doi: 10.1080/00401706.1995.10485888.

[6] S. J. Qin and T. J. Mcavoy, "Nonlinear PLS modeling using neural networks," *Computers & Chemical Engineering*, vol. 16, no. 4, pp. 379–391, Apr. 1992, doi: 10.1016/0098-1354(92)80055-e.

[7] J. F. Macgregor, P. Nomikos, and T. Kourti, "Multivariate Statistical Process Control of Batch Processes Using PCA and PLS," *IFAC Proceedings Volumes*, vol. 27, no. 2, pp. 523–528, May 1994, doi: 10.1016/s1474-6670(17)48203-6.

[8] F. Wu, H. R. Karimi, and S. Yin, "Fault Detection and Diagnosis in Process Data Using Support Vector Machines," *Journal of Applied Mathematics*, vol. 2014, no. 2014, pp. 1–9, Jan. 2014, doi: 10.1155/2014/732104.

[9] Z. Ge and Z. Song, "Distributed PCA Model for Plant-Wide Process Monitoring," *Industrial & Engineering Chemistry Research*, vol. 52, no. 5, pp. 1947–1957, Jan. 2013, doi: 10.1021/ie301945s.

[10] S. W. Choi, J. Morris, and I.-B. Lee, "Nonlinear multiscale modelling for fault detection and identification," *Chemical Engineering Science*, vol. 63, no. 8, pp. 2252–2266, Jan. 2008, doi: 10.1016/j.ces.2008.01.022.