PCA-Based Manufacturing Analytics for a Realtime Automated Filling Line

Name: Ruvimbo Chiweshe

Date: 12 January 2026

Title: PCA Based Manufacturing Analytics for a Realtime Automated Filling Line

PCA-Based Manufacturing Analytics for a Realtime Automated Filling Line.

TABLE OF CONTENTS

1.0 INTRODUCTION 2
1.1 Scope of the project 2
1.2 Objectives 2
1.3 Literature Review: 2
1.3.1 Foundations of Principal Component Analysis (PCA) 2
1.3.2 Applications of PCA 3
1.3.4 Abbott Ireland Diagnostics Division, Sligo (AIDD Sligo) 4
2.2.1 Handling Missing Values 5
2.2.2 Scaling and Standardisation 5
2.2.3 Data Export for Dashboard Integration 5
3.0 RESULT ANALYSIS 6
3.1 Principal Component Structure and Variance Explained 6
3.2 PC1: Mechanical Stability and Seal Integrity 6
3.3 PC2: Filling Accuracy 6
3.4 Key Performance Indicators (KPIs) analysis 6
3.5 Hotelling's T^2: Systematic Shifts Analysis 6
3.6 Q-Residuals: Investigation of Sudden Disturbances 7
3.7 Cluster Structure and Operational Modes 7
3.8 Anomality analysis 7
3.9 Summary of key findings 7
4.0 DISCUSSION AND CONCLUSION 8
5.0 REFERENCES 10

PCA-Based Manufacturing Analytics for a Realtime Automated Filling Line.

1.0 INTRODUCTION

1.1 Scope of the project

The purpose of this project is to apply principal component analysis (PCA) to the filling line data at Abbott Diagnostics Division (AIDD), Sligo. Through a reduction of dimensions and identification of key variables, PCA will reveal hidden patterns and sources of variability in the manufacturing process thus supporting manufacturing analytics and continuous improvement.

1.2 Objectives

This project aims to:

1. Define statistical benchmarks or normal operating ranges for the new filling line.
2. Identify variations in the process that may result in rejections.
3. Use the PCA outputs to inform predictive or preventative maintenance.
4. Provide insights for lean six sigma and business excellence ideas.
5. Translate PCA outcomes to dashboards and visualisations.

1.3 Literature Review:

1.3.1 Foundations of Principal Component Analysis (PCA)

Large data sets can be challenging to analyse. Principal component analysis (PCA) is an efficient method in analysing big data sets. PCA is a statistical method used to identify patterns in data in a way that expresses similarities and differences while reducing its dimensions. The outcome features of PCA representative of the original data in lower dimensions are referred to as principal components (Kherif & Latypova, 2020). PCA utilises key linear algebra concepts namely standardisation, covariance, eigenvectors and eigenvalues to account for most of the variability of the given dataset. This helps increase the interpretability of data whilst reducing the loss of data (Greenacre et al., 2022).

Standardisation of data is done in such a way that it has mean of 0 and standard deviation of 1 which allows the equal distribution of all features to the analysis. The covariance matrix is essential in illustrating the relationship between the features while eigenvectors and eigenvalues are the principal components of the data (Machine Learning Plus, 2023).

Procedurally, PCA starts by deducting the mean from each data dimension, followed by computing the covariance matrix and calculating the eigenvectors and eigenvalues of the covariance matrix. Once computed, the eigenvalues are ordered in ascending order and a given number of them is selected to form a feature vector (Greenacre et al., 2022).

The eigenvector with the highest eigenvalue illustrates the most significant principal component. Lastly the transpose of the feature vector is multiplied with the transpose of the original data set. The result from this is usually a biplot where original data is expressed as vectors (Smith, 2002).

1.3.2 Applications of PCA

There are several applications of PCA across various fields. Examples include the study of stock exchange prices in relation to the time of the year and common financial habits of people around that time for example from July to November most people start working hard in preparation of the new year which results in an increase in stock prices (Li & Qin, 2024). In healthcare, PCA is being utilised in pharmaceutical research and drug discovery for example at the 2013 AIChE Annual Meeting held in San Francisco Huiguan Wu presented on how PCA was useful in distinguishing the chemical and physical properties of dosage forms (Wu, 2013).

Modern day automated machines tend to generate vast amounts of real-time data that tend to reflect missing values, high correlation or low signal to noise ratio. While helpful, this data can be difficult interpret. Clarify Ferer illustrates how a PCA based multivariate statistical process control bridges lowers complexity, highlights drivers of variation in automotive manufacturing ultimately bridging the gap between data availability and data -driven decision making (Ferer, 2007). Through PCA, the complexity of data is reduced, and sources of variability are identified which can be used to provide actionable insights that drive business excellence and continuous improvement (Guo and Lu, 2022).

While PCA has been applied extensively in fields such a healthcare, finance and automotive manufacturing, its potential in the diagnostics manufacturing industry remains unmapped. At Abbott Ireland Diagnostics Division in Sligo, the filling lines is an ideal opportunity where PCA can be initiated to lower data complexity, indicate sources of variability and determine statistical scales for operating excellence. This project therefore aims to bridge this gap by demonstrating how PCA can transform real time automated machine data into actionable insights for continuous improvement and business excellence in diagnostics manufacturing.

1.3.4 Abbott Ireland Diagnostics Division, Sligo (AIDD Sligo)

AIDD Sligo was first established in 1994 to produce or manufacture bulk reagent solutions and later diagnostic reagent kits for the phased out IMx and AxSYM platforms. To date it has expanded to become the second largest manufacturing facility for diagnostic products for both the Architect and Alinity platforms globally. AIDD, Sligo manufactures bulk reagent operations (BRO) used in the chemiluminescent microparticle immunoassay technology (CMIA). These include the wash buffers, trigger, and pre-trigger solutions. AIDD, Sligo is the only site that manufactures BRO products globally (Chiweshe, 2023).

Additionally, AIDD Sligo manufactures diagnostic reagent operations (DRO) which are the key bioreagents that combine with patient blood or serum to detect and capture the antigen, antibody, or analyte in the patient's blood when screening for diseases and various medical conditions (Abbott, 2025). These reagents are filled and packaged into bottles on automated filling lines that record real time process data. Analysing this data provides opportunities to design preventive strategies, enhance quality assurance and incorporate continuous improvement into daily operations.

PCA-Based Manufacturing Analytics for a Realtime Automated Filling Line.

2.0 METHODOLOGY

2.1 Data Collection

The dataset used in this study was obtained from the automated filling line at Abbott Ireland Diagnostics Division (AIDD), Sligo. The filling line records real-time process parameters such as temperature, pressure, flow rate, torque, and machine cycle metrics. These data are logged at high frequency and exported from the equipment historian system as CSV files. Only numerical variables relevant to process performance were selected for PCA modelling. Non-process metadata (e.g., timestamps, batch identifiers) were retained for filtering and traceability but excluded from the PCA computation.

2.2 Data Preprocessing

Data preprocessing was conducted in Python using pandas and scikit-learn. The following steps were applied:

2.2.1 Handling Missing Values

Missing values were analysed and variables with missing values were excluded. The remaining gaps were imputed using mean substitution.

2.2.2 Scaling and Standardisation

Since PCA is sensitive to variable magnitude, all numerical features were standardised using the StandardScaler method:

$$z = \frac{x - \mu}{\sigma}$$

This ensured each variable contributed equally to the covariance structure.

2.2.3 Data Export for Dashboard Integration

Cleaned and scaled datasets were saved into a structured project directory (data/processed/) to support reproducibility and integration with the Streamlit dashboard.

<u>**PCA-Based Manufacturing Analytics for a Realtime Automated Filling Line.**</u>

3.0 RESULT ANALYSIS

The visual outputs for the below mentioned results are presented in the accompanying in the Stream lit Dashboard.

3.1 Principal Component Structure and Variance Explained

The PCA model reduced the dimensionality of the filling line dataset while retaining **82% of the total variance** within the first two principal components. PC1 and PC2 accounted for approximately **60%** and **22%** respectively. This suggests that these components capture the dominant patterns of variation in the process.

3.2 PC1: Mechanical Stability and Seal Integrity

Variables such as total collapse, weld force, and weld time showed the strongest loadings on PC1. This suggests that PC1 primarily reflects the mechanical performance of the sealing operation. High contributions from these variables indicate that PC1 is sensitive to changes in weld quality, seal formation, and closure integrity.

3.3 PC2: Filling Accuracy

PC2 was driven mainly by tare weight and gross weight, indicating that this component captures variation related to filling precision and material dosage. This separation of mechanical and filling behaviours across PC1 and PC2 provides a clear multivariate structure for monitoring the process.

3.4 Key Performance Indicators (KPIs) analysis

Across the 7,828 parts analysed, **516 observations (6.59%)** exceeded one or more PCA control limits, indicating periods of abnormal behaviour.

3.5 Hotelling's T^2: Systematic Shifts Analysis

A significant excursion in T^2 values occurred shortly after 03:00 on 18 December 2025 with values approaching **4 000**. This represents a substantial deviation within the PCA model space and suggests a systematic shift affecting multiple variables simultaneously.

3.6 Q-Residuals: Investigation of Sudden Disturbances

A spike in Q-residuals was observed at 01:15 on 18 December 2025 indicating a deviation not captured by the PCA model. Together, T^2 and Q-residuals provided complementary insights into both systematic and unpredictable process deviations.

3.7 Cluster Structure and Operational Modes

The PCA score plot revealed three distinct clusters, indicating multiple modes of normal operation. **Cluster 0 and Cluster 2** formed the primary normal operating region which is most likely the stable production conditions across different batches. **Cluster 1** appeared as a separate group with a noticeable negative shift along PC1.

3.8 Anomality analysis

The anomalies concentrated within Cluster 1 showed a strong negative drift on PC1. This pattern suggests that the affected parts experienced **seal integrity issues**, like insufficient weld force and potential leakers. This indicates that sealing mechanics, rather than filling accuracy, were the dominant contributors to process failures.

3.9 Summary of key findings

Metric	Value / Observation	Interpretation
Total anomalies	516 parts (6.59%)	Percentage of parts outside PCA control limits.
Dominant failure mode	Negative drift on PC1	Potential instability in welding or sealing operations
Key event	T^2 spike at 03:00 on December 18, 2025.	Suggests systematic shift for example due to calibration change
Secondary event	Q-residual spike at 01:15 on December 18, 2025.	Illustrates a sudden disturbance or mechanical fault.

4.0 DISCUSSION AND CONCLUSION

The PCA analysis provided an overview of the multivariate behaviour of the filling line. Through reducing the dataset to two principal components that captured 82% of the total variance, the model successfully separated the process into meaningful dimensions: **mechanical sealing performance** (PC1) and **filling accuracy** (PC2). This separation is valuable for a diagnostics manufacturing, where both sealing integrity and fill precision are critical to product quality.

The results demonstrated that **PC1 was the dominant driver of process variability**, with weld-related variables contributing the most. This finding aligns with known challenges in automated sealing operations, where small deviations in weld force or collapse can lead to downstream failures such as leaks or incomplete seals. The negative drift observed in Cluster 1 supports this interpretation, indicating that a subset of parts experienced mechanical instability that was not present in the main operating clusters.

The anomaly detection metrics further supported these observations. The T^2 spike at 03:00 on December 18, 2025, suggests a systematic shift affecting multiple variables simultaneously, which is characteristic of calibration drift, operator adjustments, or equipment fatigue. In contrast, the Q-residual spike at 01:15 on December 18, 2025, indicates a sudden disturbance outside the PCA model structure. This is most likely caused by a transient mechanical event or sensor irregularity. The combination of these two metrics provides a comprehensive view of both predictable and unpredictable process deviations. A potential cause for the shift on the 18th of December, calibration or maintenance could have been I motion.

The clustering behaviour observed in the PCA score plot suggests that the filling line operates in multiple stable modes, potentially corresponding to different filling heads, product types, or machine states. This may be determined by the different botte types or different products. This insight is valuable for continuous improvement, as it highlights the need to understand and standardise these modes to reduce variability. Overall, the PCA model and dashboard visualisations offer a powerful framework for monitoring the filling line. They enable early detection of abnormal behaviour, support root-cause analysis, and provide actionable insights for engineering and quality teams.

PCA-Based Manufacturing Analytics for a Realtime Automated Filling Line.

In conclusion, PCA offers a robust and scalable approach to monitoring complex manufacturing processes. The insights gained from this study demonstrate its potential to improve product quality, reduce waste, and strengthen operational excellence within the diagnostics manufacturing environment. The development of an interactive Streamlit dashboard further enhances the practical value of the PCA model by providing real-time visualisation, anomaly detection, and interpretability for operators and engineers. This tool supports data-driven decision-making and aligns with continuous improvement and Lean Six Sigma principles. Future work may include integrating real-time data streams, expanding the model to additional filling lines, and incorporating predictive maintenance algorithms to further enhance process reliability. Given that the filling line is a newly commissioned instrument, the PCA results provide valuable early indicators of how the machine is settling into routine operation. Continued PCA monitoring is recommended to track early-life machine behaviour, especially score drift on PC1, which signals sealing instability. Hotelling's T^2 should be used to detect systematic shifts such as calibration or set-point changes, while Q-residuals should flag sudden mechanical disturbances. Ongoing observation of PC1 and PC2 separation will help verify long-term stability of both sealing and filling subsystems as the new instrument settles into routine operation.

5.0 REFERENCES

Abbott. (2025) *Diagnostics: Sligo*. Abbott Ireland. Available at: https://www.ie.abbott/about-us/abbott-in-ireland/diagnostics-sligo.html (Accessed: 18 December 2025).

Chiweshe, R., 2023. *Industrial Placement and Case Studies Placement Report*. Unpublished BSc (Honours) report, Atlantic Technological University, Sligo.

Ferrer, A., 2007. Multivariate statistical process control based on principal component analysis (MSPC-PCA): Some reflections and a case study in an autobody assembly process. *Quality Engineering*, 19(4), pp.311–325. Available at: <https://doi.org/10.1080/08982110701621304> [Accessed 18 December 2025].

Greenacre, M., et al. (2022) 'Principal component analysis', *Nature Reviews Methods Primers*, 2(1). doi:10.1038/s43586-022-00184-w.

Guo, H. and Lu, W., 2022. Measuring competitiveness with data-driven principal component analysis: A case study of Chinese international construction companies. *Engineering, Construction and Architectural Management*. Available at: <https://doi.org/10.1108/ecam-04-2020-0262> [Accessed 18 December 2025].

Kherif, F. and Latypova, A. (2020) 'Principal component analysis', in *Machine Learning*, pp. 209–225. doi:10.1016/b978-0-12-815739-8.00012-2.

Li, G. and Qin, Y. (2024) 'An exploration of the application of principal component analysis in Big Data processing', *Applied Mathematics and Nonlinear Sciences*, 9(1). doi:10.2478/amns-2024-0664.

Machine Learning Plus (2023) *Principal component analysis – a deep dive into principal component analysis and its linear algebra foundations* [online]. Machine Learning Plus. Available at: https://www.machinelearningplus.com/linear-algebra/principal-component-analysis/ (Accessed: 4 December 2025).

Smith, L.I. (2002) *A tutorial on principal components analysis* [online]. Department of Computer Science, University of Otago. Available at: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf (Accessed: 2 December 2025).

PCA-Based Manufacturing Analytics for a Realtime Automated Filling Line.

Wu, H. (2013) ‘Principal component analysis for pharmaceutical process and product characterization: case study from immediate and controlled release tablet manufacturing’, *AIChE Annual Meeting Proceedings* [online]. Available at: https://proceedings.aiche.org/conferences/aiche-annual-meeting/2013/proceeding/paper/759c-principal-component-analysis-pharmaceutical-process-and-product-characterizationcase-study-1 (Accessed: 4 December 2025).