International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Batch Process Optimization Using Multi-Variate Data Analysis (MVDA)

Ravi Kiran Koppichetti

IT Data Engineer, API Manufacturing, Novo Nordisk

Abstract
Batch manufacturing in the biopharmaceutical industry involves complex processes where various factors significantly impact product quality, yield, and regulatory compliance. Traditional methods for monitoring and optimizing these processes often struggle to effectively capture intricate interactions between these variables. Multivariate Data Analysis (MVDA) presents a powerful, data-driven solution to evaluate historical and real-time process data, enabling manufacturers to identify critical parameters, detect deviations, and enhance production performance. This paper explores the implementation of MVDA for optimizing batch processes, including data collection, preprocessing, model selection, and real-time monitoring, utilizing data from sensors, Manufacturing Execution Systems (MES), and historical records to create predictive models that improve batch consistency and minimize variability.
Additionally, the paper discusses the regulatory frameworks such as the FDA's Process Analytical Technology (PAT) and Good Manufacturing Practices (GMP) to ensure MVDA applications meet industry standards. It underscores challenges related to data integration, model validation, and system scalability while also highlighting emerging trends in AI-driven analytics and IoT-based process automation. By adopting MVDA, biopharmaceutical manufacturers can achieve greater process efficiency, higher product quality, and lowered operational costs, thus fostering a robust data-driven decision-making culture and promoting continuous process improvement.

Keywords: Multi-Variate Data Analysis (MVDA), Industrial Internet of Things (IIoT), Machine Learning (ML), Industry 4.0, Biopharmaceutical manufacturing, process optimization, Process Analytical Technology (PAT), Critical Process Parameters (CPPs), Critical Quality Attributes, Manufacturing Execution System (MES), Exploration, Prediction, Monitoring, Compliance,

1. Introduction

Biopharmaceutical manufacturing is a complex and tightly regulated process that demands precise control over numerous variables to ensure product quality, efficacy, and regulatory compliance. Unlike traditional pharmaceuticals, which are generally synthesized chemically, biopharmaceuticals are derived from living cells, resulting in a manufacturing process that is inherently more variable and sensitive to fluctuations in raw materials, environmental conditions, and process parameters. To navigate these complexities, Multi-Variate Data Analysis (MVDA) has emerged as a powerful statistical tool essential for monitoring, optimizing, and controlling biopharmaceutical manufacturing processes.
MVDA is a data-driven methodology that allows manufacturers to analyze and interpret extensive, complex datasets involving multiple interrelated variables simultaneously. In contrast to univariate analysis, which investigates one variable at a time, MVDA examines the interdependencies and

correlations among numerous factors, facilitating a more comprehensive understanding of process dynamics. This capability proves particularly beneficial within the frameworks of Process Analytical Technology (PAT) and Quality by Design (QbD), where real-time monitoring and data-informed decision-making are critical for maintaining Critical Process Parameters (CPPs) and Critical Quality Attributes (CQAs) [1, 3, 24].

In the realm of biopharmaceutical manufacturing, MVDA finds application at various stages, including upstream processing (such as fermentation and cell culture), downstream purification (for example, chromatography and filtration), and final product formulation. Process data is collected from a variety of sources, including sensors, Manufacturing Execution Systems (MES), and Process Analytical Technology (PAT) tools, and then analyzed using MVDA techniques. This approach empowers manufacturers to uncover concealed patterns, detect deviations, and optimize processing conditions, ultimately enhancing yield and minimizing variability [10, 25].

A significant application of MVDA is golden batch analysis, where historical batch data is scrutinized to identify optimal conditions that yield the highest-quality products. By juxtaposing real-time batch performance against established benchmarks, manufacturers can proactively recognize process deviations and implement adjustments before issues escalate. Furthermore, MVDA is instrumental in fault detection, root cause analysis, and predictive maintenance, which collectively reduce downtime and ensure consistent product quality [23].

The integration of MVDA in biopharmaceutical manufacturing aligns with the industry's transition toward Industry 4.0 and Pharma 4.0, where data-driven methodologies and digital transformation are pivotal for enhancing operational efficiency. With regulatory agencies advocating for the adoption of advanced analytics for continuous process verification, MVDA is becoming an indispensable tool for ensuring compliance, bolstering process robustness, and accelerating the time-to-market for new biopharmaceutical products.

2. MVDA in Batch Process of Pharmaceutical Manufacturing

In the realm of pharmaceutical manufacturing, the use of Multi-Variate Data Analysis (MVDA) shows significant variation between batch and continuous processes, stemming from inherent differences in production dynamics. Batch manufacturing occurs in distinct cycles, where raw materials are processed in predetermined quantities, with each batch treated as an independent unit. This method introduces variability among batches, requiring the application of MVDA techniques such as Principal Component Analysis (PCA), Partial Least Squares (PLS), and Batch Evolution Modeling, to enable performance comparison, identify optimal process conditions, and detect deviations from historical patterns. Since data in batch processes is segmented and limited by time, MVDA is utilized for retrospective analysis, process optimization, and identifying causes of batch-to-batch variability [1].

Conversely, continuous manufacturing functions in an uninterrupted, steady-state mode where materials flow smoothly through various stages, making real-time monitoring essential. Unlike batch processes, which analyze variability post-production, continuous manufacturing depends on real-time MVDA models including Real-time Partial Least Squares (RT-PLS), Artificial Neural Networks (ANNs), and Dynamic Time Warping (DTW) to predict product quality, identify anomalies, and ensure process stability. Continuous processes require a constant flow of data from sensors, incorporating Process Analytical Technology (PAT) to enable real-time adjustments and improvements. This scenario demands more complex MVDA models capable of managing dynamic process variations, thus ensuring regulatory compliance and reducing deviations. From a regulatory standpoint, batch processes follow well-defined validation protocols, making compliance relatively straightforward [2, 4].

In contrast, continuous manufacturing demands more advanced real-time control strategies to meet regulatory standards, requiring deeper integration with PAT tools and automated decision-making processes. While batch MVDA primarily addresses process inconsistencies and optimization between batches, continuous MVDA mainly focuses on maintaining stable operations, predictive control, and immediate corrective actions to ensure quality and efficiency. As the pharmaceutical industry transitions toward Industry 4.0 and digital transformation, the significance of continuous MVDA increases, helping manufacturers achieve higher levels of efficiency, compliance, and cost-effectiveness compared to traditional batch processing. Implementing MVDA in batch pharmaceutical manufacturing entails a systematic approach that incorporates data collection, preprocessing, modeling, and analysis to enhance process efficiency, improve product quality, and ensure adherence to regulatory standards [3].

3. Data Collection for MVDA

Data for MVDA is collected from various sources, including:

3.1 Process Sensors:

Process sensors are vital in pharmaceutical manufacturing as they continuously monitor key process parameters to ensure product quality, regulatory compliance, and operational efficiency. By providing real-time data on critical process parameters (CPPs) that affect critical quality attributes (CQAs), these sensors help manufacturers adhere to standards set by the FDA, EMA, and cGMP. In addition to maintaining quality, manufacturers leverage process sensors to minimize waste, maximize yield, and reduce costs through real-time monitoring and control systems, which enable quick adjustments for stability. Overall, process sensors are crucial for enhancing automation and supporting data-driven decisions in the industry [5, 6, 7, 8, 9].

Types of Process Sensors in Pharmaceutical Manufacturing

3.1.1 Physical Parameter Sensors: These devices gauge essential physical attributes impacting manufacturing processes.

- Temperature Sensors: Assess the temperature during processes to guarantee ideal reaction settings (e.g., thermocouples and RTDs).
- Pressure Sensors: Keep track of pressure in bioreactors, filtration systems, and chromatography activities.
- Flow Sensors: Monitor liquid and gas flow rates within pipelines to ensure precise dosing and mixing.
- Level Sensors: Identify liquid levels in tanks and vessels to avert overflow or dry states.
- pH Sensors: Maintain pH levels in bioreactors and chemical reactions to support optimal biological performance.

3.1.2 Chemical Composition Sensors: These sensors evaluate the chemical characteristics of process streams.

- Conductivity Sensors: Monitor ion concentration during buffer preparation and purification phases.
- Dissolved Oxygen (DO) Sensors: Measure oxygen concentrations in fermentation and bioreactors to support cellular growth.
- Carbon Dioxide (CO_2) Sensors: Observe CO_2 concentrations in cell culture and fermentation processes.
- Total Organic Carbon (TOC) Sensors: Ensure water purity by identifying organic contaminants.

3.1.3 Spectroscopic and Analytical Sensors (Process Analytical Technology Tools): These sensors facilitate real-time, non-destructive examination of pharmaceutical processes.

- Near-Infrared (NIR) Sensors: Assess moisture content, active pharmaceutical ingredient (API) concentration, and blend uniformity.
- Raman Spectroscopy Sensors: Monitor the chemical composition and identify impurities present in drug formulations.
- Ultraviolet-Visible (UV-Vis) Spectroscopy Sensors: Evaluate protein concentration during the purification of biopharmaceuticals.
- Fourier Transform Infrared (FTIR) Sensors: Determine chemical bonds in complex molecules for the analysis of raw materials.

3.1.4 Biological and Microbiological Sensors: These instruments are primarily employed in biopharmaceutical production to sustain aseptic conditions.

- Viable Particle Counters: Utilized to detect microbial contamination in cleanroom environments.
- Endotoxin Sensors: These sensors quantify bacterial endotoxins in injectable products and Water for Injection (WFI).
- Cell Density Sensors: Designed to monitor cell growth in bioreactors, thereby optimizing the timing of harvest.

3.1.5 Smart and IoT-Enabled Sensors: Modern pharmaceutical facilities are adopting intelligent sensors to enable real-time analytics and automate operations.

- Wireless Temperature and Humidity Sensors: These sensors monitor the storage and transport conditions of sensitive pharmaceuticals.
- Automated Vision Sensors: These sensors detect defects in tablets, capsules, and vials, ensuring quality assurance.
- IoT-Enabled Process Sensors: Designed to connect with cloud systems, these sensors support predictive maintenance and process optimization.

3.2 Manufacturing Execution Systems (MES):

A Manufacturing Execution System (MES) is a crucial software tool in the pharmaceutical sector that optimizes manufacturing operations in real time while ensuring compliance with regulatory standards like FDA Part 11 and EU GMP Annex 11. By integrating with systems like Enterprise Resource Planning (ERP) and Supervisory Control and Data Acquisition (SCADA), MES enhances efficiency, reduces cycle times, and minimizes waste. It provides real-time visibility into production data, improving decision-making and quality control, thus decreasing batch failures and facilitating cost reduction through better resource utilization. Ultimately, MES supports the digital transformation of pharmaceutical manufacturing, aligning operations with Industry 4.0 and Pharma 4.0 initiatives [10, 11, 12, 13].

Key Features of MES in Production Planning

3.2.1 Real-Time Monitoring: The Manufacturing Execution System (MES) offers immediate visibility into production processes, allowing stakeholders to oversee operations, identify bottlenecks, and make well-informed decisions swiftly.

3.2.2 Regulatory Compliance: MES systems in the pharmaceutical industry are specifically designed to uphold stringent regulatory compliance and ensure product integrity. These systems meticulously track and document the entire manufacturing process, including batch tracking, genealogy, and adherence to good manufacturing practices (GMP).

3.2.3 Data Integration: MES functions as a centralized system, effectively interfacing with other manufacturing systems and departments, such as operations, quality assurance, maintenance, and inventory control.

3.2.4 Electronic Batch Records (EBR): The MES facilitates the development and management of electronic batch records, thereby minimizing paper-based processes and their associated errors.

3.2.5 Quality Control: These systems prioritize real-time quality control, recipe management, and batch management to guarantee product consistency and safety.

3.2.6 Resource Management: The MES systematically tracks all production-related resources, including materials, equipment, and personnel, within a comprehensive electronic batch recording system.

3.2.7 Process Optimization: By delivering accurate and current information, the MES empowers manufacturers to enhance scheduling, resource utilization, and manufacturing flexibility.

3.3 Historical Batch Data:

Historical batch data comprises the documented records of all batch production activities, including process parameters, quality attributes, and any deviations from previous batches. This information is essential for optimizing processes, ensuring regulatory compliance, maintaining quality control, and promoting continuous improvement in pharmaceutical manufacturing. It serves as a cornerstone for data-driven decision-making and is critical for batch release, troubleshooting, and predictive analysis analytics [14, 15].

Components of Historical Batch Data

3.3.1 Process Parameters: Temperature, pressure, pH, flow rate, mixing speed, and other parameters are gathered from Process Control Systems (PCS), Distributed Control Systems (DCS), and Manufacturing Execution Systems (MES).

3.3.2 Quality Control Data: Critical Quality Attributes (CQAs) include assays, potency, impurity profiles, and dissolution rate. In-process controls (IPCs) involve testing intermediate products during manufacturing, while Finished Product Testing ensures compliance with regulatory standards.

3.3.3 Batch Production Records (BPRs) / Electronic Batch Records (EBRs): These records provide a detailed overview of the manufacturing process, highlighting operator interventions, the equipment utilized, and specific raw material batches involved.

3.3.4 Raw Material & Equipment Data: Details of suppliers, batch information of excipients, active pharmaceutical ingredients (APIs), solvents, and records of equipment used, including calibration and preventive maintenance.

3.3.5 Deviations: All process deviations, out-of-specification (OOS) results, and quality issues that arise during batch manufacturing are recorded for root cause analysis (RCA) as well as for corrective and preventive actions.

3.3.6 Yield & Efficiency Metrics: Examine batch yield by assessing the expected product output against the actual output, including cycle time assessments for each manufacturing process.

4. Data Preprocessing for MVDA

Data preprocessing is crucial in Multi-Variate Data Analysis (MVDA) as it guarantees the accuracy, consistency, and reliability of analytical results. In the biopharmaceutical manufacturing sector, raw process data is often collected from various sources, including sensors, Process Analytical Technology (PAT) tools, Manufacturing Execution Systems (MES), and Laboratory Information Management Systems (LIMS). This data frequently suffers from noise, incompleteness, or inconsistencies, making a methodical preprocessing workflow necessary before applying MVDA techniques. By establishing a structured data preprocessing pipeline, MVDA models in biopharmaceutical manufacturing can provide precise insights for process optimization, fault detection, and predictive quality control. This leads to improved decision-making, enhanced batch consistency, and more robust regulatory compliance [15, 16, 17, 5, 4, 14, 18].

4.1 Data Acquisition and Integration

4.1.1 Sources of Data: Data is obtained from real-time sensors, historical batch records, quality control laboratories, and environmental monitoring systems.

4.1.2 Data Standardization: Diverse sources may utilize varying formats and units, so data standardization is imperative to uphold consistency. This procedure encompasses unit conversions, including the transformation of temperature from Fahrenheit to Celsius, alongside the synchronization of time for coherent analysis.

4.1.3 Data Fusion: Combining various data types, such as structured MES data and unstructured lab reports, to form a complete dataset for MVDA.

4.2 Handling Missing Data

4.2.1 Deletion Methodology: In cases where a limited number of values are absent, the corresponding records can be eliminated.

4.2.2 Imputation Methods: In cases of significant missing data, various techniques such as mean or mode imputation, regression-based imputation, and interpolation methods (for example, forward-fill or spline interpolation) are utilized to estimate the missing values.

4.2.3 Advanced Techniques: Machine learning algorithms such as K-Nearest Neighbors (KNN) and Expectation-Maximization (EM) possess the capability to infer missing values by analyzing correlations within the dataset.

4.3 Data Cleaning and Noise Reduction

4.3.1 Outlier Detection: Involves identifying and managing outliers through methods like Z-score analysis, Principal Component Analysis (PCA), or robust regression models to avoid skewed MVDA results.

4.3.2 Smoothing Techniques: Moving averages, Savitzky-Golay filters, and wavelet transformations can eliminate noise from continuous process signals (e.g., temperature, pH, or pressure trends).

4.3.3 Anomaly Detection: Techniques such as Hotelling's T^2 statistic, Mahalanobis distance, or unsupervised clustering can help identify unusual variations that may suggest sensor malfunctions or process deviations.

4.4 Feature Engineering and Transformation

4.4.1 Normalization & Scaling: Data must be normalized (e.g., min-max scaling, and Z-score standardization) to ensure a fair comparison across variables with varying magnitudes.

4.4.2 Feature Selection: To enhance model interpretability, irrelevant or redundant variables can be eliminated through correlation analysis, Recursive Feature Elimination (RFE), or LASSO regression.

4.4.3 Dimensionality Reduction: PCA or Partial Least Squares (PLS) can be used to reduce high-dimensional data while maintaining essential process variability.

4.5 Data Structuring for MVDA Models

4.5.1 Time-Series Data Alignment: Ensuring consistent timestamps across various process variables to prevent misalignment issues in batch comparisons.

4.5.2 Batch Data Structuring: Data from multiple batches should be formatted into a batch-wise matrix, with rows representing batches and columns representing process variables.

4.5.3 Categorical Data Encoding: Variables such as batch IDs, material sources, or equipment types may require transformation using techniques like one-hot encoding, ordinal encoding, or embedding before analysis.

4.6 Ensuring Data Integrity and Compliance

4.6.1 Audit Trails: All preprocessing steps must be recorded to guarantee data traceability and compliance with FDA 21 CFR Part 11 and Good Automated Manufacturing Practice (GAMP).

4.6.2 Validation Assessments: It is imperative to conduct statistical tests, such as the Shapiro-Wilk test for normality and Levene's test for homogeneity, to verify data quality prior to the application of Multivariate Data Analysis (MVDA).

5. Models for Multivariate Data Analysis

Multi-variate Data Analysis (MVDA) employs statistical and machine learning techniques to examine intricate datasets that involve several variables. These techniques reveal concealed patterns, correlations, and crucial elements that influence biopharmaceutical manufacturing processes. The key MVDA models applied in this field include exploratory, predictive, and monitoring models. In biopharmaceutical manufacturing, MVDA utilizes these models to enhance production, elevate quality, and maintain compliance with regulations. By employing models like PCA, PLS, Neural Networks, and Statistical Process Control, manufacturers can actively manage variability, identify anomalies, and boost overall process efficiency. The appropriate MVDA model selection is guided by the specific process objectives, data intricacy, and regulatory standards [6, 19, 20, 13, 21, 22, 23].

5.1 Exploratory Models

5.1.1 Principal Component Analysis (PCA): This method aims to reduce dimensionality while maintaining the crucial variance in the dataset. In the biopharmaceutical industry, it helps identify key factors influencing batch variability, recognize trends, and classify different processing conditions. A concrete application of this approach is in comparing batches during purification or fermentation processes.

5.1.2 Partial Least Squares (PLS) Regression: This process aims to establish connections between input (process) and output (quality) variables. This is essential in the biopharmaceutical field for developing soft sensors used in real-time process monitoring. For example, it can predict protein yield based on critical process parameters such as pH, temperature, and dissolved oxygen levels.

5.1.3 Cluster Analysis: This process categorizes analogous observations based on data patterns. It is especially relevant in the biopharmaceutical industry for segmenting diverse production conditions and identifying prevalent failure modes. For example, it proficiently identifies abnormal production batches that diverge from established historical norms.

5.2 Predictive Models

5.2.1 Multiple linear Regression: This model establishes relationships between independent variables and a dependent variable. This model is often used in the biopharmaceutical sector to predict drug potency or yield based on process variables. An example illustrates its use in modeling the impact of raw material variability on the quality of the final product.

5.2.2 Artificial Neural Networks (ANNs): This approach aims to capture complex, non-linear relationships between input and output variables. It is applicable in the biopharma sector, including advanced process control, soft sensors, and predictive maintenance. For example, it can forecast deviations in bioreactor performance by analyzing historical process data.

5.2.3 Support Vector Machines (SVMs): This classification is used to identify optimal decision boundaries for data points, particularly in the biopharma sector. It detects process deviations and categorizes product batches as acceptable or non-acceptable. An example of its application is distinguishing between high-quality and defective drugs.

5.3 Monitoring & Fault Detection Models

5.3.1 Hotelling's T^2 and Squared Prediction Error (SPE) Statistics: The goal of monitoring process deviations is to identify anomalies, especially in the biopharma sector, where it acts as an early warning system for critical deviations in process parameters. For example, it can detect unexpected changes in chromatography columns.

5.3.2 Multivariate Statistical Process Control (MSPC): This method improves upon traditional Statistical Process Control (SPC) techniques by analyzing multivariate datasets, especially in the biopharmaceutical sector. It is crucial for monitoring batch consistency and ensuring compliance with regulatory standards, including the identification of process drifts in sterile drug production.

5.3.3 Dynamic Time Warping (DTW) & Batch Evolution Modeling: This analysis compares time series data from various process runs. It is particularly useful in biopharma for assessing batch similarity and identifying deviations from the established "golden batch." A practical example is aligning real-time fermentation data with historical high-yield records.

6. Real-time Monitoring of MVDA

Real-time monitoring in batch production refers to the continuous tracking and analysis of process parameters, equipment performance, and product quality during the manufacturing cycle. It enables immediate detection of deviations, process optimization, and proactive decision-making to ensure batch consistency, regulatory compliance, and operational efficiency. The data points from the PAT Sensors, MES, SCADA, and LIMS are continuously collected and streamed into historical databases or cloud platforms for analysis [18, 7, 8, 20].

6.1 Batch Evolution Tracking: This process involves monitoring and analyzing the progress of batch processes over time to ensure they follow expected trajectories, detect deviations, and identify faults. This is valuable because consistency and quality in batches are essential. Real-time MVDA models evaluate batch performance by comparing current batch profiles with historical golden batch trajectories. Time-series analysis and multivariate control charts facilitate early detection of process drifts, while outlier detection models identify deviations from expected behavior, enabling timely corrective actions to prevent failures.

6.2 **Anomaly Detection & Alerts**: Anomaly detection in batch processing using Multivariate Data Analysis (MVDA) involves identifying deviations from normal batch behavior by analyzing multiple process parameters simultaneously. By leveraging historical batch data, statistical models, and machine learning algorithms, manufacturers can detect early warning signs of process inefficiencies, quality issues, and equipment failures. Principal Component Analysis (PCA) and Partial Least Squares (PLS) detect process deviations beyond normal operating ranges, while Machine Learning-based anomaly detection identifies subtle deviations that could foreshadow future process failures. Automated alarms and notifications ensure operators are alerted to critical deviations, facilitating timely intervention.

7. Regulatory Compliance of MVDA

Multivariate Data Analysis (MVDA) constitutes an instrumental methodology within the realm of pharmaceutical manufacturing, serving purposes such as process monitoring, optimization, and quality control. Nevertheless, to secure regulatory acceptance and enhance effectiveness, MVDA models must adhere to rigorous industry guidelines. Such compliance ascertains that data-driven decisions are consistent with Good Manufacturing Practices (GMP), pertinent regulatory requirements, and the risk management frameworks established by regulatory authorities including the FDA (U.S. Food and Drug Administration), EMA (European Medicines Agency), and ICH (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use) [5, 1].

8. Conclusion

Multivariate Data Analysis (MVDA) has emerged as a transformative instrument in the optimization of batch processes within biopharmaceutical manufacturing. It empowers manufacturers to enhance process efficiency, product quality, and adherence to regulatory requirements. By utilizing historical batch data, monitoring processes in real-time, and employing predictive modeling, MVDA delivers profound insights into Critical Process Parameters (CPPs) and Critical Quality Attributes (CQAs), thereby ensuring consistent product yield and minimizing process variability.

Implementing MVDA in batch manufacturing facilitates anomaly detection, root cause analysis, and continual process improvement, establishing it as a pivotal enabler of Process Analytical Technology (PAT) and Quality by Design (QbD) frameworks. Integrating real-time sensor data, Manufacturing Execution Systems (MES), and advanced analytics fosters proactive decision-making, reduces instances of batch failure, and optimizes resource utilization.

Notwithstanding its benefits, integrating MVDA presents challenges concerning data assimilation, model validation, and compliance with regulations. Ensuring data integrity, model transparency, and alignment with industry guidelines—including the FDA’s 21 CFR Part 11, ICH Q8-Q12, and Good Manufacturing Practice (GMP) regulations—is imperative for the successful implementation of MVDA.

Looking forward, advancements in artificial intelligence (AI), machine learning (ML), and Internet of Things (IoT)-enabled biopharmaceutical manufacturing are anticipated to further augment the capabilities of MVDA, facilitating autonomous process control and predictive optimization. As the industry advances towards smart manufacturing and digital transformation, MVDA will remain an essential strategy for maximizing operational efficiency, ensuring regulatory compliance, and fostering ongoing innovation in the optimization of batch processes.

9. References

1. Beckett, L. Eriksson, E. Johansson, and C. Wikström, “Multivariate data analysis (MVDA),” in *Pharmaceutical Quality by Design: A Practical Approach*, 2018, pp. 201–225.
2. T. Williams et al., “Machine learning and metabolic modelling assisted implementation of a novel process analytical technology in cell and gene therapy manufacturing,” *Scientific Reports*, vol. 13, no. 1, p. 834, 2023.
3. Corredor, D. Bu, and G. McGeorge, “Applications of MVDA and PAT for Drug Product Development and Manufacturing,” in *Multivariate Analysis in the Pharmaceutical Industry*, Academic Press, 2018, pp. 211–234.
4. K. Kylefors, “Evaluation of leachate composition by multivariate data analysis (MVDA),” *Journal of Environmental Management*, vol. 68, no. 4, pp. 367–376, 2003.
5. S. Goldrick et al., “Multivariate data analysis methodology to solve data challenges related to scale-up model validation and missing data on a micro-bioreactor system,” *Biotechnology Journal*, vol. 15, no. 3, p. 1800684, 2020.
6. J. Glassey, “Multivariate data analysis for advancing the interpretation of bioprocess measurement and monitoring data,” in *Measurement, Monitoring, Modelling and Control of Bioprocesses*, 2013, pp. 167–191.
7. M. Amazouz and R. Pantea, “Use of Multivariate Data Analysis for Lumber Drying Process Monitoring and Fault Detection,” in *Proceedings of DMIN*, June 2006, pp. 329–332.
8. S. Rathore, S. Mittal, M. Pathak, and A. Arora, “Guidance for performing multivariate data analysis of bioprocessing data: pitfalls and recommendations,” *Biotechnology Progress*, vol. 30, no. 4, pp. 967–973, 2014.
9. S. Barimani and P. Kleinebudde, “Evaluation of in–line Raman data for end-point determination of a coating process: Comparison of Science–Based Calibration, PLS-regression and univariate data analysis,” *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 119, pp. 28–35, 2017.
10. H. Alosert et al., “Data integrity within the biopharmaceutical sector in the era of Industry 4.0,” *Biotechnology Journal*, vol. 17, no. 6, p. 2100609, 2022.
11. V. Steinwandter, D. Borchert, and C. Herwig, “Data science tools and applications on the way to Pharma 4.0,” *Drug Discovery Today*, vol. 24, no. 9, pp. 1795–1805, 2019.
12. K. Isoko, J. L. Cordiner, Z. Kis, and P. Z. Moghadam, “Bioprocessing 4.0: a pragmatic review and future perspectives,” *Digital Discovery*, vol. 3, no. 9, pp. 1662–1681, 2024.
13. A. Puranik, M. Saldanha, N. Chirmule, P. Dandekar, and R. Jain, “Advanced strategies in glycosylation prediction and control during biopharmaceutical development: Avenues toward industry 4.0,” *Biotechnology Progress*, vol. 38, no. 5, p. e3283, 2022.
14. O. Kirdar, K. D. Green, and A. S. Rathore, “Application of multivariate data analysis for identification and successful resolution of a root cause for a bioprocessing application,” *Biotechnology Progress*, vol. 24, no. 3, pp. 720–726, 2008.
15. J. Emerson, B. Kara, and J. Glassey, “Multivariate data analysis in cell gene therapy manufacturing,” *Biotechnology Advances*, vol. 45, p. 107637, 2020.

16. R. Johnson et al., “Applications of multivariate data analysis in biotech processing,” *Biopharm International*, vol. 20, no. 10, p. 130, 2007.
17. L. Eriksson, T. Byrne, E. Johansson, J. Trygg, and C. Vikström, *Multi-and Megavariate Data Analysis: Basic Principles and Applications*, vol. 1, Umetrics Academy, 2013.
18. Pretzner et al., “Multivariate Monitoring Workflow for Formulation, Fill and Finish Processes,” *Bioengineering*, vol. 7, no. 2, p. 50, 2020.
19. Buvé et al., “Application of multivariate data analysis for food quality investigations: An example-based review,” *Food Research International*, vol. 151, p. 110878, 2022.
20. S. M. Mercier, B. Diepenbroek, M. C. Dalm, R. H. Wijffels, and M. Streefland, “Multivariate data analysis as a PAT tool for early bioprocess development data,” *Journal of Biotechnology*, vol. 167, no. 3, pp. 262–270, 2013.
21. P. Pořízka et al., “On the utilization of principal component analysis in laser-induced breakdown spectroscopy data analysis, a review,” *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 148, pp. 65–82, 2018.
22. K. Ritchie, “Application of multivariate data analysis in biopharmaceutical production,” Ph.D. dissertation, Newcastle University, 2016.
23. P. P. Lapointe-Garant, R. Kamyar, R. Paravitorghabeh, and Z. Wang, “Multivariate Analysis for Process Understanding, Continuous Process Verification, and Condition Monitoring of Lyophilization Processes,” in *Principles and Practices of Lyophilization in Product Development and Manufacturing*, Springer International Publishing, 2023, pp. 577–594.
24. M. Schad, S. Gautam, T. A. Grein, and F. Käß, “Process Analytical Technologies (PAT) and Quality by Design (QbD) for Bioprocessing of Virus-Based Therapeutics,” in *Bioprocess and Analytics Development for Virus-Based Advanced Therapeutics and Medicinal Products (ATMPs)*, Springer International Publishing, 2023, pp. 295–328.
25. K. Isoko, J. L. Cordiner, Z. Kis, and P. Z. Moghadam, “Bioprocessing 4.0: a pragmatic review and future perspectives,” *Digital Discovery*, vol. 3, no. 9, pp. 1662–1681, 2024.