Multivariate Approach to small-scale Model Qualification (SSMQ) in Upstream (USP) Adeno-Associated Virus (AAV) Production

Apezteguia A., Iglesias M., and Keune J.

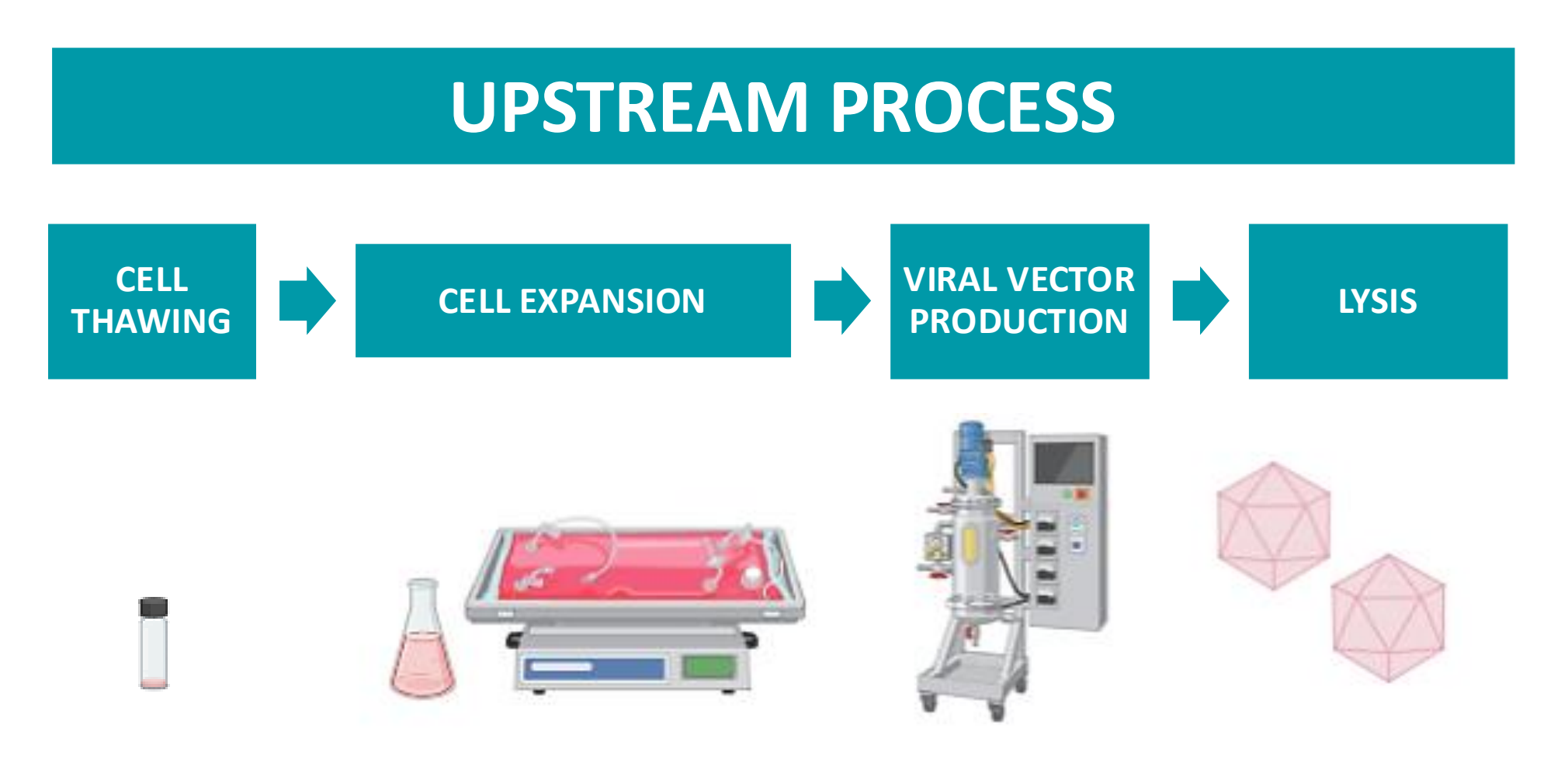
Abstract

AAV gene therapy manufacturing requires scalable processes and reliable small-scale models (SSMs) to support process characterization and regulatory expectations. Ensuring that SSMs accurately represent large-scale operations is essential for defining robust process acceptance ranges. However, the high dimensionality and multivariate nature of upstream AAV processes limit the use of traditional univariate similarity assessments.

This work presents a multivariate framework for qualifying small-scale cell-expansion models using historical large-scale data. By applying dimensionality-reduction and multivariate modeling techniques, we evaluate the alignment between scales and demonstrate the suitability of SSMs for upstream characterization.

Introduction

The field of cell and gene therapy (CGT) continues to expand, driven by the therapeutic potential of adeno-associated virus (AAV)–based treatments for a wide range of genetic disorders. As AAV programs progress from early development to commercialization, manufacturing must operate reliably across multiple scales. To ensure consistent product quality during scale transitions, Process Characterization (PC) activities are used to identify critical process variables and define Proven Acceptable Ranges (PAR). Small-scale models (SSMs) play a central role in PC by enabling controlled experimentation that is impractical at large scale. A key requirement for using SSMs in PC is demonstrating that they accurately represent the corresponding large-scale process—a framework referred to as Small-Scale Model Qualification (SSMQ).



UPSTREAM PROCESS

CELL THAWING → **CELL EXPANSION** → **VIRAL VECTOR PRODUCTION** → **LYSIS**

However, upstream AAV manufacturing is inherently high-dimensional, with each batch generating hundreds of measurements related to cell growth, metabolism, nutrient consumption, gas transfer, and environmental conditions. These variables exhibit strong correlations and dynamic interactions, reflecting the biological complexity of mammalian cell culture. Traditional univariate tools, such as Two One-Sided Tests (TOST), evaluate parameters independently and therefore fail to capture the multivariate structure of the process. This motivates the adoption of multivariate analytical approaches that can evaluate correlated variables simultaneously and provide a more comprehensive measure of similarity between scales(4). The objective of this study is to apply such a multivariate framework to qualify small-scale cell-expansion models against large-scale manufacturing data.

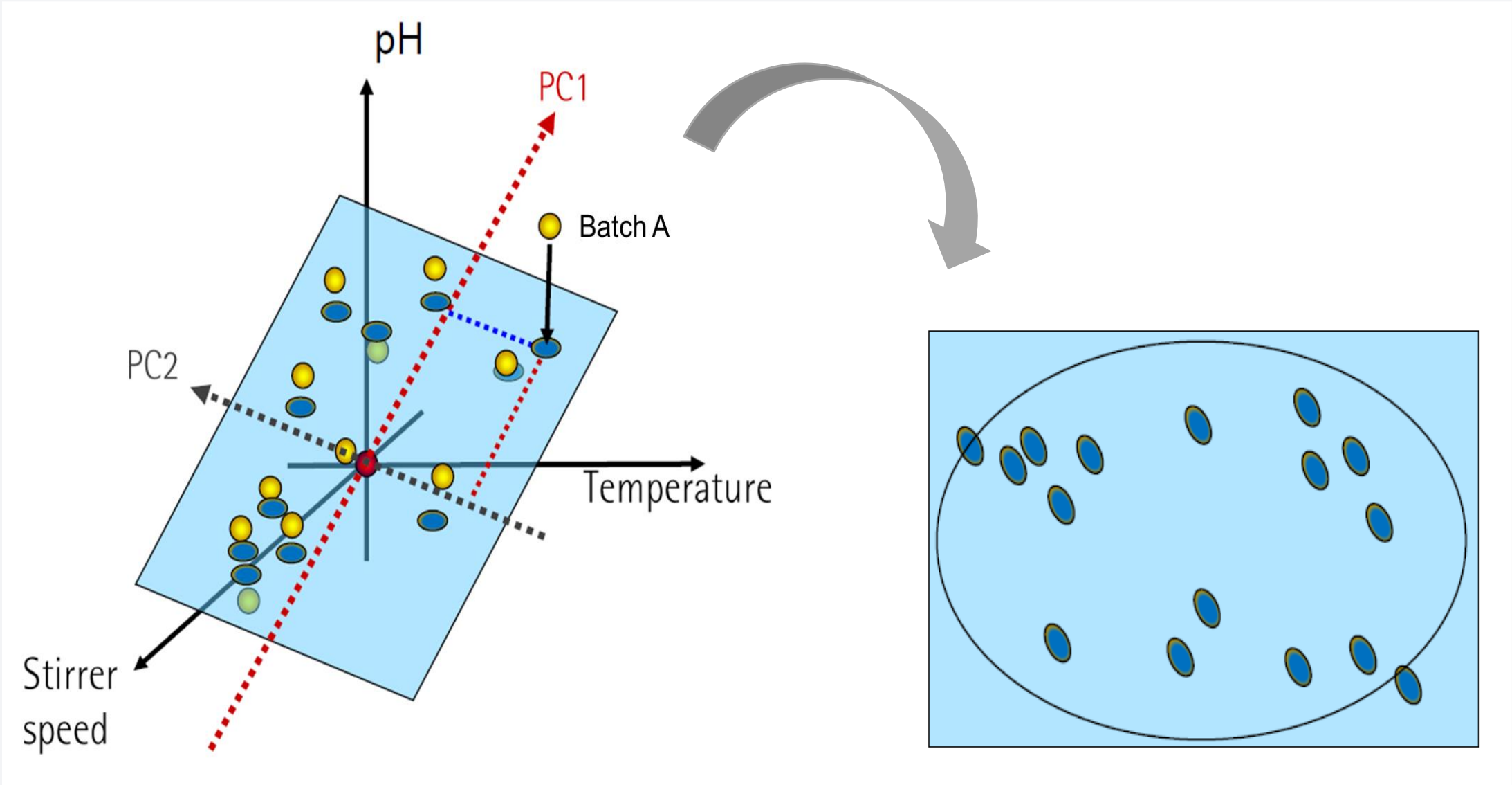
Methods

Multivariate data analysis was conducted using Python(1) and SIMCA(2) software.

Data collection:

- Two datasets constructed for the cell-expansion stage in the production bioreactor:
  - 22 large-scale historical batches
  - 6 small-scale batches (3 executed to date)
- Included in-line and off-line process parameters
- Off-line measurements in the transfection bioreactor taken at least daily
- Timestamp metadata incorporated to capture temporal evolution of variables

(3)



Dimension-reduction techniques such as Principal Component Analysis (PCA) and Partial Least Squares (PLS) were employed to effectively model the high-dimensional process data.

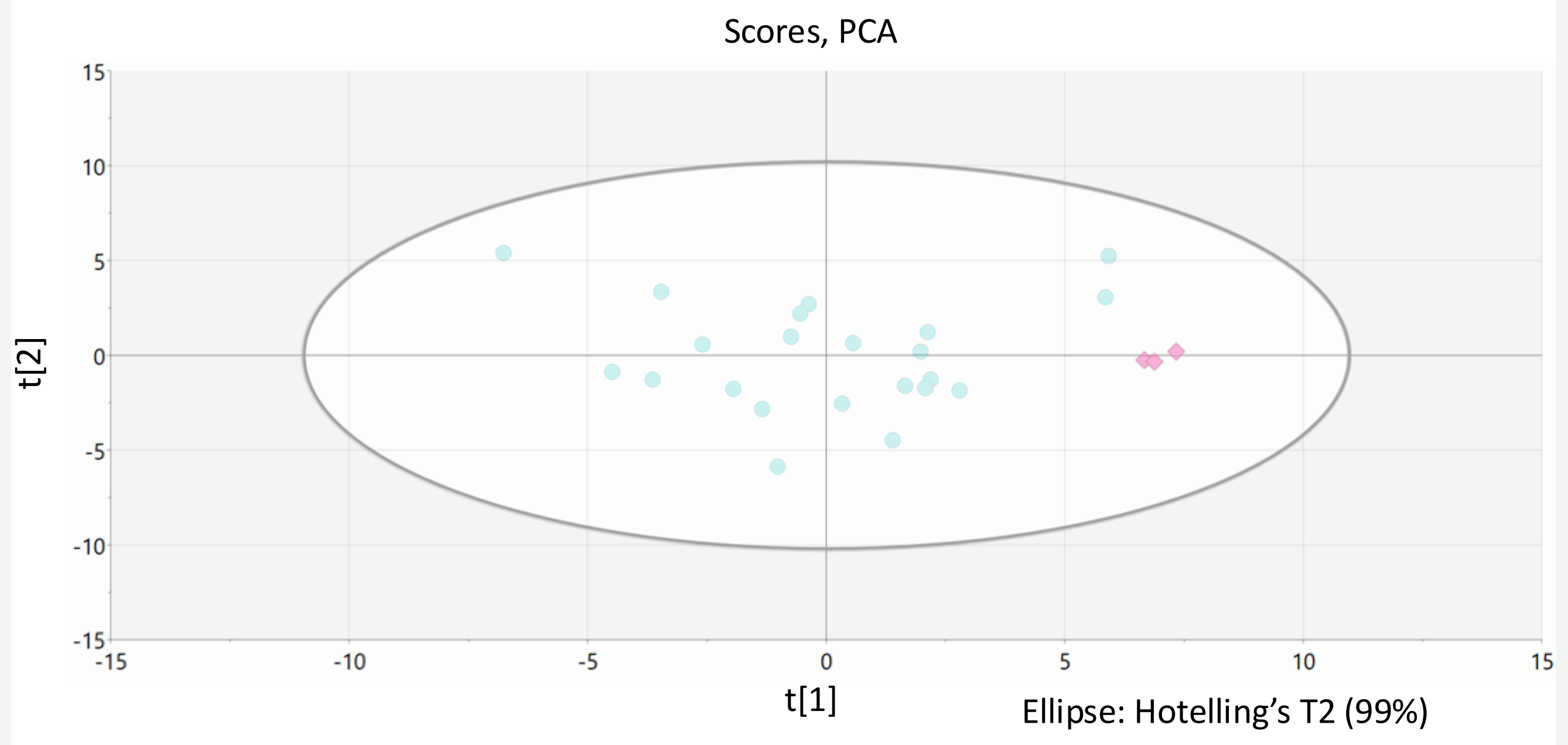
By collecting data from small-scale cell-culture batches and projecting this data into the established multivariate model space, the study assesses how well these smaller batches mimic the larger-scale processes based on their proximity to the large-scale batches within the model space (4).

Results

Batch-level Modelling

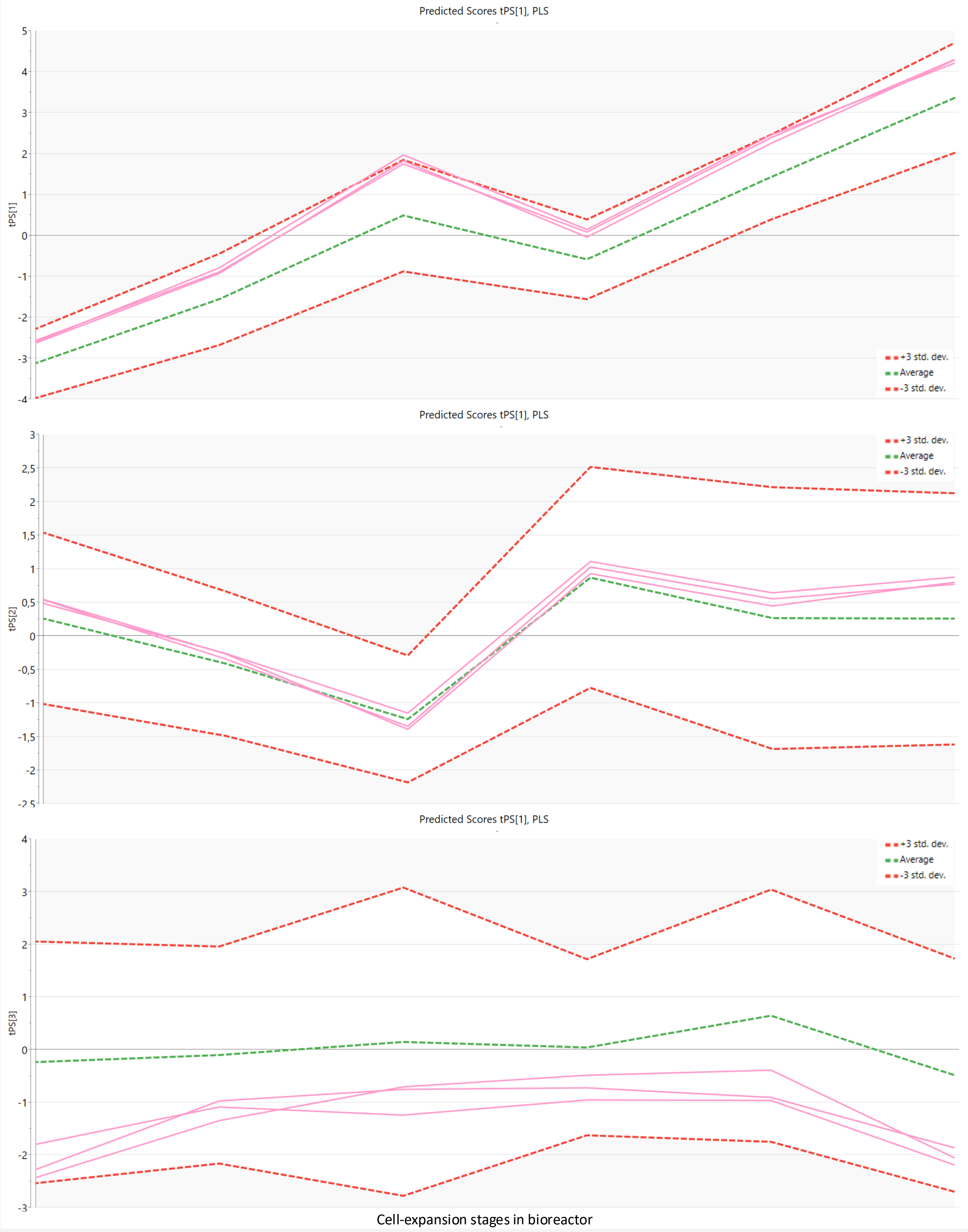
A PCA score plot was generated to evaluate process comparability between the small and the large scales.

The PCA model captures over 80% of the total variance in the dataset (R^2X(cum) = 0.803). By projecting the original p-dimensional variables into a reduced k-dimensional latent space, PCA summarizes the dominant sources of variation using the first k principal components, where $k \ll p$ (4).



The 99% Hotelling's T^2 confidence ellipse was used to identify potential excursions from the multivariate model. All small-scale batches (pink markers) project within this tolerance region, indicating strong comparability between small- and large-scale operations.

Batch Evolution Modelling



A PLS model was used to determine process limits in the time-transformed domain. Statistical control limits were calculated using ±3 standard deviations from the transformed large-scale reference data (4).

This approach provides a high-level multivariate assessment of process behavior across scales throughout the batch trajectory. The resulting model demonstrated strong explanatory and predictive performance, with R^2X(cum) = 0.889, R^2Y(cum) = 0.926, and Q^2(cum) = 0.92.

Principal component scores calculated from the small-scale batch data (pink lines) were overlaid onto the multivariate batch trajectory derived from large-scale operations to support visual identification of atypical behavior and to assess comparability across scales. While overall trajectory alignment was observed, one small-scale batch exhibited a localized deviation at a single time point along PC1. Contribution plot analysis was therefore applied to identify the variables driving differences between the small- and large-scale projections within the latent space.

The variable identified in the contribution plot is not classified as critical for process control or product quality. It is measured to enhance process understanding, and the observed difference is regarded as a secondary effect rather than an underlying cause of the deviation observed in the multivariate projection.

Conclusion

Overall, contingent upon confirmation of the observed behavior in the remaining small-scale batches, these findings illustrate how multivariate approaches can be applied to support SSMQ considerations. Furthermore, they contribute to an improved understanding of scalability in cell-culture processes and describe a framework that may help guide future process characterization activities. Given that transfection is a key operation in the manufacturing process and that multivariate correlations are present, applying the multivariate approach to the acceptance criteria for the transfection unit operation parameters and attributes in the future could provide valuable insights. However, this would require more product-specific data, as the quality and process performance may vary across different genes of interest (GOI) or capsid serotypes. The three remaining small-scale batches still to be manufactured will be evaluated using the same multivariate approach applied to the batches presented here, ensuring a consistent assessment of scale comparability

References

1) Python Software Foundation. (n.d.). Python. https://www.python.org
2) Umetrics. (n.d.). SIMCA. https://www.umetrics.com/simca
3) Sartorius Stedim Biotech. (n.d.). Sartorius Stedim Biotech. https://www.sartorius.com
4) BioPhorum. (2021). Justification of Small-Scale Models: An Industry Perspective. https://www.biophorum.com/download/justification-of-small-scale-models-an-industry-perspective/