# SCALE-UP/DOWN OF A MONOCLONAL ANTIBODY MANUFACTURING BIOPROCESS USING DATA ANALYTICS

Pierantonio Facco[*,a], Simeone Zomer[b], Ruth C. Rowland-Jones[b], Douglas Marsh[b], Paloma Diaz-Fernandez[b], Gary Finka[b], Fabrizio Bezzo[a], and Massimiliano Barolo[a]
[a] CAPE-Lab – Computer-Aided Process Engineering Laboratory
Department of Industrial Engineering, University of Padova, Padova (Italy)
[b] GlaxoSmithKline Research & Development
Ware (U.K.)

*Abstract*

The scale up/down of biopharmaceutical processes is still a very challenging task. Different cell lines, although clones of the same cell, display different performance in terms of drug productivity. In drug development, the effect of process parameters on cell line performance is often not completely understood and there is a lack of sound science-based methodologies to address this issue. However, the Industry 4.0 revolution is changing the biopharmaceutical industry standards. Extensive digitalization is determining that, even during the process development and scale-up, a significant amount of data can be collected and exploited.

In this study we consider a monoclonal antibody manufacturing bioprocess and we focus on two main objectives: the possibility of identifying the most promising cell lines in terms of drug productivity and performance stability from the early development stages, and the prediction of cell lines performance across scales during scale-up/down. This is possible by taking advantage of the information available in the data using multivariate, multiway and multiblock statistical techniques.

*Keywords*

Biopharmaceutical industry, scale-up/down, cell line selection, multivariate statistical techniques.

## Introduction

In the last decade the Quality by Design (QbD) initiative gained large acceptance as an approach towards development and manufacturing of therapeutic products (Rathore, 2009; Mandenius et al, 2009). This suggests to design and control biopharmaceutical processes in such a way as to consistently achieve assigned product quality. Application of QbD entails in-depth understanding of the manufacturing process, although extended experimental campaigns on pilot-scale or commercial-scale plants are often not feasible. High throughput micro-well bioreactors are commonly utilized to carry out exhaustive experimental campaigns in a cost/time-effective manner. Unfortunately, micro-bioreactors cannot replace the larger production scales. In fact, several biological factors (e.g., clones mutation, contamination), chemical factors (e.g., pH) and physical factors (e.g., configuration, aeration, fluid-dynamics, mixing) cannot be studied at the same time in the micro scales. For this reason, a lot of passages are always utilized for bioprocess scale-up/down from the micro-scales and the intermediate ones up to the pilot and the commercial scale plant.

In this study multivariate statistical methodologies are used to exploit the wealth of information available in the

[*] To whom all correspondence should be addressed: pierantonio.facco@unipd.it

data collected at all scales to both accelerate bioprocesses scale-up/down and aid cell line selection.

**Biopharmaceutical process and available data**

The biopharmaceutical process under investigation is the manufacturing of a monoclonal antibody (IgG1). Nine development scales are considered at increasing size: "static" scales #1 to #6 (micro-well bioreactors) and "shaken/stirred" scales #7 to #9 (stirred bioreactors of increasing volumes). The equipment used is (by increasing scale): (*i*) 24-well plates, 6-well plates, and T25 for the static scales (culture volumes in the range of mL-cL); experiments may be "regular" or "late" at this scales, where late means prolonged; (*ii*) AMBR 15, shake-flask, and 2-liter bioreactors for the shaken/stirred scales (volumes in the range of dL-L). Time profiles of the measurements are available for the (two-week) experiments at these scales in $T = 8$ time points.

The scale-up is carried out on different cell lines. The available data may refer to biological variables or chemical variables. The biological variables (usually related to drug productivity and cell lines stability) are related to cell concentration in the culture (titre), their vitality (viability) and productivity. The chemical variables are related to the observation of the cell status in terms of environmental conditions (pH), availability of nutrients (glucose concentration) and cell "health" (lactate concentration).

**Mathematical methodologies**

The mathematical methodologies used in this study are multiway principal component analysis (MPCA; Nomikos and MacGregor, 1994) for realtime cell line performance monitoring, and joint-Y projection on latent structures (JY-PLS; Garcìa-Muñoz et al., 2005) for cell line performance prediction across scales.

*MPCA*

Principal component analysis (PCA) is a correlative methodology that summarizes the dataset $\mathbf{X}$ $[N\times M]$ of $N$ observations on $M$ variables, by projecting it onto a reduced space of $A<<M$ orthogonal principal components, which describe the direction of maximum variance of $\mathbf{X}$:

$$\mathbf{X}=\mathbf{T}\mathbf{P}'+\mathbf{E} \quad . \tag{1}$$

$\mathbf{P}'$ is the transpose of the $[M\times A]$ loadings matrix, namely the eigenvectors of the covariance of $\mathbf{X}$; $\mathbf{T}$ is the $[N\times A]$ score matrix, i.e., the coordinates of the samples projected onto the PCs; $\mathbf{E}$ is the $[N\times M]$ residual matrix minimized in a least-squares sense. The data in $\mathbf{X}$ are autoscaled, i.e., mean centered and scaled to unit variance.

This methodology can be easily extended to study the dynamics of experiments for which time profiles of the variables are available. To this purpose, multiway PCA (MPCA; Nomikos and MacGregor, 1995) is used. MPCA is a PCA on a multidimensional matrix $\underline{\mathbf{X}}$ $[N\times M\times T]$, where $M$ variables are collected for $N$ experiments in $T$ time instants. MPCA deals with $\underline{\mathbf{X}}$ by performing a PCA on the experiment-wise unfolded matrix, where each experiment corresponds to a different cell line to be tested.

*JY-PLS*

JY-PLS (García Muñoz et al., 2005) relates input datasets $\mathbf{X}_i$ $[N_i\times M_i]$ and $\mathbf{X}_j$ $[N_j\times M_j]$ from different scales $i$ and $j$, through the space of the corresponding response variables $\mathbf{Y}_i$ $[N_i\times U]$ and $\mathbf{Y}_j$ $[N_j\times U]$. The correlation of data within each scale is studied through the relation between $\mathbf{X}_i$ and $\mathbf{Y}_i$, whereas the correlation between scales is assessed through the relation between $\mathbf{Y}_i$ and $\mathbf{Y}_j$. The JY-PLS mathematical formulation is:

$$\mathbf{Y}_J=\mathbf{T}_J\mathbf{Q}_J'+\mathbf{F} \quad . \tag{2}$$

$\mathbf{Q}_J'$ is the transpose of the $[U\times A]$ loadings matrix of the common latent space of $\mathbf{Y}_J$, $\mathbf{T}_J$ is the $[(N_i+N_j)\times A]$ joint score matrix obtained concatenating vertically the scores of each scale, $\mathbf{F}$ is the $[(N_i+N_j)\times A]$ matrix of the residuals. JY-PLS does not pose restrictions on the number and the type of clones $N_i$ and variables $M_i$ measured at each scale $i$. The only constraint is that the responses $\mathbf{Y}_i$ and $\mathbf{Y}_j$ must follow the same statistical distribution. Data pretreatment is performed by autoscaling and dividing each matrix by the square root of its number of elements, to consider the different dataset dimension.

**Results**

*Cell line selection*

An MPCA model was built for the AMBR 15 scale dataset (cell biological and chemical markers) to assess the dynamic behavior of different cell lines at each time sample. Two LVs explained ~ 50% of data variability.

The score space of this model is an effective map of the dynamic behavior of the clones. In Figure 1, MPCA separates the condition of the "golden" (i.e., high productivity) cell lines (green circles), from low productivity ones (black triangles), the latter being typically associated to the smallest values of PC1. These maps are effective in the sense that the most productive cell lines are identified considering jointly the effect of several biological markers (e.g., cell titre and viability). Furthermore, it was found that usually the most productive cell lines are those which are exposed to a larger amount nutrients and that display the lowest lactate concentration (details are not reported for conciseness).

This performance mapping strategy can also be utilized to monitor the cell line cultivations (red and blue open squares of the validation cell lines) over the duration

of the experiment (i.e., along time instants 1-8). Two examples are shown in Figure 1.

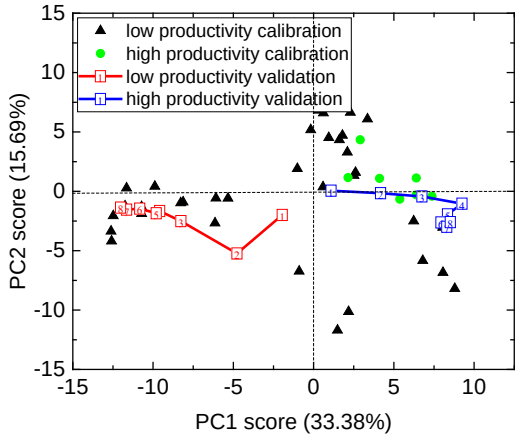


*Figure 1. MPCA model to monitor cell lines productivity in the AMBR15 scale and select the "golden cell lines". Numbers 1-8 within the squares (validation clones) indicate the time sequence of the measurements for a single experiment on a cell line.*

The open red squares refer to a new cell line, whose performance is monitored during an AMBR15 experiment by plotting it onto the model space: the cell line consistently projects onto the low productivity half-plane. The experiment starts at time point 1 close to the origin of the model space, but evolves towards low values of PC1, where poor productivity is typically observed. The conclusion is that this cell line should not be progressed to the upper scales. In a different experiment, a good-performing cell line (open blue squares in Figure 1) starts from similar conditions, but evolves towards high PC1 values, where high productivity lines are typically mapped.

Another valuable indication is obtained by observing that after 4-5 time points (~50% of the experiment duration, or 1 week), the cell line performance stabilizes in a given zone of the score space. This indicates that the experiment might be shortened with no loss of information on cell line selection.

*Prediction of cell line behavior*

It was found that all the most important productivity variables can be estimated from the other variables measured at the same scale, if the information available from other scales is used jointly (JY-PLS modeling). Prediction aims at reducing the number of measurements required in an experimental campaign. A leave-one-out procedure was used to test prediction across scales. Only results on the static scales #1 to #6 are provided here, and cell viability is estimated. The estimation accuracy is satisfactory: the determination coefficients in validation range between 0.86 and 0.98 for all scales. Furthermore, the estimation error $e$ is much smaller than the intrinsic variability of the response variable (standard deviation $s_Y$ of cell viability; Table 1).

*Table 1. Estimation accuracy of the JY-PLS cell viability estimation across static scales.*

| static scale # | 100 ($e/s_Y$) |
|---|---|
| 1 | 10.4 |
| 2 | 9.8 |
| 3 | 3.0 |
| 4 | 2.0 |
| 5 | 3.1 |
| 6 | 1.7 |

The proposed methodology is effective and especially useful when the number of experiments within one scale is very small (< 6), allowing for a significant reduction on the number of experiments.

## Conclusions

In this study, multivariate statistical approaches supporting biopharmaceutical process scale-up/down were presented. In particular, a methodology for the selection of the "golden cell lines" was developed together with a methodology for the estimation of the most important biological variables across different scales within the development process.

The golden cell lines selection identified the most promising cell lines with respect to drug productivity by means of a monitoring model based on multiway principal component analysis. Furthermore, it was found that the duration of an experiment can be reduced.

Finally, a method to estimate the most important productivity variables from the information embedded in each scale demonstrated to be promising in accelerating the process scale-up/down.

## References

García Muñoz, S., J. F. MacGregor, T. Kourti (2005). Product transfer between sites using Joint-Y PLS, Chemom. Intell. Lab. Syst., 79, 1–2, 101–114.

Mandenius, C. F., K. Graumann, T. W. Schultz, A. Premstaller, M. Olsson, E. Petiot, C. Clemens, M. Welin (2009). Quality-by-Design for biotechnology-related pharmaceuticals, Biotechnol. J., 4, 600–609.

Nomikos, P., J. F. MacGregor (2005). Multivariate SPC Charts for Monitoring Batch Processes», Technometrics, 37, 1, 41–59.

Rathore, A. S. (2009). Roadmap for implementation of quality by design (QbD) for biotechnology products, Trends Biotechnol., 27, 9, 546–553.