

# BioPharm INTERNATIONAL

October 2007

The Science &amp; Business of Biopharmaceuticals



## Applications of Multivariate Data Analysis in Biotech Processing

Rob Johnson, Oliver Yu, Alime Ozlem Kirdar,  
Alagappan Annamalai, Sanjeev Ahuja, Kripa Ram, Anurag S. Rathore

### ABSTRACT

Multivariate data analysis (MVDA) is quickly gaining popularity both in basic research and applied scientific fields as a statistical method of choice for examining variable interactions that were previously undefined. Multivariate data analysis by means of projection and regression methods overcomes challenges associated with applications such as multidimensionality of the data set, missing data, and variation introduced by disturbing factors such as experimental error and noise. This article presents how four of the major biotech companies, Amgen, Genentech, Wyeth Biotech, and MedImmune, are using multivariate analysis to solve problems encountered in biotech processing.



**Anurag S. Rathore, PhD**, is the director of process development and

**Alime Ozlem Kirdar** is a senior engineer, both at Amgen, Inc., Thousand Oaks, CA, 805.447.1000, asrathore@yahoo.com.

**Rob Johnson** is an engineer II at Genentech Inc. **Oliver Yu** is a senior engineer at Zymergi LLC. **Alagappan Annamalai** is an associate director at Wyeth Biotech, Andover, MA. **Sanjeev Ahuja** is a senior scientist and **Kripa Ram** is a vice president, both at MedImmune, Gaithersburg, MD.



Listen to a podcast interview with Anurag Rathore at [biopharminternational.com/biopharmnow](http://biopharminternational.com/biopharmnow)

As the biotech industry moves toward implementing the initiative of Quality by Design, using statistical tools for design of experiments and for data analysis is becoming necessary. It has been pointed out that biopharmaceutical manufacturing data is complex and univariate or bivariate analysis can often be inefficient and result in misleading conclusions.<sup>1,2</sup> Principal component analysis (PCA), partial least squares (PLS), and multiple regression are some of the commonly used projection and regression methods in MVDA. Additionally, multivariate statistical process control (SPC) charts are useful in routine monitoring of manufacturing processes.

Recently, several studies have addressed the topic of performing multivariate analysis on data from fermenta-

tion and cell culture operations.<sup>3-5</sup> This article is the tenth in the "Elements of Biopharmaceutical Production" series and presents how four of the major biotech companies, Amgen, Genentech, Wyeth Biotech, and MedImmune, are using multivariate analysis to solve problems encountered in biotech processing.

### USE OF MVDA TO OPTIMIZE LARGE-SCALE PRODUCTION CULTURE PERFORMANCE

**Rob Johnson and Oliver Yu, Genentech, Inc.**

The multiple linear regression of historical data from a licensed antibody process revealed opportunities for improved production culture performance and more consistent product quality by optimizing process parameters in their acceptable ranges stated in the product filing. This application demonstrates the utility of MVDA for capitalizing on these opportunities for a signifi-

cant increase in production culture yield and a sustained decrease in process variability.

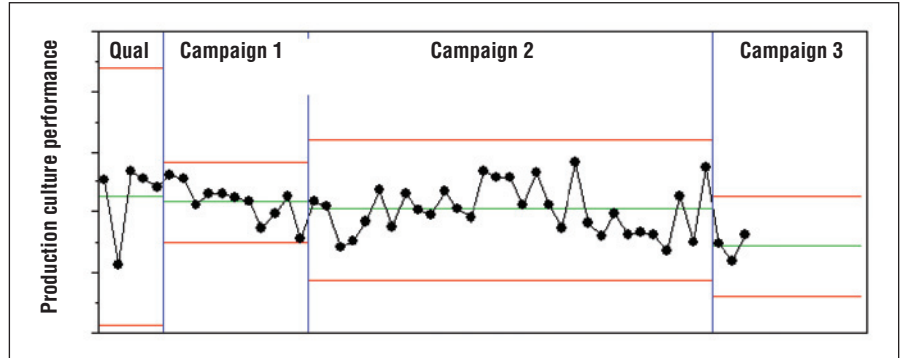
At the start of the third manufacturing campaign of a commercial antibody process, production culture productivity was 70% of the previous campaign average, as shown in Figure 1. This observed decrease in culture productivity and increase in process variability led to an effort to understand and reduce the sources of variability in the process.

The multivariate approach used here was simple multiple linear regression (MLR) achieved in an iterative backwards stepwise fashion against data from the previous campaign. The assumption of linearity is often acceptable since nonlinear relationships are well-approximated by a line when constrained in a small range, as is the case for commercial production. The iterative backwards stepwise approach minimizes expert bias because the computer iterates toward maximum correlation and minimum model error.

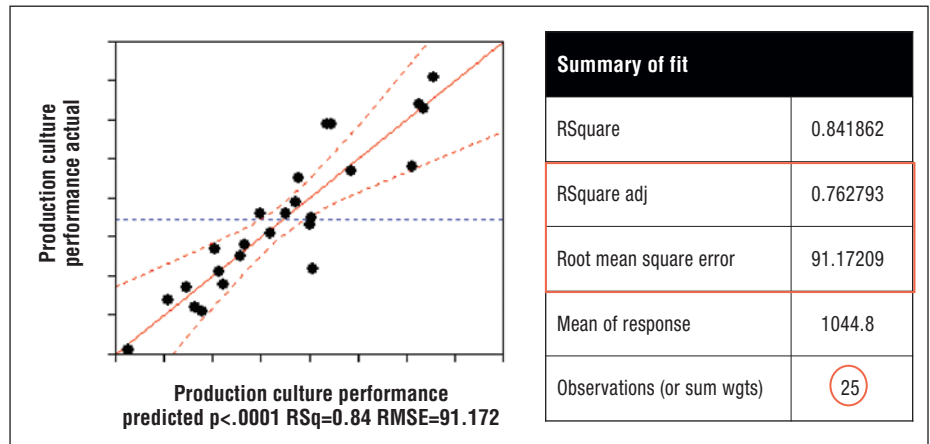
Expert bias exists in the selection of initial factors to include in the model, but this bias is often constrained to a set of actionable parameters, which can be addressed on the production floor to mitigate unexpected process behavior. The inclusion of nonactionable factors in the model, such as maximum lactate production or initial glucose uptake rate, yields scientifically relevant ideas for further investigation, but is not as useful for immediately improving production culture performance and achieving production campaign targets. Actionable process parameters were studied from 25 batches of campaign 2 and were used to build the model shown in Figure 2, Table 1A, and Table 1B.

We know from our historical experience that models with an adjusted  $R^2$  of 0.76 and the root mean square error (RMSE) of 91 are capable of supporting

**Figure 1.** Control chart of an antibody process shows production cultures performing at 70% of the previous campaign



**Figure 2.** Multiple linear regression model built on data from campaign 2 (model generated using JMP 6.0)



**Table 1A.** Parameter estimates for data presented in Figure 2.

Parameter estimates				
Term	Estimate	Std error	t Ratio	Prob> t
Intercept	24830.71	13303.51	1.87	0.0804
Initial concentration	21.231987	7.051392	3.01	0.0083*
Culture timing 1	-155.3026	174.566	-0.89	0.3868
Culture timing 2	546.73332	188.1104	2.91	0.0103*
Culture timing 3	1.9409029	1.699762	1.14	0.2703
Operating parameter 1	-3590.111	1901.787	-1.89	0.0773
(Initial concentration:13.5665) x (Culture timing 3:242.563)	0.8945541	0.47628	1.88	0.0787
(Initial concentration:13.5665) x (Operating parameter 1:6.95574)	-906.6274	411.3248	-2.20	0.0425*
(Culture timing 1:0.76612) x (Culture timing 3:242.563)	-29.51671	10.61991	-2.78	0.0134*

\*Significant main effects and interactions that impact cell culture productivity

data-based decisions. The linear equation was solved to determine the magnitude and direction of parameter change necessary to maximize culture performance.

We re-optimized the parameters to maximize culture productivity and minimize a product quality attribute (in this case with a maximum but no minimum specification).

The model achieved an optimum and prescribed specific targets for these parameters.

Figure 3 indicates the actions required to achieve this optimum:

- Increase initial concentration of a production culture parameter
- Decrease the time at which culture operation 1 occurs

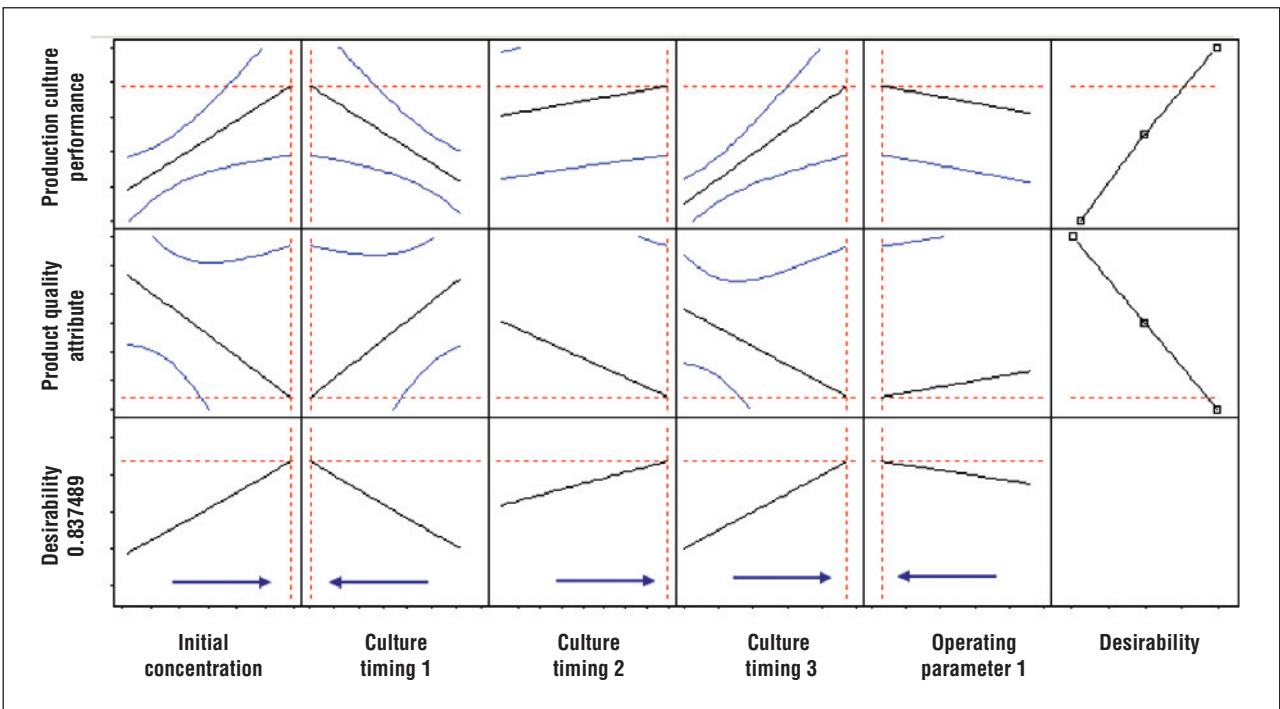
**Table 1B.** Scaled estimates for data presented in Figure 2.

Scaled estimates					
Continuous factors centered by mean, scaled by range/2					
Term	Scaled estimate		Std error	t Ratio	Prob> t
Intercept	1019.2476		20.20316	50.45	<.0001*
Initial concentration	150.61441		50.02081	3.01	0.0083*
Culture timing 1	-47.56141		53.46083	-0.89	0.3868
Culture timing 2	211.03038		72.60763	2.91	0.0103*
Culture timing 3	91.222438		79.88882	1.14	0.2703
Operating parameter 1	-77.48656		41.0469	-1.89	0.0773
(Initial concentration:13.5665) x (Culture timing 3:242.563)	298.24994		158.7947	1.88	0.0787
(Initial concentration:13.5665) x (Operating parameter 1:6.95574)	-138.8108		62.97662	-2.20	0.0425*

\*Significant main effects and interactions that impact cell culture productivity

- Increase the time at which culture operation 2 occurs
  - Increase the time at which culture operation 3 occurs
  - Decrease the set point of operating parameter 1.
- The parameter changes were implemented incrementally and thus resulted in a gradual improvement in culture productivity.

**Figure 3.** The model generates the direction and magnitude for each factor in order to maximize production culture performance and minimize a product quality attribute.



**Figure 4.** Performance of campaign 3 shows initial lower-than-expected productivity that was overcome by process parameter changes. The variability of culture performance was 25% RSD.

mization to reduce process variability and achieve predictable performance.

**USE OF MVDA FOR ESTABLISHING PROCESS COMPARABILITY AND TROUBLESHOOTING**

**Alime Ozlem Kirdar and Anurag Rathore, Amgen Inc.**

This application involved multi-variate analysis of data from small-scale (2-L) and large-scale (2000-L) cell culture batches.<sup>3</sup> A commercially available MVDA

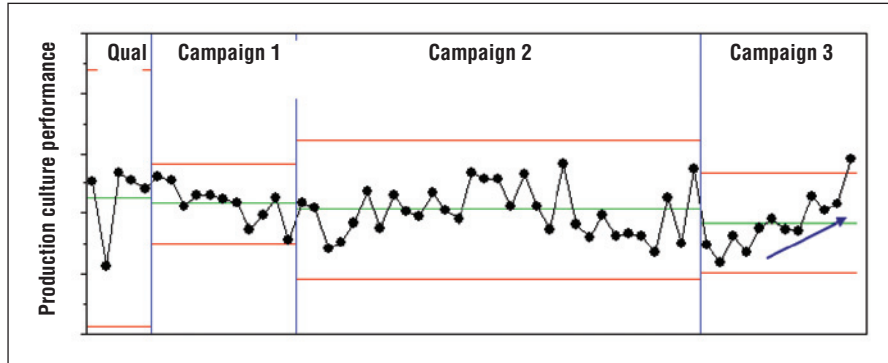


Figure 4 demonstrates the gradual improvement in culture performance during campaign 3, which ultimately resulted in production of 104% of the campaign goal. While campaign 3 of this antibody process met goals through incremental parameter changes, the process variability was unacceptably large at 25% relative standard deviation (RSD). For the fourth campaign, the process parameters were evaluated and targets were reset with capability in mind. Most parameters remain changed in the direction prescribed by the MLR, but the magnitude decreased in order to maintain capability and minimize process variability.

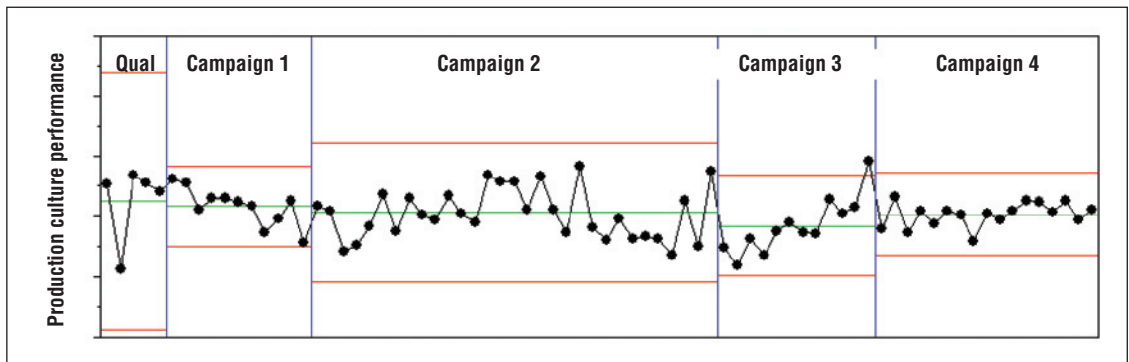
Campaign 4 reliably achieved its production goals and produced 110% of the campaign goal with a productivity improvement of 10% over campaign 3, sustained product quality, and a 15% decrease in process variability. This is shown in Figure 5.

As shown in this application, MVDA can be a useful tool for continuous process improvement and long-term process understanding. Furthermore, it can be used for process opti-

software package, SIMCA P+ 11 version 11.0.0.0 (Umetrics AB, Kinnelon, NJ), was used to perform the multivariate analysis. Daily offline metabolic and cell growth measurements from 14 center point runs (2-L scale) and 11 center point runs (2000-L scale batches) were analyzed separately by partial least squares (PLS) modeling. Several input parameters (pCO<sub>2</sub>, pO<sub>2</sub>, glucose, pH, lactate, ammonium ions) and output parameters (percent purity, viable cell density, percent viability, osmolality) were included in the analysis.<sup>3</sup> Loadings plot and variable importance for the projection plots were used to evaluate process comparability across scales.

The loadings plot shows the PLS loadings computed for each of the x variables. The variables with the largest absolute values of principal components (p1 or p2) are situated far away from the origin (on the positive or negative side) on the plot and dominate the projection. The farther we are from the center (0,0) in the loadings plot, the greater the impact of input parameters on the performance of the cell culture or the greater the

**Figure 5.** Performance of campaign 4 shows sustained high culture performance with low variability; Relative standard deviation = 9.8%.





impact of the cell culture process on the output parameter. Also, variables near each other (in the same quadrant) are positively correlated and those opposite to each other (opposite quadrants) are negatively correlated.<sup>3</sup>

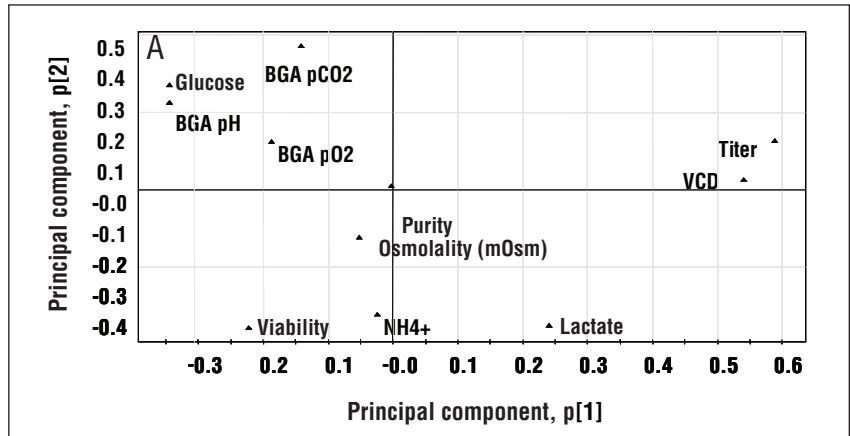
Figure 6A presents this plot for data from the 2-L scale. It is seen that the cell culture process has a significant impact on viable cell density (VCD), titer, and viability. Titer and VCD are in the same quadrant and this is supported by the observation that both of them are lower in the earlier stages of the culture and increase as the culture progresses. In contrast, viability and VCD are in the opposite quadrants and this implies that at earlier stages of the culture, when the VCD is lower, the viability is higher and this reverses in the latter part of the process. Of all the input parameters examined, pH, pCO<sub>2</sub>, glucose, and lactate levels have a significant effect on the performance of the cell culture process. Also, pH, glucose, and pCO<sub>2</sub> levels have a similar effect on the performance of the cell culture process, whereas lactate has the reverse effect. Figure 6B presents the loadings plot for the data from the 2000-L scale. For most of the output parameters—namely titer, VCD, and purity—the plot in Figure 6B is quite comparable

with that in Figure 6A. However, differences are seen for the loadings of pO<sub>2</sub>, osmolality, NH<sub>4</sub><sup>+</sup>, and lactate levels, suggesting changes in the cell-culture metabolism upon scale up. It is well known in the literature that gas transfer is less efficient at large scale leading to a build up of CO<sub>2</sub> in the vessel. This results in an increased use of base in order to maintain pH at the intended set point, and the increased base addition leads to higher osmolality. Metabolic response of cells can also be amplified, which may lead to higher NH<sub>4</sub><sup>+</sup> and lactate levels.<sup>6–7</sup>

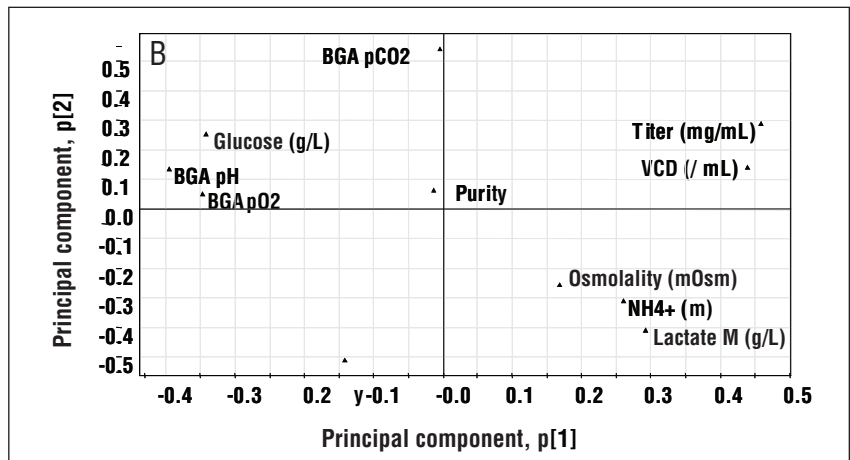
These observations led to further investigation of the differences observed during scale up and correction of some of those differences.

Variable importance for the projection (VIP) plot shows the relative importance of each variable included in the analysis. Figure 7A presents this plot for the 2-L scale data set. Consistent with the observations made from the corresponding loadings plot (Figure 6A), it is seen that this cell-culture process has a strong influence on the

**Figure 6A.** Partial least squares loadings plots for a 2-L bioreactor<sup>3</sup>



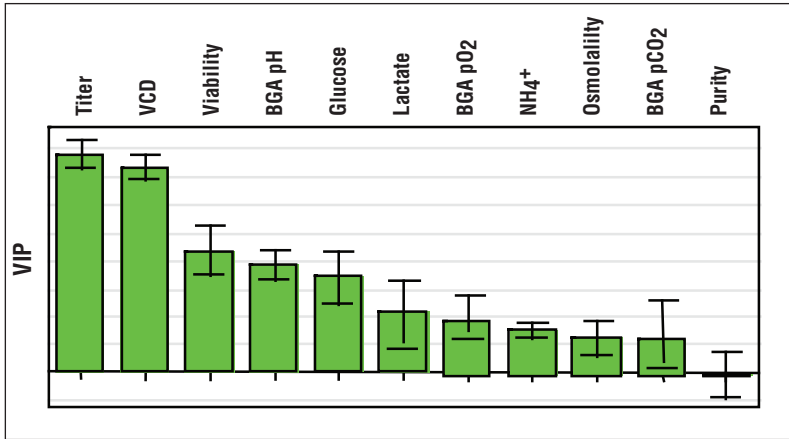
**Figure 6B.** Partial least squares loadings plots for a 2,000-L bioreactor<sup>3</sup>



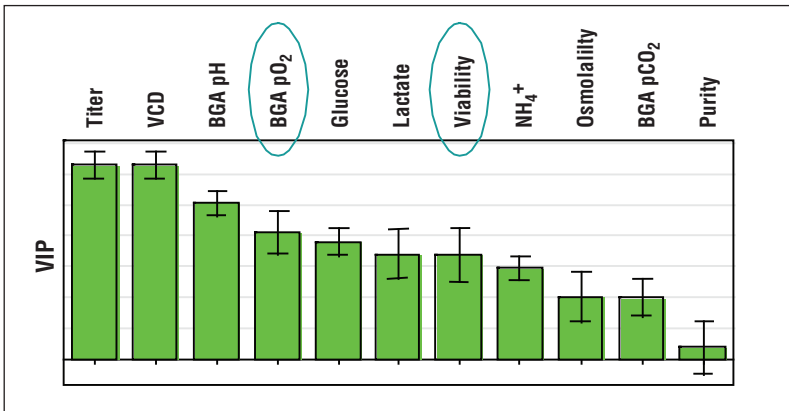
titer, VCD, and viability of the broth at the end of the process. Of the input parameters, pH, lactate, and glucose levels were found to have the greatest effect on process performance. On the contrary, osmolality has one of the lowest loadings in both principal component directions (p1 and p2). Figure 7B presents the VIP plot for the 2000-L scale data set. Comparison of the VIP plot for the 2-L and 2,000-L data yields similar conclusions as mentioned above for discussion of the loadings plots. The most significant difference is seen in the VIP score for pO<sub>2</sub> for the reasons mentioned above.

In summary, it is shown that MVDA can be a useful tool for evaluating process comparability across scales, equipment, or facilities. Although the loadings plot provides a qualitative assessment, the VIP plot is more quantitative. Data analysis can easily be used for troubleshooting issues encountered during scale up or technology transfer by identifying the differences and helping focus the investigation.

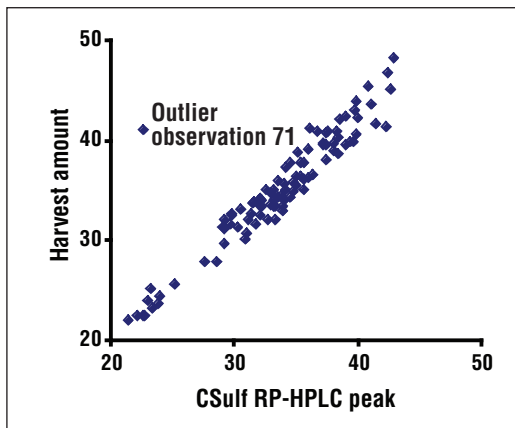
**Figure 7A.** Variable importance for the projection (VIP) plots for a 2-L bioreactor. Adapted from reference 3.



**Figure 7B.** Variable importance for the projection (VIP) plots for a 2,000-L bioreactor. Adapted from reference 3.



**Figure 8.** Bivariate plot showing an outlier



**USE OF MVDA FOR ROUTINE MONITORING OF MANUFACTURING PROCESSES**

**Alagappan Annamalai, Wyeth Biotech**

In this multivariate statistical process control (SPC) application, MVDA was performed on the data

from a legacy protein purification process. Custom software was developed at Wyeth Biotech in the form of an Excel add-in to create two kinds of multivariate control charts, Hotelling's  $T^2$  and multivariate exponentially weighted moving average (MEWMA).<sup>8-10</sup> Hotelling's  $T^2$  monitors individual process observations while MEWMA monitors shifts and drifts in the process. Process data residing in Excel worksheets can be used directly with the software. The software interfaces Excel with custom-developed functions and the run-time version of MATLAB (The MathWorks, Inc., Natick, MA) through a component object model (COM) object.<sup>11</sup> Excel serves as the user interface while the MATLAB functions perform the multivariate calculations. The Excel add-in was developed using Visual Basic for Applications (VBA) and the COM object was developed using Visual C++.

Eight parameters of the protein purification process were examined initially. Those parameters were harvest volume, harvest amount, CSulf RP-HPLC, B-sepharose recovery, overall recovery, specific activity, peptide map sub-unit percent, and DS rapid acidic C4 RP-HPLC. We used 100 process vectors in the training set and then monitored the next one hundred vectors. Hotelling's  $T^2$  produced one signal for the monitoring observation 71. The bivariate plot between harvest amount and CSulf RP-

HPLC in Figure 8 shows that observation 71 undoubtedly is an outlier. Harvest amount and CSulf RP-HPLC measure the amount of protein between two successive process steps and exhibit a very high correlation. Because of the high correlation, we repeated the multivariate statistical process control calculations after excluding the parameter "harvest amount" but using the remaining seven parameters. Interestingly, this too flagged observation 71 as an outlier. During the seven-parameter training phase, the two observations above the training  $T^2$  limit in Figure 9 were discarded and the remaining 98 observations were used to estimate the covariance matrix. It is surprising to note that the theoretical MEWMA limit for a typical ARL of 370 ( $\alpha = 0.0027$ ) with  $\gamma = 0.1$  was found to be too tight in Figure 10. In Figures 11 and 12, both the first 100 training observations and the next 100 monitoring observations are plotted. As noted above,  $T^2$  for the monitoring observation 71 exceeds the control limit in Figure 11. Despite the theoretical

MEWMA limit being too tight, Figure 12 indicates that the process has drifted from the process mean calculated with the training data.

The mean square of successive differences (MSSD) was used to remove the influence of process drifts in the covariance estimates.<sup>12–14</sup> Hotelling's  $T^2$  chart in Figure 13 based on MSSD suggests that a process shift occurred during the second half of the training data and persisted throughout. MSSD appears to be too sensitive for  $T^2$  but we do not know if this will be true for a large training data set or for other biopharmaceutical processes. At present, a control limit for MEWMA calculated with MSSD can be set only arbitrarily, as done in Figure 14.

We have shown that multivariate statistical process control (SPC) tools are useful in ongoing

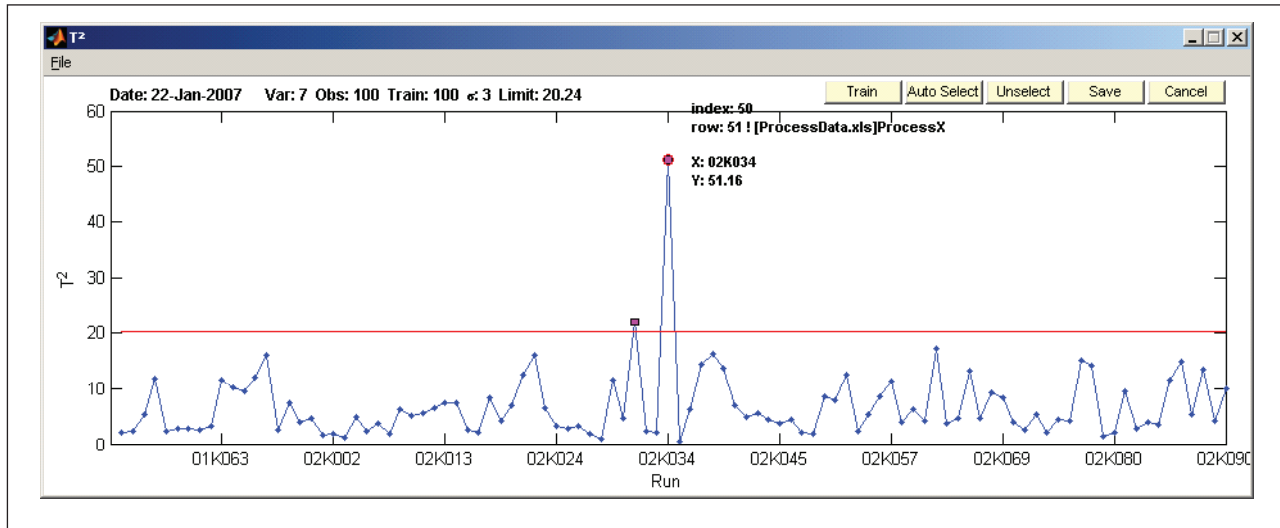
monitoring of manufacturing processes. These tools can provide early warning of process problems before they become severe. In this case study, Hotelling's  $T^2$  identified an unusual production batch (observation 71) during monitoring that would have otherwise gone unnoticed. This scenario is possible for multiple consecutive batches also. MEWMA revealed small process drifts that were previously hidden. Understanding the origins of these drifts will provide opportunities to improve the process further.

## USE OF MVDA FOR RAW MATERIAL CHARACTERIZATION AND SCREENING

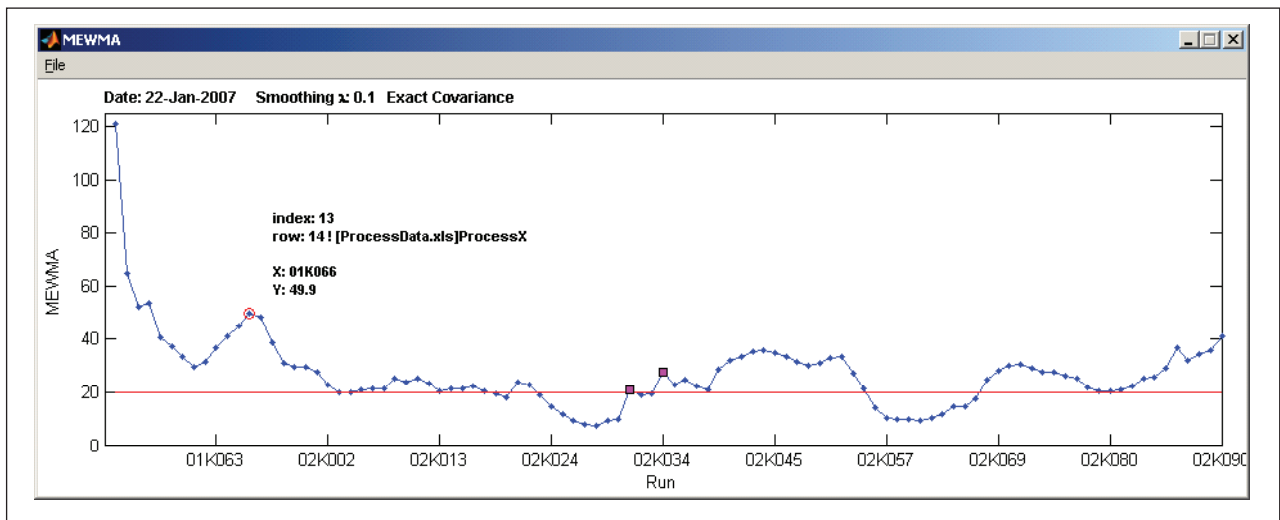
**Sanjeev Ahuja and Kripa Ram, MedImmune**

Many industrial mammalian-cell-growth media rely upon the inclusion of a serum fraction,

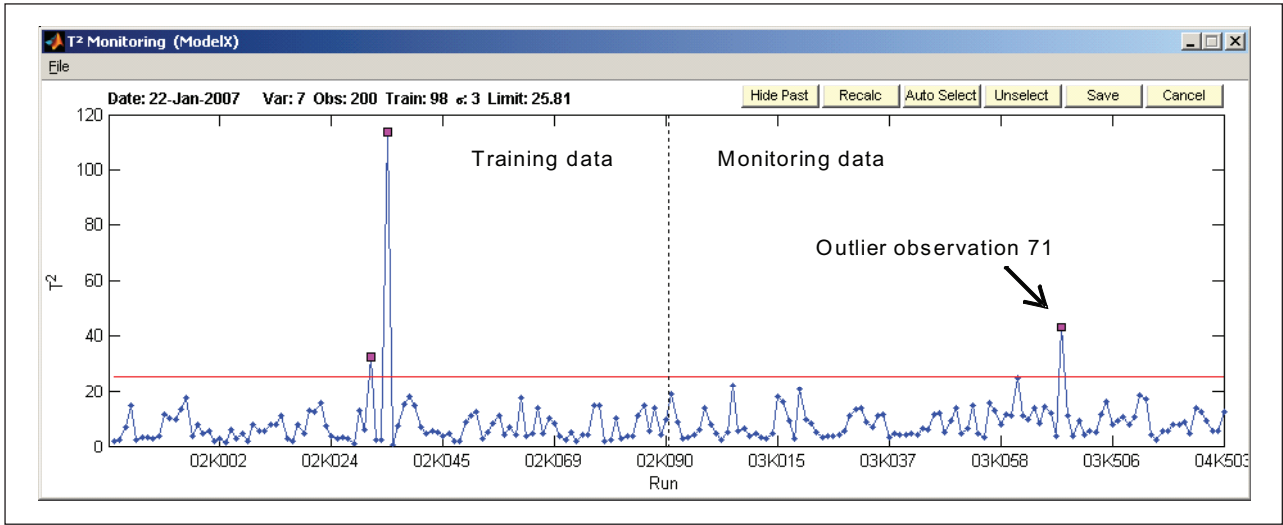
**Figure 9.** Hotelling's  $T^2$  chart during training



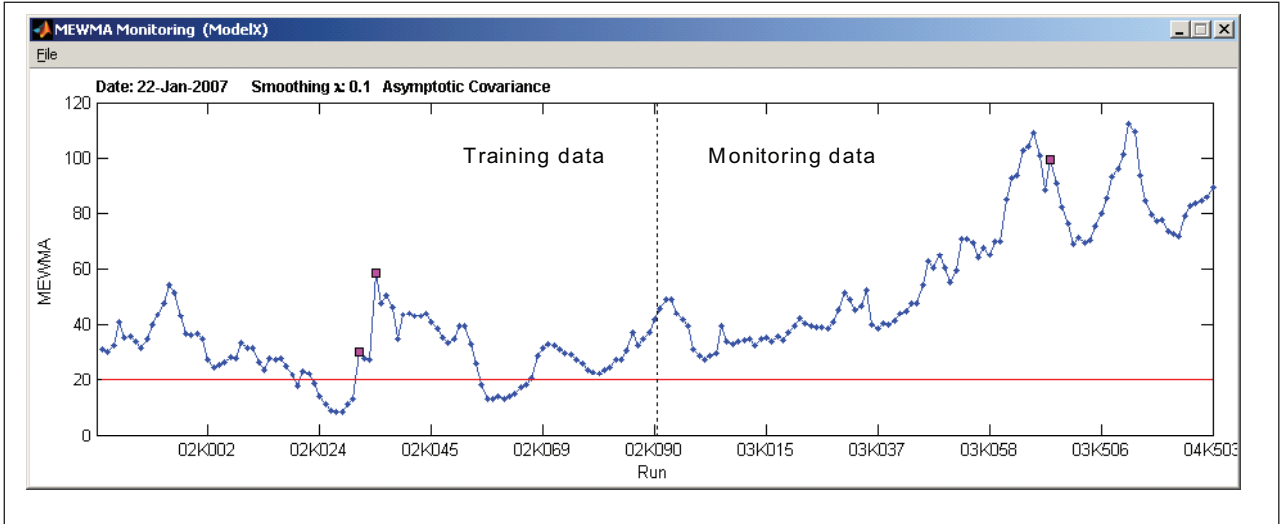
**Figure 10.** Multivariate exponentially weighted moving average chart during training



**Figure 11.** Hotelling's  $T^2$  chart during monitoring



**Figure 12.** Multivariate exponentially weighted moving average chart during monitoring



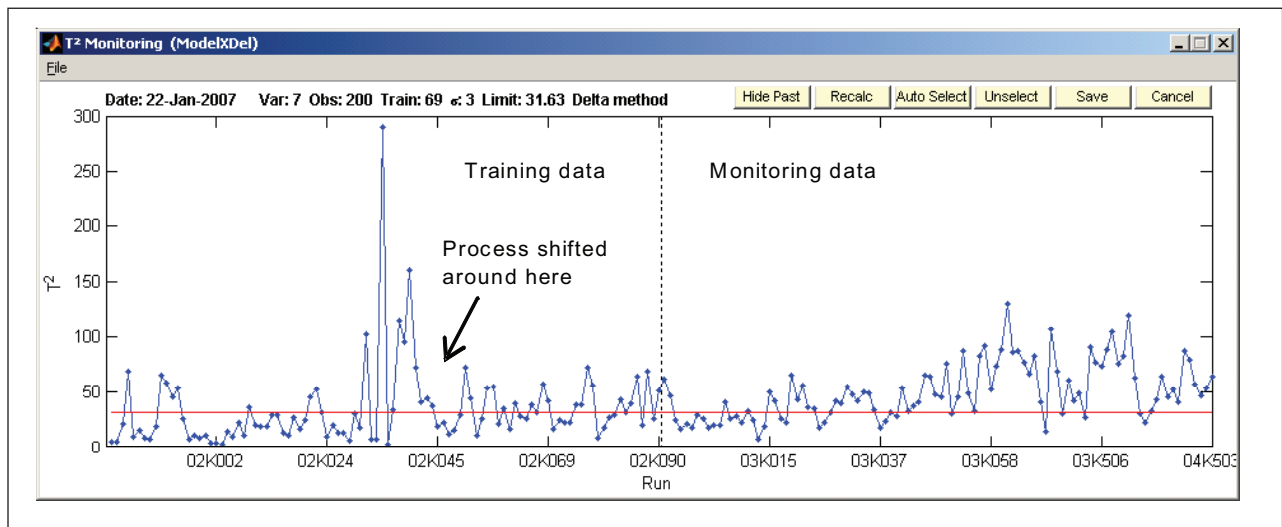
commonly known as the lipoprotein fraction, to ensure the availability of cholesterol to the cells. The cholesterol contained in such serum fractions is associated with various lipoproteins that act as carriers of cholesterol. The lipoproteins also provide a means by which the cholesterol can be solubilized in a hydrophilic environment. Since these fractions are derived from serum, it is likely that the compositions of these fractions are highly variable, which, in turn, can contribute to the variation in process outcome and product quality attributes. To understand this raw material further and how it might affect process productivity, detailed analytical characterization and cell culture experiments were carried out. This case study shows how multivariate analysis can be used to understand various lipoprotein frac-

tion products and to assist in raw material selection and control for a cell culture process.

Analytical characterization of various lipoprotein fraction lots involved four distinct assays: lipid profiling, fatty acid analysis, lipoprotein analysis, and lipid oxidation. A brief description of these tests is given below.

- Lipid profiling results were obtained via an HPLC-based method for cholesterol esters (CE), free fatty acids (FFA), free cholesterol (FC), and phosphatidylcholine (PC). The total cholesterol (TC) values were derived from the results of free cholesterol and cholesterol esters. The levels of total cholesterol for all the tested lots lay in the pre-specified range required for raw material release.
- Fatty acid analysis was performed via a GC



**Figure 13.** Mean square of successive differences based Hotelling's  $T^2$  chart during monitoring

method and quantitative results were obtained for non-essential, n-6 essential, and n-3 essential fatty acids. The results for [n-6]:[n-3] ratios and total fatty acids (TF) were subsequently calculated.

- Various lipoproteins were identified by an SEC technique. The three distinct species in these fractions included high density lipids (HDL), low density lipids (LDL), and very low density lipids (VLDL). The ratios LDL:HDL and VLDL:HDL were derived from the original data.
- The extent of lipid oxidation (LO) was also determined because it can affect the cell productivity.

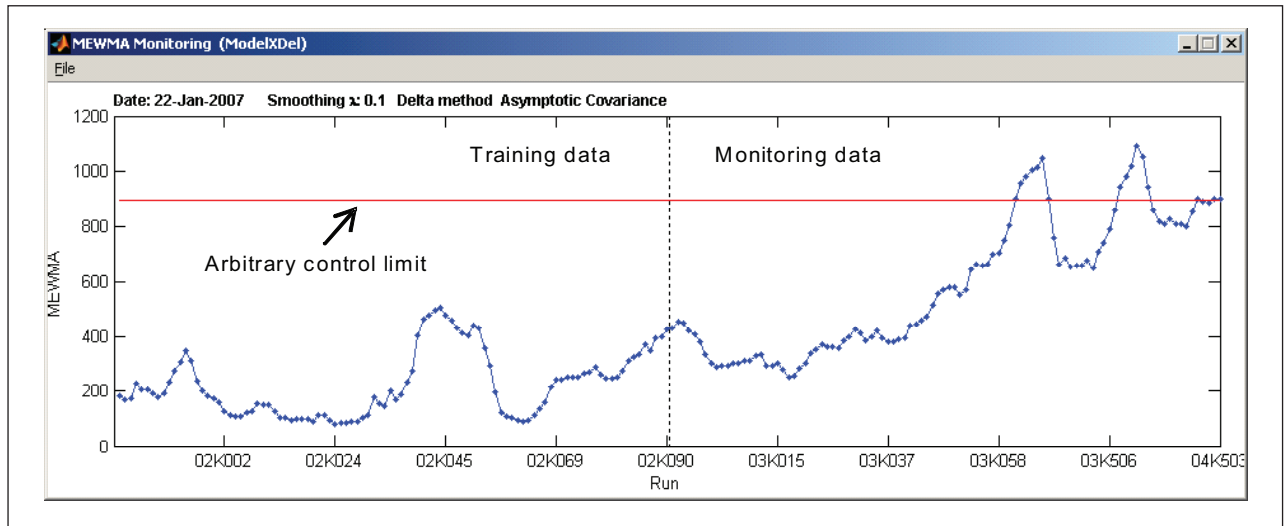
A total of 16 analytes were tested for each lot using the above-mentioned assays.

It is difficult to predict how the composition might be related to cell productivity because the individual effects of each of the tested analytes on the cell culture performance is poorly understood. Also, it is impractical to conduct experiments to evaluate the effects of various components because of their biological complexity and solubility issues. The use of multivariate analysis, however, offers a methodology by which the lot-to-lot variability of the above-mentioned complex raw material can be assessed. It also provides a method to integrate all the analytical data and to understand the key analytical differences between products from different vendors. Moreover, the number of available lots and analytical and experimental resources are often limited, which makes the use of multivariate analysis particularly useful. Finally, as will be discussed later, it can be used

to identify the key raw material quality attributes required for desired productivity. These quality attributes can help focus efforts in the right direction to continually improve a raw material and hence the manufacturing process it is used in. For the multivariate analysis described in this case study, the software **SIMCA-P+** (Version 11) from Umetrics, Inc. (Kinnelon, NJ) was used.

To characterize the two products that have been successfully used for an antibody-producing cell culture process, principal component analysis was used. The data from a total of 17 lots—12 lots from vendor A (called product A) and 5 lots from vendor B (called product B)—were used for this analysis. The raw material type (product A or product B) was treated as a qualitative  $X$  variable and all the analytical data was used as continuous  $X$  data. The first principal component was able to explain 61% of variation in  $X$  data ( $R^2X = 0.61$ ), and this component was able to predict 54% of the variation in data ( $Q^2 = 0.54$ ). The addition of a second principal component resulted in the cumulative goodness of fit ( $R^2X$ ) and predictability ( $Q^2$ ) values of 0.77 and 0.50. Because the addition of second principal component decreased the model predictability to 0.50, this component was ignored and only the first component was used for analysis. The principal component analysis showed that the two products were distinct and belonged to different clusters in  $M$ -space. The correlation structure was analyzed using the loading plot. Product B, n-6 fatty acids, [n-6]:[n-3] ratio, and free fatty acids clustered together and appeared diago-

**Figure 14.** Mean square of successive differences based multivariate exponentially weighted moving average chart during monitoring



nally opposite to the cluster containing LDL, LDL:HDL, VLDL, VLDL:HDL, and PC. These results suggested that compared to product A, product B was richer in n-6 fatty acids and free fatty acids and had a higher [n-6]:[n-3] ratio. On the contrary, product A was richer in n-3 fatty acids and phosphatidylcholine. Product A also had a higher content of LDL and VLDL and had higher LDL:HDL and VLDL:HDL ratios compared to product B. It is likely that the dissimilarities between products A and B are due to the differences in starting sera that in turn may be related to the feeding practices of different herds used for serum production.

To assess the feasibility of using an alternate raw material for the process, the analytical data of four lots from a new vendor (called product C) was integrated in the PCA model developed earlier. The resulting PCA model resulted in two principal components and had a cumulative goodness of fit ( $R^2(X)$ ) and predictability ( $Q^2$ ) values of 0.71 and 0.57, respectively. Remarkably, the new product did not appear similar to either product tested earlier and belonged to a distinct cluster in M-space (Figure 15). A closer analysis showed that the new product appeared closer to product A in fatty acid composition and closer to product B in phosphatidylcholine content and lipoprotein analysis. The loading plot (Figure 16) showed that product A was uniquely identified by phosphatidylcholine content, LDL, LDL:HDL ratio, VLDL, and VLDL:HDL ratio. Also, product B was uniquely identified by the content of n-6 fatty acids and the [n-6]:[n-3] ratio. Because

product C appeared analytically different from the other two products, a cell culture use test was performed. Two lots of products A and B each and one lot of product C were tested in parallel. Results indicated that product C resulted in significantly lower titer compared to products A and B. The average titer with product C was approximately 30% lower than that obtained with products A and B. Because of lower titer obtained with product C, it was considered unacceptable for use.

Although product C was deemed unacceptable for use in manufacturing, the experimental data provided an opportunity to establish the critical components required to attain desired productivity. To identify these components, a partial least squares (PLS) model was developed using the data in the experiment described earlier. The PLS model used 10 experimental observations (five lots described above, tested in duplicate) and analytical data of 16 analytes. The model resulted in three principal components that could explain 96% of the variation in  $X$  data ( $R^2(X) = 0.96$ ) and resulted in cumulative  $R^2(Y)$  and cumulative  $Q^2(Y)$  values of 0.96 and 0.89, respectively. The critical quality attributes were identified by reviewing the coefficient plot (Figure 17) and the variable importance plot (Figure 18). Note that the coefficients in the coefficient plot indicate the scaled and centered data with confidence intervals derived from jackknifing. The variable importance for the projection (VIP) values reflect the importance of terms in the model with respect to  $Y$ , i.e., its correlation to titer, and with respect to  $X$  (the

projection). Terms with VIP values greater than 1 are the most relevant in explaining Y. For the PLS model described above, the terms with VIP values larger than 1 were n-3 fatty acids, LO, HDL, and FC. However, a review of the coefficient plot pointed to the large variation associated with HDL and FC data. As a result, only n-3 fatty acids and lipid oxidation (LO) were determined to be the critical quality attributes. The effects of n-3 fatty acids and lipid oxidation on process productivity are plausible as literature references of such effects on certain cell types are available.<sup>15-16</sup>

In summary, multivariate analysis for raw material selection and control serves multiple purposes. First, it categorizes an incoming product based on the analytical data alone, without the need to evaluate it using time-consuming cell culture studies. This classification is quite important because scale-down studies may not result in picking up the differences in raw materials even though they are different analytically. Performing the multivariate analysis on the analytical data provides another criterion for deciding if a raw material from a particular vendor or source is acceptable for use in manufacturing. Second, MVDA can be applied to the combined experimental

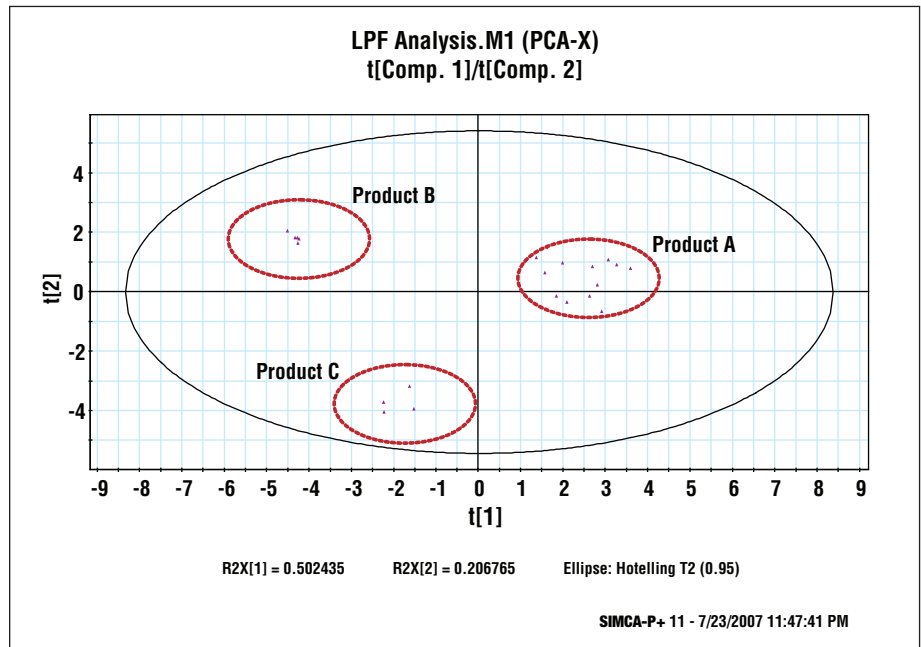
and analytical data to identify the critical components required for desired outcome, e.g., productivity. After sufficient analytical and experimental data are gathered, multivariate analysis can be used as the sole criterion for assessing the raw material quality. It can also assist in directing the efforts to improve the quality of a suboptimal raw material (e.g., product C in the current study). Finally, the multivariate analysis also helps limit the scope of analytical testing for raw material control. For example, in the case study described here, only

two assays may be needed for future products (or lots) instead of the four used to develop the model earlier.

### SUMMARY

This article demonstrates the usefulness of the MVDA with respect to various activities involved in biopharmaceutical manufacturing, including scale up, process comparability, process optimization, process monitoring, and raw material testing. Currently, a lot of data collected at small and large scale do not

**Figure 15.** Principal component analysis (t[1] versus t[2])



**Figure 16.** Loading plot (p[1] versus p[2])

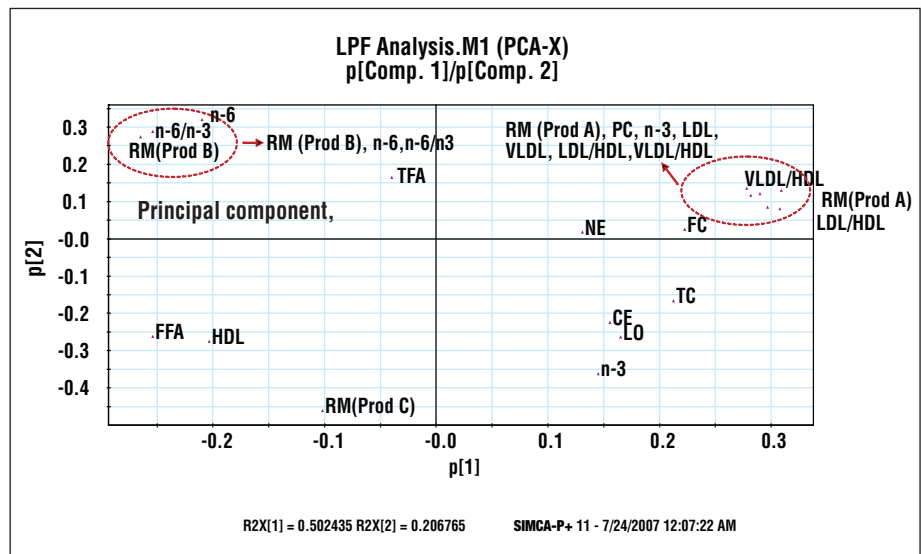


Figure 17. Coefficient plot (partial least squares model of titer)

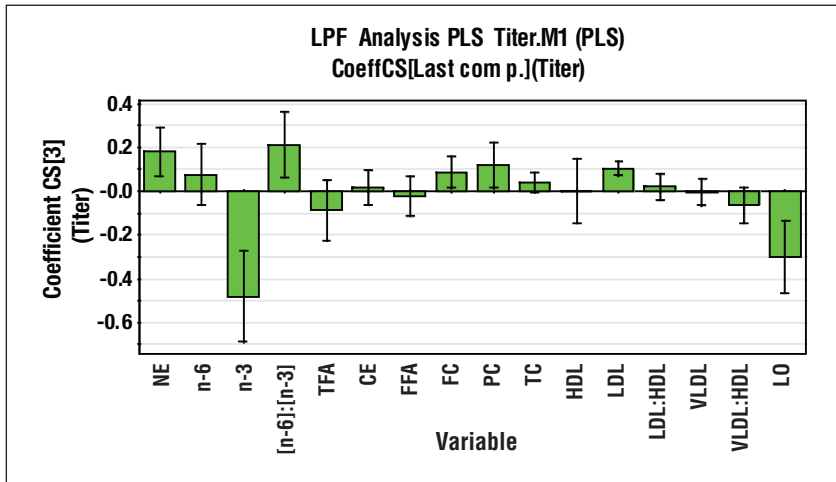
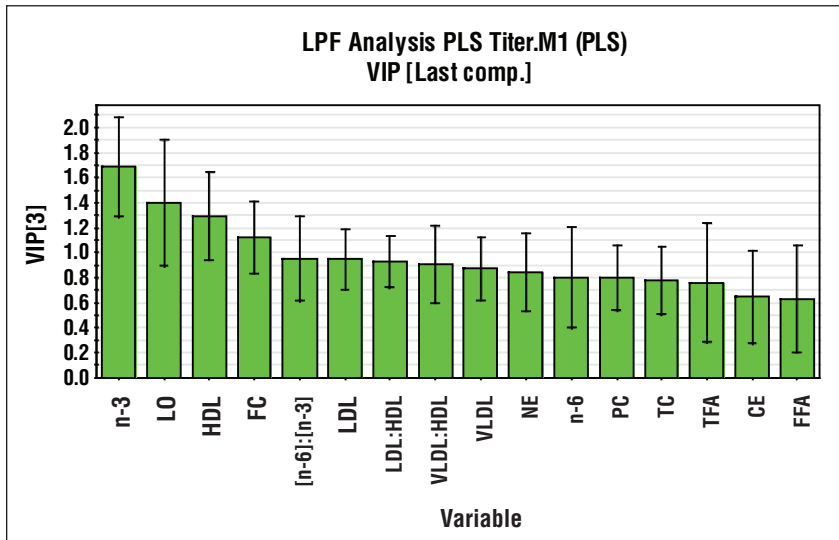


Figure 18. Variable importance plots (partial least squares model of titer)



undergo the rigorous data analysis presented here. We hope to convince the readers that MVDA allows us to extract useful process information through analysis of the readily available data, in order to maximize our understanding of the process. As the biotech industry implements Quality by Design, multivariate analysis will become a necessity. ♦

**REFERENCES:**

1. Kourti T. Process analytical technology and multivariate statistical control. *Process Analytical Technol.* Part 1: 2004;1(1):13–19. Part 2: 2005;2(1):24–28. Part 3: 2006;3(3):18–24.
2. Martin EB, Morris AJ. Enhanced bio-manufacturing through advanced multivariate statistical technologies. *J*

3. Kirdar AO, Conner JS, J. Baclaski J, Rathore AS. Application of multivariate analysis toward biotech processes: case study of a cell-culture unit operation. *Biotechnol Prog.* 2007;23(1):61–67.
4. Cunha CCF, Glassey J, Montague GA, Albert S, Mohan P. An assessment of seed quality and its influence on productivity estimation in an industrial antibiotic fermentation. *Biotech Bioeng.* 2002;78(6):658–669.
5. Ündey C, Ertunc S, Cinar A. Online batch/fed-batch process performance monitoring, quality prediction, and variable-contribution analysis for diagnosis. *Ind Eng Chem Res.* 2003;42(20):4645-4658.
6. Mostafa SS, Gu X. Strategies for improved dCO2 removal in large-scale fed-batch cultures. *Biotechnol Prog.* 2003;19(1):45–51.
7. Zhu MM, Goyal A, Rank DL, Gupta SK, Boom TV, Lee SS. Effect of elevated pCO2 and osmolality on growth of CHO cells and production of antibody-fusion protein B1: a case study. *Biotech Prog.* 2005;21(1):70–77.
8. Hotelling H. *Multivariate quality control, techniques of statistical analysis.* Eisenhart C, Hastay HW, Wallis WA, editors. New York: McGraw-Hill; 1947. p. 111–184.
9. Mason RL, Young JC. *Multivariate statistical process control with industrial applications.* Philadelphia: ASA-SIAM; 2002.
10. Lowry CA, Woodall WH, Champ CW, Rigdon SE. A multivariate exponentially weighted moving average control chart. *Technometrics.* 1992;34(1):46–53.
11. Annamalai A, Lewis J. Statistical process control using multivariate exponentially weighted moving average and MATLAB-to-Excel software interface. *Proc Amer Statistical Assoc, Quality and Productivity Section.* 2006;1776–81.
12. Holmes DS, Mergen AE. Improving the performance of the T<sup>2</sup> control chart. *Qual Engin.* 1993;5(4):619–625.

13. Scholz FW, Tosch TJ. Small sample uni- and multivariate control charts for means. *Proc Amer Statistical Assoc, Quality and Productivity Section.* 1994;17–22.
14. Williams JD, Woodall WH, Birch JB, Sullivan JH. On the distribution of Hotelling's T<sup>2</sup> statistic based on the successive difference covariance matrix estimator. *J Quality Technol.* 2006;38(3):217–229.
15. Danesch U, Weber PC, Sellmeyer A. Differential effects of n-6 and n-3 polyunsaturated fatty acids on cell growth and early gene expression in Swiss 3T3 fibroblasts. *J Cell Physiol.* 1996;168(3):618–624.
16. Priault M, Bessouille JJ, Grelaud-Koq K, Camaugrand N, Manon S. Bax-induced cell death in yeast depends on mitochondrial lipid oxidation. *Eur J Biochem.* 2002;269(22): 5440–5450.