Pharmaceutical Technology®

Maximizing PAT Benefits from Bioprocess Modeling and Control

Nov 1, 2006

By: Michael A. Boudreau, Gregory McMillan, Grant Wilson


PHOTOS.COM

Process analytical technology (PAT), according to the US Food and Drug Administration's Web site (www.fda.gov/cder/OPS/PAT.htm#Introduction), is defined as: a system for designing, analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality. It is important to note that the term analytical in PAT is viewed broadly to include chemical, physical, microbiological, mathematical, and risk analysis conducted in an integrated manner.

Many tools are available for PAT. This article focuses on the knowledge discovery available from the creation and operation of a virtual plant environment. Such an environment can involve the synergistic use of model predictive control, first-principal models, neural networks, and principal component analysis, as described in New Directions in Bioprocess Modeling and Control (1).

Advantages of a virtual plant

The virtual plant is a relatively new concept that should not be confused with existing simulation methods for process design, configuration checkout, and operating training systems. Most existing batch-process simulations are off-line and noninteractive, and most real-time dynamic-process simulations were designed originally for continuous processes. These real-time process simulations can develop severe numerical errors or even fail for the extreme conditions of batch operations. They also may require interfaces for communication of input–output, inventory controls, and coordination with the control system of speedup, slowdown, pause, and resume. The control-system engineer is probably most familiar with tieback simulations because these have been used predominantly for configuration checkout and operating training systems. The process response in these tiebacks is mimicked by the trial-and-error adjustment of ramp rates triggered by the opening and closing of valves or the turning on or off of pumps.

The first key feature that distinguishes a virtual plant from process simulators is its ability to use the actual configuration,

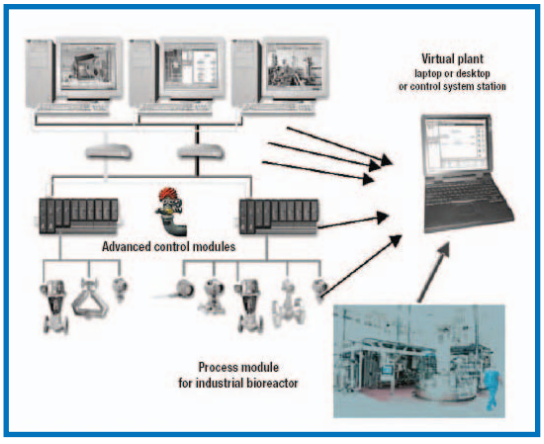

Figure 1: Virtual plant with imported configuration from actual plant and embedded advanced control tools and process simulation.

historian, displays, and advanced-control tool set of the real plant without translation, emulation, special interfaces, or custom modifications. The configuration database from the real plant can be exported and then imported and downloaded into a personal computer or a control-system computer just as if it were an actual hardware controller. Files for operator graphics, process-history charts, and data history from the real plant can be copied to the computer for the virtual plant so that the user has the entire control system of the real plant on a computer (see Figure 1).

Most dynamic, high-fidelity, process-simulation software allows users to build a basic control strategy or sequence inside the simulation environment. Nonetheless, simulation developers tend to have a process rather than a control background and focus. It is unrealistic to expect the process and batch-control capability offered by simulation software to be in the same realm as the control capability of a distributed control system's (DCS) software developed from hundreds of years of experience by process-control experts. The overall control functionality in process simulators is primitive compared with the capabilities offered in the modern DCS, which offers capabilities such as sequential function charts and basic function blocks, a batch manager, and advanced control tools such as multivariable model predictive control. These are almost nonexistent in simulators. Duplicating even a simplified version of a control system in a dynamic simulation is a large effort. At best, a user may end up with two control systems with no assurance of how well they match up and with no way to automatically manage changes between them.

Consequently, most simulation software now offers a standard or custom open-process control interface. For the simulations with material balances to run independently of the DCS for development and testing, however, the simulations still must have internal pressure and level loops set up to prevent volumes from running dry, overflowing, overpressurizing, or developing flow reversals from pressure-gradient reversals that can lead to fatal numerical errors. Tables must be mapped that transfer control from these internal simulation loops to the DCS loops and initialize the proper controlled variables, set points, and manipulated variables. DCS loops that do not have a counterpart in the simulation still must have their controller outputs initialized. Using standard blocks for split-ranged control, velocity limiting, signal characterization, and signal selection makes proper initialization problematic for external simulations.

The virtual plant also can simultaneously stop and start the execution and restore and replay simulated conditions of all control and simulation modules. Emerging is the ability to replay actual plant data-history files at high speeds for adaptation and testing of the process modules without a connection to the actual plant.

Most simulations used for control-system checkout and operator training become obsolete after startup. The investment is lost. The virtual plant offers a better chance of keeping the control system up-to-date simply by importing the most current configuration and enabling the simulation to match process changes better by a nonintrusive automatic adaptation when running the virtual plant in real-time, on-line, and in a read-only mode. An innovative use of model predictive control has been demonstrated to adjust process model parameters (manipulated variables) automatically to provide a better match between key process variables in the actual plant (set points) and in the virtual plant (controlled variables).

In a virtual plant, everything is done in the same configuration environment used for the actual control system. The focus can be more on the application than on learning the inevitable undocumented features and tips and techniques associated with any new simulation software and interface. In summary, a virtual plant offers the following advantages:

- The control system and graphics do not need to be duplicated, emulated, or translated.
- Special data interfaces and tables and initialization issues are avoided.
- All batch, basic, and advanced control tools can be tested readily.
- Controls and simulation can run in unison at the same real-time multiplier.
- Controls and simulation scenarios can be saved, restored, and played back.
- Actual plant data can be played back at high speeds for testing and adaptation.
- Simulations can handle extreme conditions of batch operations and failures.
- Simulations can incorporate dynamics important for tuning and performance.
- Controls and simulation can stay up-to-date and have a longer life cycle.
- Engineers can work in the same environment and focus on the application.

Applications

The most familiar use of a virtual plant is for testing and training. For the checkout of batch sequences and the training of operators, it is important to be able to simulate batch phases repetitively and rapidly. The ability to stop, start, save, restore, and replay scenarios and record operator actions is critical. For first-pass testing and familiarization of sequences and graphics, an automated tieback simulation may be sufficient, but to test and learn the interaction and performance of control strategies and the process, the higher fidelity of dynamics offered by process modules is important. It opens the door for upgrading the process and control skills of the technology, maintenance, and configuration engineers who will support the operations.

Before the configuration starts at the beginning of a project, the process modules can be used to evaluate control strategies and advance control tools. In the past, this was achieved with off-line dynamic simulations. Today, ready access to an industrial tool set for basic and advanced control and simulations adapted to benchtop or pilot-plant runs offers the opportunity for rapid prototyping that can lead to control definitions with better detail and potential performance. Benchtop or pilot-plant systems with a miniature version of the industrial DCS that greatly facilitate the development and scale-up of the control system now

are available (2). Benchtop systems with all the functionality of the main manufacturing systems are scarce, however, because the expertise (and, more important, the interest) to configure, maintain, and engineer these systems is not traditionally in the development groups of these types of companies.

The virtual plant can be demonstrated with industrial-batch, basic, and advanced control systems used in a benchtop or pilot-plant system. The process modules can be adapted by means of a connection to a benchtop or pilot-plant system or by high-speed playback of process data from experimental runs. An opportune time to take advantage of the PAT initiative is during the research and process-development phase. The use of benchtop and pilot-plant systems with an industrial DCS can improve project schedule, cost, and effectiveness because it allows process control to be designed into the plant at an early stage and puts the biochemist, process technology, and configuration engineer on the same page (2). Nonetheless, the best return on investment for PAT is realized by the eventual implementation of advanced process analysis and control in large-scale manufacturing processes.

A virtual plant allows manufacturers to try out and tune innovative strategies such as effective switching of the controller output for set-point response optimization. Advanced control tools such as adaptive control, autotuning, model predictive control (MPC), neural networks (NN), principal component analysis (PCA), and partial least squares or projection to latent structure (PLS) can be demonstrated, adjusted, and evaluated faster than in real time.

Although actual plant operation is the best source of data, the long batch-cycle time and the desire to minimize disruptions from the introduction of perturbations severely restricts the amount of useful plant data for the development of MPC, NN, and PCA. As a general rule, five changes are needed for each process input to develop an experimental model of a process output. For MPC, this corresponds to a minimum of five step changes in each process input, at least one of which is held long enough for the process output to reach a steady condition. For NN and PCA, it means a minimum of five batches per process input in which the respective process input differs from the normal value. Developing a PCA with four inputs for the detection of an abnormal batch requires at least 20 batches with varying inputs and at least five batches with normal inputs. If the batch-cycle time is about two weeks, it would take approximately a year of plant production to have enough data. If the spectrum of variability in the inputs cannot be optimized, then it may take several years of production runs.

It is important to keep in mind that process control transfers variability from a controlled variable to a manipulated variable. When a loop for tight control is tuned, a straight line is generated for the controlled variable, but fluctuations appear in the manipulated variable on a long-term trend. Process control does not make variability disappear. One cannot keep both a controlled variable and manipulated variable constant. In advanced control, variability is transferred from a higher-level process variable (e.g., product-formation rate) to a lower-level process variable (e.g., substrate flow or concentration). The level of control and transfer of variability has a profound effect on the choice of variables for NN and multivariable statistical process control.

The profiles to be analyzed depend upon the level of control. When proportional-integral-derivative (PID) loops are added and put in automatic, the transfer of variability by feedback control means that the batch profile is seen in the controller output rather than in the controller input. For example, if dissolved oxygen is not controlled, then the dissolved oxygen shows the batch profile, and the difference between the measured and reference dissolved-oxygen profile is the PCA input. If a PID loop for dissolved-oxygen control is implemented and put in automatic to manipulate airflow, then the dissolved oxygen draws a relatively straight horizontal line (i.e., the process variable stays relatively close to a constant set point and thus makes a straight horizontal line on the trend plot) and airflow now shows the batch profile and is used for the PCA. If the dissolved-oxygen set point is not constant but is manipulated by MPC to control product formation rate, both the dissolved-oxygen set point and airflow show the batch profile.

Perturbations can be automated and introduced to a virtual plant running faster than real time so that in a couple of weeks there are enough data for the identification of models for MPC, the training of NN, the development of latent variables and discriminant analysis for PCA, and to predict economic variables using PLS. The predictive ability of MPC, NN, PCA, and PLS then can be verified and evaluated by the high-speed playback of previous plant batches.

Conventional PCA assumes that all process inputs, other than those used for the PCA, are fixed. A virtual plant running in real time and synchronized with the actual plant can predict the effect of variations in other process inputs by using the model-based and super

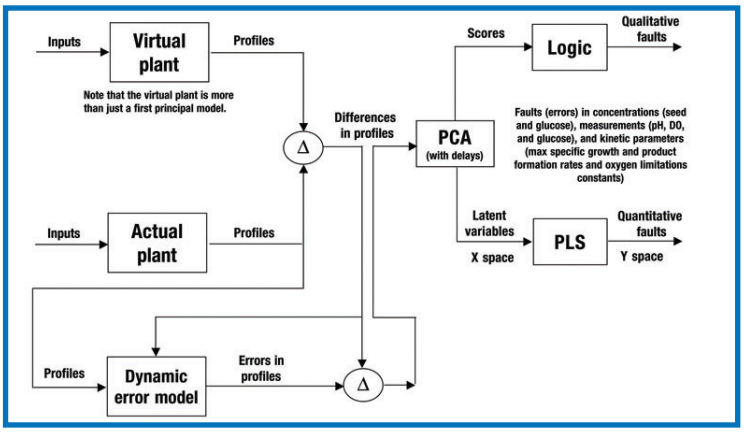

Figure 2: Super model-based dynamic principal component analysis and partial least squares or projection to latent structures for fault detection and prediction. Abbreviations: PCA is principal component analysis and PLS is projected to latent structure.

model-based PCA algorithms. Figure 2 shows the setup of super model-based control (3). In this setup, a dynamic-error model implementing techniques such as dynamic PLS is used to predict the structured errors from an imperfect model. These structured errors are subtracted from the differences in the profiles between the actual plant and the virtual plant before being sent to the dynamic PCA for batch fault detection and PLS for fault prediction.

With a virtual plant, manufacturers can explore more optimal operating conditions and investigate "what-if" scenarios. These scenarios are important to identify the cause of an abnormal batch. To date, PCA for batch fault detection only identifies a batch as abnormal; logic must be added to diagnose the fault. By creating scenarios, a virtual plant can help develop these rule sets off-line faster than real time and can evaluate the rule sets by the high-speed playback of previous batches. An on-line virtual plant synchronized with the real plant can be sped up and run to batch completion to predict and analyze abnormal situations on the basis of current batch conditions. A virtual plant also can help create the predictive capability of the PLS "y space" of economic variables from the PCA "x space" of process variables.

A virtual plant can provide inferential real-time measurements of important process outputs. The built-in material balances, in conjunction with kinetic models, can predict biomass and product concentrations and the slope of batch profiles such as biomass growth rate and product-formation rate. Oxygen-uptake rates can be used to predict metabolic rates. These predictions then are delayed so that the values are synchronized with analysis on-line, at-line, or in a laboratory. The prediction is shifted by a bias that is a fraction of the difference between the inferential and actual measurement in a fashion similar to what is done in the feedback correction of an NN for property estimation (4). The inferred growth, production, or metabolic rates can be controlled variables for optimization by MPC that manipulates substrate and dissolved oxygen or dissolved carbon dioxide concentrations. Figure 3 shows the general setup for the adaptation of the virtual plant process-model parameters by an MPC and the optimization of the batch profiles by another MPC using inferential measurements of growth and production rates from a virtual plant.

The biomass-growth and product-formation rates for mammalian cell cultures are so slow that the signal-to-noise ratio is often too low to allow calculation of these rates from the differences between concentration measurements. An inferential measurement can have significantly less noise and more sensitivity. It is important to remember, however, to correct the inferential measurement by a change in analysis only over a sufficient time interval (e.g., 24 hours) so that the true change in concentration is larger than the scatter in the analysis, as determined by the repeatability of the analyzer.

A virtual plant can be run faster than real time to batch completion, thereby providing an online prediction of key performance indicators such as batch-cycle time and yield based on current batch conditions and inferential measurements. The virtual plant can accelerate the benefits from PAT by offering the ability to use process and endpoint monitoring and control, continuous improvement, and knowledge-management tools in an integrated and accelerated manner.

In summary, the potential applications of a virtual plant are as follows:

- testing of configuration and process interactions;
- process and control education of operators, technicians, and engineers;
- virtual experimentation for exploration of optimums and "what-if" scenarios;
- rapid prototyping of innovative control strategies and advanced controls;
- evaluation of tuning settings;
- identification of MPC models;
- training of NN;
- development of latent variables and reference trajectories for PCA;
- development of logic for fault analysis by PCA;
- online prediction of abnormal situations;
- inferential real-time measurements of important concentrations;
- optimization of batch profiles;
- on-line prediction of key performance indicators for the batch such as cycle time and yield.

Virtual plant implementation

Probably the biggest obstacle to the implementation of a virtual plant is a lack of knowledge of the kinetic equations needed to calculate the growth rates and product-formation rates as a function of batch operating conditions, including temperature, pH, dissolved oxygen, and substrate concentration. In

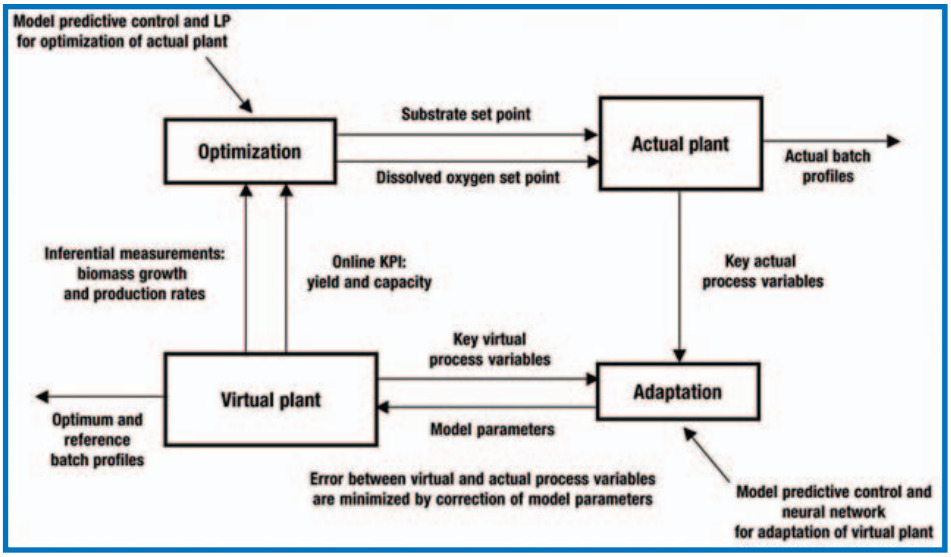

Figure 3: Setup of a virtual plant and use of model predictive control for adaptation and optimization.

some cases, the general form of the equations can be obtained from the literature or from internal research, but benchtop experiments usually are needed to confirm the relationships and quantify the parameters. This requires collaboration between research, process-development, and process-control groups. If the kinetics remain unknown, then an NN may be able to predict rates from inferential and actual measurements. The growth rates and product-formation rates then are included in the net calculation of the rate of change of biomass, nutrient, and product mass in the material balance and integrated to obtain a new accumulation or concentration of the component as described. On-line calculations of oxygen-uptake rate (OUR) and carbon-dioxide evolution rate (CER) as NN inputs enhance the predictive capability of the NN. The result is a hybrid first-principal and NN model (5). In general, batch time should not be an NN input because it creates an undesirable dependence of the predictions on time.

A charge balance is critical for computing the pH that is important for the kinetics. Process simulations in the literature for bioreactors generally use empirical relationships for pH that do not show the effect of alternative operating conditions and upsets. The OURs and CERs can be used in lieu of kinetic equations, but it may be difficult to differentiate between the metabolic rates associated with biomass growth and the rates for product formation.

MPC example

In the exponential-growth phase, the response of biomass or product concentration is a one-sided integrator because it always increases as it ramps up. Because it is undesirable and impossible to decrease the biomass and product in this phase, it is not possible to control these concentrations directly. If the controlled variable is translated from a concentration to a growth or product-formation rate, however, then the one-sided integrating response becomes a two-sided, self-regulating response in which a steady state can be approached from either direction.

Figure 4 shows the automated test of a virtual plant running 500 times faster than real time. The identification algorithm for MPC was able to handle the nonstationary behavior associated with batch processes. Figure 5 shows the self-regulating process responses identified for an embedded 2 × 2 MPC block where the controlled variables are biomass growth rate and product-formation rates and the manipulated variables are substrate concentration and dissolved oxygen. Note that the process gain is much larger for the manipulation of the substrate compared with the manipulation of dissolved oxygen, as would be expected. The MPC seeks to maintain the biomass growth rate and product-formation rate at a set point (see Figure 6). In this case the set point is constant, but it could be a function of batch-phase progression. The penalty on error (a standard tuning parameter in the MPC block) is increased for the product-formation rate to make it more important relative to the biomass growth rate. With this setup, the MPC can provide more consistent and optimized batch profiles for biomass and product concentration that lead to a reduction in batch-cycle time and an increase in product yield.

Figures 7 and 8 show two batches for antibiotic production in a virtual plant 1000 times faster than real time. In the second batch, the embedded MPC defined in Figure 5 was switched to auto when the product-formation rate was reaching its peak. Because the MPC set point was tracking the corresponding process variable, the MPC then sought to hold the product-formation rate at its maximum. Note that in Figure 8 the prediction of the batch-cycle time and final product yield is updated when the product concentration reached about 60% of its normal end point. As a result of the MPC, the second batch has a more constant product-formation rate, shorter cycle time, and

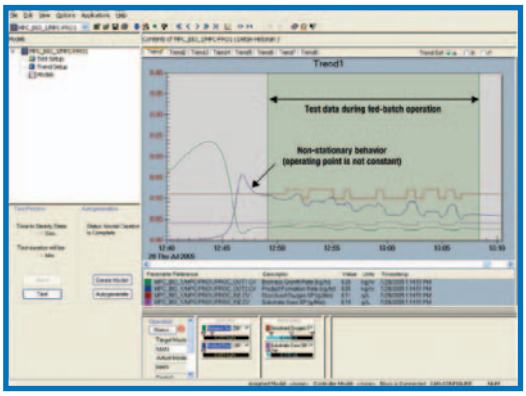

Figure 4: Automated test of a virtual plant.

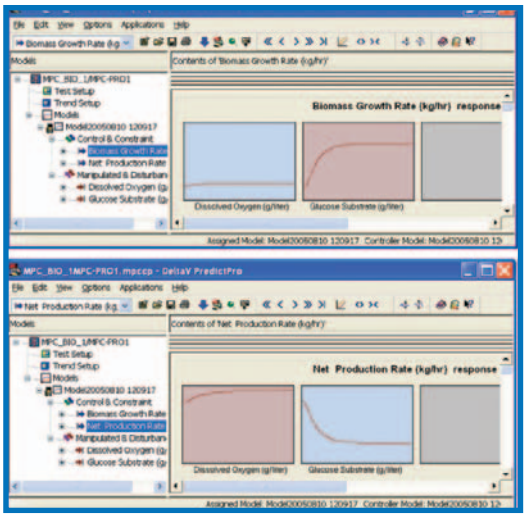

Figure 5: Identified responses for batch-profile control.

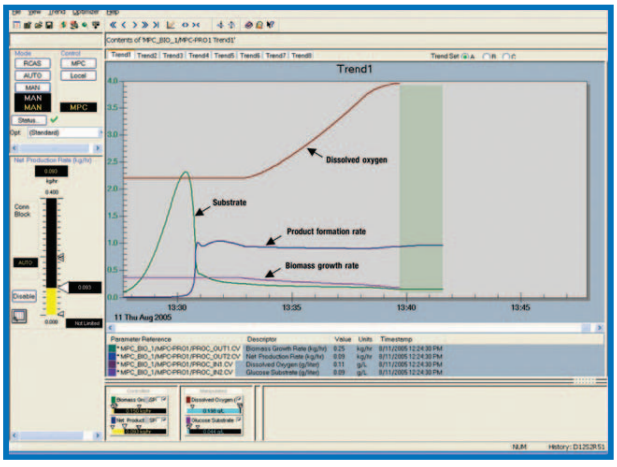

Figure 6: Model predictive control of growth rate and product-formation rate by manipulation of dissolved oxygen and substrate.

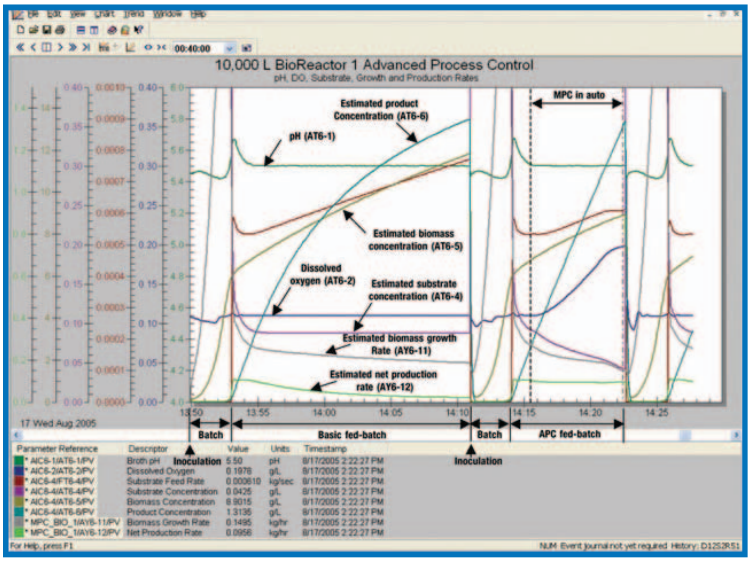

Figure 7: Optimization of batch profiles by model predictive control.

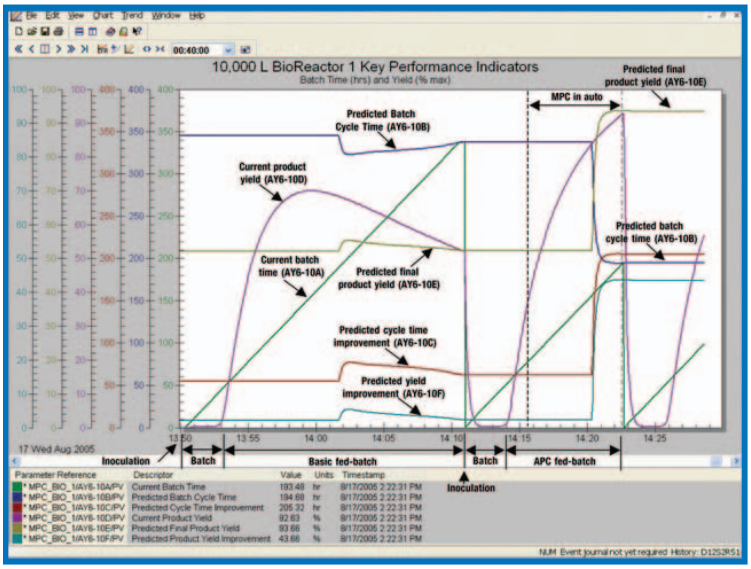

Figure 8: Improvement in key performance parameters by model predictive control.

greater yield. Also, the prediction of batch-cycle time is much more accurate because of the more repeatable batch profile. Alternately, for a fixed batch-cycle time, the prediction of final product concentration is more accurate. If a reliable online substrate-concentration measurement is not available, then the MPC can directly manipulate the substrate feed rate.

Summary

It is advantageous for modeling to be conducted in the process-development and early-commercialization stages to increase process efficiency and enable ongoing opportunities for process control improvement. When benchtop and pilot-plant systems use the same industrial control systems and configuration expertise used in manufacturing, applications of modeling and control can be developed as an integral part of the process definition and ported for industrial production via the control definition. Technologies such as model predictive control, first-principal models, neural networks, and multivariate statistical process control in a virtual plant environment are important for getting the most benefits from process analyzers and tools. The synergistic knowledge discovery is consistent with the intent in the "Process Analyzer and Process Control Tools" sections of the FDA's Guidance for Industry (6).

Michael A. Boudreau is a control systems engineer at Emerson Process Management.

Gregory McMillan* is a consultant in process simulation and control of bioreactors and can be reached at Emerson Process Management, 12301 Research Blvd., Research Park Plaza, Bldg. III, Austin, TX 78759, tel. 512.832.3029, fax 512.418.7500, greg.mcmillan@emersonprocess.com

Grant Wilson is the vice-president of technology at Emerson Process Management's Process Systems and Solutions division.

*To whom all correspondence should be addressed.

Keywords: advanced control, bioprocess, dynamic simulation, first principal, model predictive control, modeling, neural networks, PAT, process efficiency, process control, process analytical technology, virtual plant

References

1. M. Boudreau and G. McMillan, New Directions in Bioprocess Modeling and Control: Maximizing Process Analytical Technology Benefits (ISA, Research Triangle Park, NC, 2006).

2. BioNet Bioreactor Control System, Broadley-James Bulletin L0016B (2003).

3. L. McPherson, J. Morris, and E. Martin, "Super Model-Based Techniques for Batch Performance Monitoring," in European Symposium on Computer Aided Process Engineering (Elsevier Science B.V., Amsterdam, 2002).

4. G. McMillan, "Process Modeling," in A Guide to the Automation Body of Knowledge, V.L. Trevathan, Ed. (ISA, Research Triangle Park, NC, 1st ed., 2005).

5. R.G. Silva et al., "A Hybrid Feedforward and Neural Network Model for the Cephalosporin C Production Process," Brazil J. Chem. Eng. (Dec. 2000).

6. US Food and Drug Administration, Department of Health and Human Services, Guidance for Industry: PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, Sept. 2004.