# Controlling the living factory

## Modeling a mammalian cell culture for an online process control system

JESSICA WHELAN, DANIEL KEOGH – **Pharmaceutical products have been with us for many decades. Relatively new, however, are biopharmaceuticals. These are medical drugs produced using living organisms such as microbial cells, mammalian cell lines and plant cell cultures. Each cell is a living factory transforming nutrients into a protein drug product via their metabolic pathways. The bioreactors in which this production takes place are inordinately complex environments, with a nonlinear, dynamic mix of billions of cells and nutrients, vulnerable to temperature change, pH, inhomogeneity and so on. There is significant interest in improving process control capabilities in bioreactors, one of the main drivers being pressure from the regulatory authorities. The US Food and Drug Administration's Process Analytical Technology (PAT) initiative is an example of this. Utilizing ABB's xPAT product, ABB is collaborating with Irish universities and leading biopharmaceutical players in an initiative funded by Enterprise Ireland to construct models to set up and evaluate the benefits of PAT-enabled, model-based control of a fed-batch mammalian cell culture.**

**Title picture**
Biopharmaceutical production Is somewhat more complicated than that of regular pharmaceuticals. Just how do process engineers keep track of the complex organic processes involved in their manufacturing?

In contrast to pharmaceuticals, which may be thought of as being "chemical" based, biopharmaceuticals are medical drugs produced by (but not simply extracted from) living organisms. The science is relatively new and it was only in 1982 that one of the first products, biosynthetic "human" insulin made via recombinant DNA technology, came on to the market. Since then, the biopharmaceutical sector has become a significant and growing division of the pharmaceutical industry. The number of biopharmaceuticals currently on the market is just in excess of 200 and in 2009 they generated $99 billion in sales. The market is predicted to grow by between seven and 15 percent annually over the next several years, and by 2013 four of the five top-selling drugs will be protein-based products. Mammalian cells, particularly Chinese hamster ovary (CHO) cells, and bacterial systems, such as Escherichia coli (E. coli), produce the bulk of the products on the market but alternative systems such as yeast and plant cells are also used [1].

## Fed-batch mode

The vast majority of current industrial bioprocesses are operated in batch or fed-batch mode. Fed-batch mode refers to a process whereby a concentrated nutrient feed solution is added either continuously or periodically over the course of a batch and no product is harvested before the endpoint is reached. The improvement in productivity achieved in fed-batch processes is due mainly to an increase in the integral of viable cells and a resultant increase in volumetric productivity. Fed-batch mode is popular because of reliability, ease of scale-up, significant increase in production levels and ease of process characterization and validation.

## Controlling the process

In contrast to the chemical and traditional pharmaceutical sectors, process control for bioprocesses is in its infancy due, in part, to the challenges associated with bioreactor control: complex growth media, inadequate measurement of relevant process parameters, the limited and noisy nature of experimental data and difficulties inherent in controlling bioprocesses, which are dynamic, complex and nonlinear. Process control of bioreactors seeks to influence the individual complex intracellular reactions of billions of cells by controlling their extracellular environment [2]. Traditionally, parameters such as temperature, pH and dissolved oxygen (DO) are measured using in-situ probes and are controlled using PID loops to adjust gas or alkali flows. Control of nutrient levels is still usually manual. Generally, bolus feeds are introduced at 24-hour intervals based on offline analysis of daily process samples and a priori process knowledge.

The market is predicted to grow between seven and 15 percent annually over the next several years, and by 2013 four of the five top-selling drugs will be protein-based products.

## Improving control – the PAT initiative

Currently, there is significant interest in improving the process control capabilities of bioreactors. One of the main drivers for this is pressure from the regulatory authorities. In 2003, the US Food and Drug Administration (FDA) launched the Process Analytical Technology (PAT) initiative. The FDA defines PAT as a mechanism to design, analyze, and control pharmaceutical manufacturing processes through the measurement of critical process parameters (CPP), which affect critical quality attributes (CQA). The objective is to understand the processes by defining their CPPs, and to monitor them in a timely manner, preferably in-line or online, thus reducing final testing requirements, reject rates and instances of over-processing while enhancing consistency and product quality.

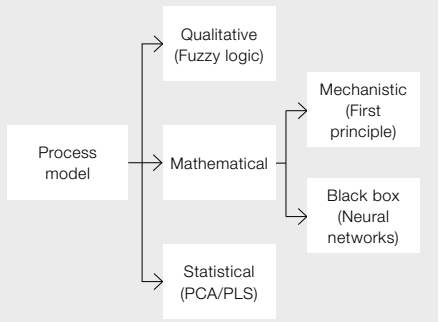
As there is variability both in raw materials and in operation of equipment, a static batch process will produce a variable product. By increasing process understanding and control potential, the PAT initiative aims to design quality into the process, rather than relying on testing the CQAs of the final product, by facilitating a dynamic manufacturing process that can compensate for these underlying variations. PAT has increasingly gained worldwide acceptance as a proven method of ensuring product safety and quality by many industry experts.

Associated benefits of greater process understanding, apart from improved product quality, include faster process optimization and speed to market, improved product titers, decreased process variability, shorter cycle times and reduced waste.

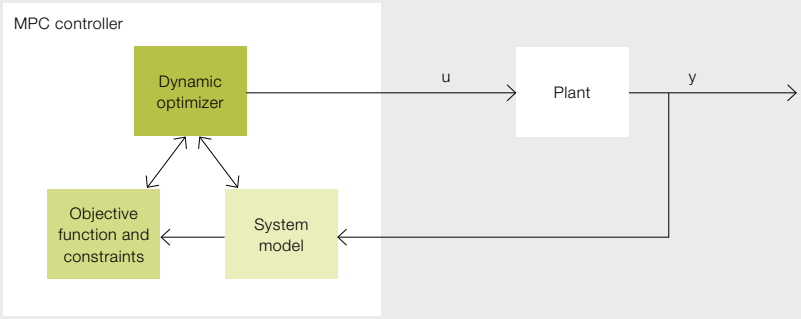
## PAT-facilitating instrumentation

Facilitating the PAT initiative is the increased availability of online measurements. In order to effectively control a parameter, it must first be measured. Spectroscopic techniques such as near infrared (NIR), mid-IR and Raman are online monitoring tools for nutrient and metabolite concentrations. Automated multifunction analyzers such as the Nova or YSI systems use a combination of enzymatic, amperometric, potentiometric and coulter counter or CCD camera analyzers to perform simultaneous quantification of

1 Process model categories




2 Model predictive control schematic




nutrient and metabolite concentrations, cell density and viability, dissolved oxygen and carbon dioxide, and pH at-line.

Automated at-line analysis must also be able to withdraw an automated, sterile sample from the bioreactor. There are a number of systems capable of delivering this requirement currently being developed and brought to market. Closed-loop feedback control based on at-line samples is possible if the sample interval is less than 25 percent of the dominant system response time [3].

## Model types and control strategies

There is a wide range of potential applications for models of the mammalian cell culture process: investigation of underlying process mechanisms; analysis and prediction of experimental results facilitating swifter process optimization; advanced control strategies; decision support systems; and soft sensors. Choice of model type depends on both the intended application and also the quantity, quality and nature of experimental data available upon which to build the model.

Process models form the basis for most of the modern or advanced control strategies. The discussion here will focus on model-based control for mammalian cell processes. → 1 illustrates a nonexhaustive categorization of process models. They can be delineated into qualitative, mathematical and statistical models. Mathematical models can be further subdivided into mechanistic and black box models. For control applications, generally, quantitative models are required. As such, the most useful model types for the control of bioprocesses are first-principle models, neural-network-based models and PCA- or PLS-based multivariate statistical models.

## First-principle models

The first-principle models discussed here refer to engineering, rather than biological, first principles. They combine mass balance, rate and yield equations to describe the dynamic profiles of biomass, nutrients and metabolites.

When these models move from treating the cell as a single unit to more detailed levels, considering for example, amino acids, nucleotides, proteins and lipid pools, the complexity, time for development and required computational power increase greatly.

To replicate the relationship between substrate and metabolite concentrations, cell growth and product formation, it is not necessary to use expensive theoretical models. Instead, parameters within the series of equations can be optimized in order to fit the equations to experimental data.

Typically, these semi-empirical models have some extrapolation capability and so first-principle models of this nature are most useful for optimization and control applications. Some experimental data is required but it is usually of the order of three to five batches depending on the number of parameters and range of operating conditions within the model.

## Neural networks

Neural networks involve a type of black box modeling, where little or no understanding of the underlying mechanisms of the process is required. The neural network is programmed to predict an

Mammalian cells, particularly Chinese hamster ovary (CHO) cells, and bacterial systems such as E. coli, make up the bulk of the products on the market.

3 The Applikon 15 L pilot scale bioreactor with ancillary equipment such as feed pumps and gas delivery equipment

The PAT initiative aims to design quality into the process. Process models form the basis for most of the modern or advanced control strategies.

output or outputs based on a set of inputs by training it with a set of experimental data and/or deterministic models. The neural network learns by processing data without rules. At its core it is a nonlinear regression. A series of data inputs are applied to a number of functions or nodes, which form a layer within the neural network. The inputs are weighted, summed and then processed by a transfer or threshold function before being output. This output may be the final result or simply data to be passed on to further layers within the neural network. During training, the weights applied to each of the inputs are adjusted in order to minimize the difference between the predicted and actual results.

The development of a robust neural network model requires a much greater number of batches than first-principle models, typically of the order of 20 to 30.

The quality of the experimental data is also very important. It is necessary to identify all inputs that affect the output to be predicted, and also to avoid correlated inputs. The training data set must contain enough variability to span the entire operating region and the data must be preprocessed effectively to remove outliers and poor-quality data as well as to scale and normalize the inputs. However, in the case of bioprocesses where there is a high degree of process complexity, the low level of process understanding needed can be a considerable advantage, particularly if there is access to historical datasets.

## Multivariate process control

Traditional manufacturing industries have used univariate statistical process control (SPC) charts to monitor and improve their processes for some time. In bioprocesses, many variables are recorded and can potentially have an effect on the output of the system. Multivariate statistical methods have been developed to allow similar analysis of these large datasets. Detection of abnormal conditions and root cause analysis can be conducted by applying Principal Component Analysis (PCA) and Partial Least Squares or Projection on Latent Structures (PLS) methods.

Generally speaking, PCA is most often used to detect batch abnormalities, while PLS can be used to predict output parameters such as product quality or batch endpoint based on input variables and the current process states.

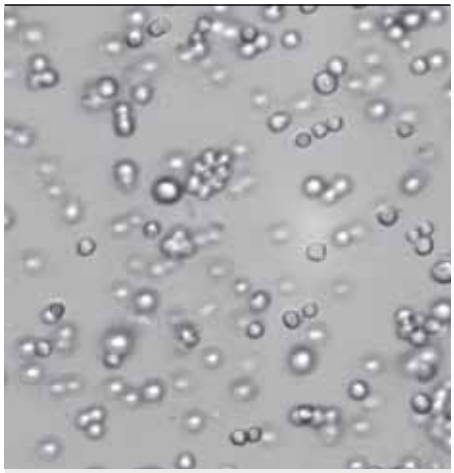
Statistical models are data driven and, therefore, it is important that the set of good reference batches captures the variation in the process. Too little variation results in false alarms and too much variation causes the model to be insensitive. A set of ten or more batches is necessary to build a multivariate statistical model.

Once the effort to build a good process model has been expended, it is desirable to exploit that model to maximize the benefit. Model-based control strategies represent one potential application. There are numerous forms of model-based control including model predictive control (MPC).

## Model predictive control

MPC → 2 is a multiple-input multiple-output (MIMO) control algorithm based on the repeated solution of a finite-horizon optimal control problem subject to a per-

4 CHO cell lines



Chinese hamster ovary cell lines are the workhorses of the mammalian cell biopharmaceutical industry. They are used to manufacture a number of licensed therapeutic proteins such as erythropoietin (EPO), CD20, tumor necrosis factor alpha and HER2. They are robust cells, which can easily be adapted to meet the requirements of large-scale protein production.

They are easily adapted to grow in suspension at high viable cell densities, which simplifies large-scale culture in stirred tank bioreactors. They are also capable of high levels of protein expression. Also, their DNA is easily modified in order to have the cell line produce the protein product of interest.

formance specification, constraints on states and inputs, and a system model. It can use a mathematical model such as first principle or neural network models, or statistical models such as PCA or PLS to create a future trajectory of the batch based on multiple measured process inputs. It seeks to minimize the square of the error between the predicted trajectory and desired trajectory over a user-defined prediction horizon and then calculates a controller action for each of its outputs. In contrast to a traditional PID controller, which aims to minimize the instantaneous error between process variable and setpoint, the longer view taken by MPC reduces the impact of unknown disturbances, erratic signals and noise. MPC can also deal well with systems with a long dead time, though it is not robust in situations where the dead time changes significantly.

Statistical models can be used as an aid for online process evaluation and decision making, as well as a tool for identifying variables likely to be responsible for deviations.

MPC is inherently suited to optimization. The presence of a dynamic optimizer, objective function and constraints within the framework means that MPC can predict future violations of constraints, handle complex interactions and smoothly adjust the manipulated variables. MPC has the widest application of all advanced control strategies in industrial applications.

### The apPAT project

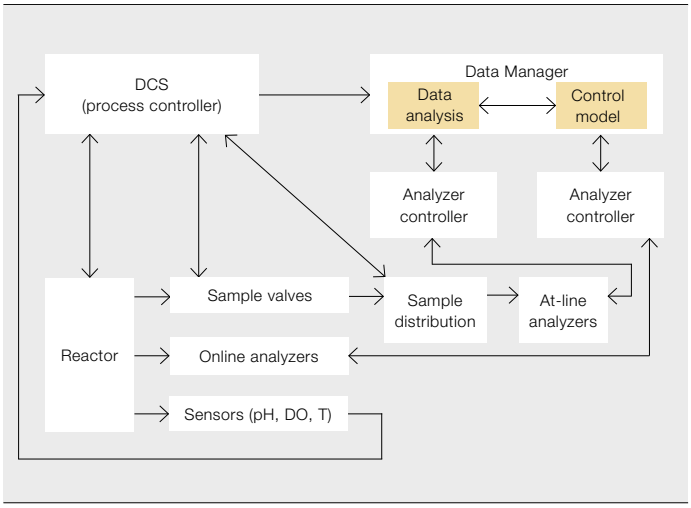
Currently, a project to investigate the potential of the bio-application of PAT is underway. Funded by Enterprise Ireland and led by Professor Brian Glennon in the School of Chemical and Bioprocess Engineering at University College Dublin, research groups at UCD, Dublin City University, the Tyndall Institute at University College Cork and the National Institute for Bioprocessing Research and Training (NIBRT) are collaborating with ABB to set up and evaluate the benefits of PAT-enabled, model-based control of a fed-batch mammalian cell culture. A number of Irish-based multinationals such as Pfizer, Eli Lilly, Jannsen Biologics and Merck as well as a group of indigenous SMEs including BioUETIKON, Technopath and BioImages, among others, form an industrial advisory board that meets quarterly to comment on and guide the research strands.

The project is based on an ABB Exhended Automation System 800xA control system and an xPAT system comprised of FTSW800 analyzer controllers and data management system.

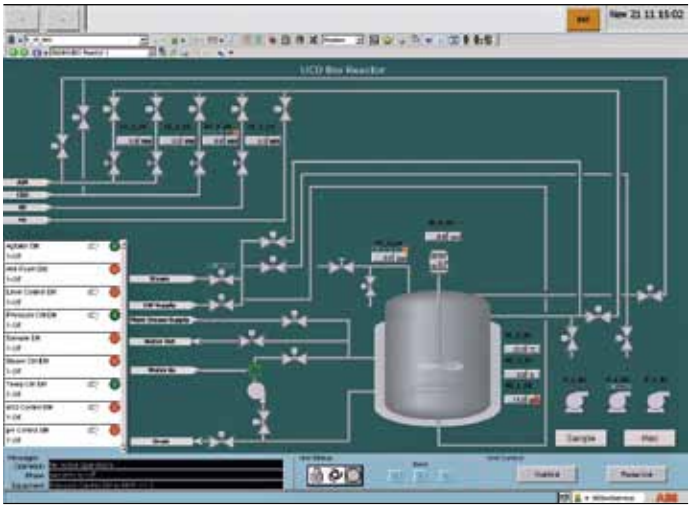
An industry-specific application built on the System 800xA infrastructure, xPAT (Industrial IT eXtended PAT) is a next-generation PAT solution that harnesses the System 800xA operations and engi-

The longer view taken by MPC reduces the impact of unknown disturbances, erratic signals and noise. MPC can also deal well with systems with a large dead time.

5 Block diagram of setup




6 apPAT control graphical interface



neering environment and integration capability to provide significant improvements in the overall process and end-product quality. It provides life sciences users with a single system to access and examine online, real-time process data directly from the manufacturing operation. The configurable Windows-based system collects data from ABB and/or third-party vendors' analytical instruments and analyzes the data to determine the actual condition of the process. It then passes the resulting information to the ABB or third-party control system, and to other applications that support the drug manufacturing process.

ABB is collaborating with Irish universities and leading biopharmaceutical players to construct models for PAT-enabled, model-based control of a fed-batch mammalian cell culture.

ABB is providing the engineering services required to install and configure the system at UCD → 3. A CHO cell line → 4 is being used as the model system because it is the most common industrial expression system. A number of PAT technologies both in-line and at-line – such as mid-IR and Raman spectroscopies for the monitoring of substrate and metabolite concentrations and flow cytometery, Canty imaging systems and the Beckman Coulter Vi-Cell for the determination of many cell parameters such as cell density, viability and cell cycle – have been evaluated and developed for the process under investigation.

Researchers at the Tyndall Institute and at UCD are developing a sample valve assembly that is capable of taking an automated, sterile sample from the reactor to facilitate at-line analysis.

A first-principle, semi-empirical model describing the biomass, substrate and metabolite trajectories is currently being utilized in an MPC framework in order to control the feed rate of nutrients to the reactor.

Experimental work is ongoing to implement, optimize and evaluate the practicalities and benefits of installing a PAT-enabled advanced control system for a mammalian cell process → 5–6. Work to date has shown a 15 percent increase in the maximum viable cell density achieved in the MPC-controlled fed-batch bioreactor when compared with standard bolus fed-batch bioreactor runs, a significant increase. An increase in viable cell density means that there are a greater number of production units within the bioreactor and so more of the biopharmaceutical is produced.

A control system such as the xPAT system, capable of integrating a wide variety of PAT instruments and managing the data that they produce, opens up the possibility of implementing advanced model-based control strategies. These advanced strategies can potentially increase productivity and process robustness, as well as decrease process variation, all extremely desirable outcomes in the biopharmaceutical industry.

**Jessica Whelan**
School of Chemical and Bioprocess Engineering,
University College Dublin
Dublin, Ireland
jessica.whelan@ucd.ie

**Daniel Keogh**
ABB Control Technologies
Dublin, Ireland
daniel.keogh@ie.abb.com

**References**

[1] Walsh, G. (2010). Biopharmaceutical benchmarks 2010. Nat Biotechnol 28 (9):917–924.
[2] Boudreau, M., McMillan, G. (2006). New directions in bioprocess modeling and control – Maximizing process analytical technology Benefits. ISA.
[3] Boudreau, M., Benton, T. (2008). Bioprocess control: What the next 15 years will bring. Retrieved February 9, 2012 from http://www.pharmamanufacturing.com/articles/2008/062.html