

Chapter 1

Opportunities

1-1. Introduction

The cell is the ultimate reactor. It is the essential and in some cases the only possible route to the production of complex compounds [1]. Bioreactors (fermenters) are the key unit operation in biopharmaceutical, brewing, biochemical, biofuel, and activated sludge processes. Each bioreactor relies on the performance of billions of these individual reactors or cells. Process control attempts to influence the sophisticated metabolic reactions inside the cell by controlling the environment immediately outside of the cell [2].

Process control attempts to influence the individual sophisticated internal reactions of billions of cells by controlling their extracellular environment.

In order to control a process variable, we must measure it directly or infer it from other measurements. The key process variables of the cells' environment that are measured are compositions and conditions, such as temperature, pH, dissolved oxygen and carbon dioxide, and substrate. In process control, the compositions and conditions in the broth or vent gas are termed *process outputs* and the quantity, composition, and condition of feeds or charges into the bioreactor or of coolant into coils or jacket are termed *process inputs*. Seed cultures, nutrients, substrates, air, and oxygen are process inputs. In basic feedback control, process outputs that are measured or inferred are controlled by manipulating the process inputs. The goal of process control is to transfer variability from important process outputs to the process inputs designed to be manipulated. How well the process outputs are controlled is determined by loop dead time, measurement resolution, repeatability, noise, the tuning of the controllers, and the resolution of the control valves and variable speed drives [3].

Section 1-2 of this chapter discusses the major sources of variability, the definition of process inputs and outputs, the availability of measurements, the differences between experimental and first-principle models, the quality and quantity of data, and the different modes of batch operation. Section 1-3 discusses the different levels of control and the effect each has on the selection of model inputs and on the setup of basic feedback loops for different types of cell cultures. Section 1-4 introduces the important topics of online yield and capacity performance indicators. Section 1-5 outlines the use of model predictive control for optimization. Section 1-6

summarizes the reasons behind, the drivers of, and the tools used for the Process Analytical Technology (PAT) initiative.

Models offer benefits before they are put on line. It is the authors' experience that significant improvements result from the process knowledge and insight that are gained when building the experimental and first-principle models for process monitoring and control. The benefits often come in ways not directly attributable to the associated technologies, such as changes in batch set points, end points, and phases.

Significant improvements result from the process knowledge and insight that are gained when building the experimental and first-principle models for process monitoring and control.

Doing modeling in the process development and early commercialization stages is advantageous because it increases process efficiency and provides ongoing opportunities for improving process control. When bench-top and pilot plant systems use the same industrial control systems and configuration expertise that are employed 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. The advanced technologies discussed in this book—model predictive control (chapter 4), the virtual plant (chapter 5), first-principle models (chapter 6), neural networks (chapter 7), and multivariate statistical process control (chapter 8)—are important tools for maximizing the benefits from process analyzers and tools. The synergistic discovery of knowledge is consistent with the intent behind the Process Analyzer and Process Control Tools sections of the FDA's "Guidance for Industry PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance" document, as the following excerpts show.

The synergistic knowledge discovered through the integration of analyzers, models, and controls is the essence of the process analytical technology (PAT) opportunity.

In the Process Analyzer section of the FDA's guidance document we read:

"For certain applications, sensor-based measurements can provide a useful process signature that may be related to the underlying process steps or transformations. Based on the level of process understanding these signatures may also be useful for the process monitoring, control, and end point determination when these patterns or signatures relate to product and process quality."

In the Process Control Tools section of the guidance document we read:

“Strategies should accommodate the attributes of input materials, the ability and reliability of process analyzers to measure critical attributes, and the achievement of process end points to ensure consistent quality of the output materials and the final product. Design and optimization of drug formulations and manufacturing processes within the PAT framework can include the following steps (the sequence of steps can vary):

- Identify and measure critical material and process attributes relating to product quality
- Design a process measurement system that allows real-time or near real-time (e.g., on-, in-, or at-line) monitoring of all critical attributes
- Design process controls that provide adjustments to ensure all critical attributes are controlled
- Develop mathematical relationships between product quality attributes and measurements of critical material and process attributes

Within the PAT framework, a process end point is not a fixed time; rather it is the achievement of the desired material attribute.... Quality decisions should be based on process understanding and the prediction and control of relevant process/product attributes.”

The discovery and implementation of more optimal batch profiles, end points, and cycle times are encouraged by the PAT guidance document. The proper integration of technologies can actually speed up rather than slow down the time to market. Ideally, the optimization process is coincident with the commercialization process. Waiting until a pharmaceutical is out of patent protection increases the risk of not having the lowest cost position. The return on investment (ROI) gained by optimizing the production process for a new drug can be enough to justify the use of analyzers, models, and control. The goal of this book is to provide the basis for taking advantage of this opportunity.

Making the optimization process coincident with the commercialization process can lead to a low-cost position, an important factor for a competitive advantage when a drug goes off of patent protection.

Learning Objectives

- A. Be able to track down the major sources of variability.
- B. Recognize how instruments aggravate or mitigate variability.
- C. Understand how to reduce variability through different levels of control.
- D. Know the basic setup of loops for different types of bioreactors.
- E. Appreciate the implications of missing online and infrequent lab measurements.
- F. Become familiar with the requirements and capabilities of various models.
- G. Understand how the level of control affects the selection of model inputs.
- H. Discover optimization opportunities.
- I. Be aware of the business drivers for the Process Analytical Technology (PAT) Initiative.

1-2. Analysis of Variability

In the pulp and paper industry, poor feedback control actually increases the variability of important process outputs. This phenomenon is rare for bioreactors because of the large volume, the slowness of the kinetics, and the lack of interactions. However, the actual performance the bioreactors achieve still depends on how well controllers are set up and tuned to deal with the sources of variability. Most feedback control implementation problems involving bioreactors can be traced to nonrepresentative measurements or mechanical design limitations in equipment, agitation, piping, and injection.

The bulk velocity in a bioreactor for animal cell cultures is extremely low and ranges from 0.01 to 0.1 ft/sec (scale of agitation ranges from 0.1 to 1.0, and power input per unit volume ranges from 10 to 100 watts/cubic meter). For bioreactors with fungi and bacterial cultures, the degree of agitation is comparable to chemical reactors with gas dispersion: bulk velocity is 0.8 to 1.0 (ft/sec) or larger (scale of agitation ranges from 8 to 10 and power-per-unit volume is 1000 watts/cubic meter or more) [4] [5] [6]. For ethanol, the scale of agitation ranges from 1 to 2 in the bioreactor and 2 to 4 in the mix tanks.

Low fluid velocities result in larger mixing time delays, less broth uniformity and gas dispersion, and lower mass transfer rates. The response time of a clean electrode and thermowell is several times larger in an animal cell culture. An even more important consideration is the increased propensity for coatings caused by low fluid velocity. A 1 millimeter coating on a pH measurement electrode can cause its response time to increase from 10 seconds to 7 minutes [7]. Chapter 2 on process dynamics discusses how these time delays and time lags add up to a total loop dead time that determines the loop's ultimate performance.

A coating on the reference electrode junction can cause a drift of 0.5 pH or more during the batch. Since the peak in the growth rate of the biomass with pH is relatively sharp, the pH drift can affect batch performance. Even if electrodes could be withdrawn, cleaned, and calibrated during a batch, the result would be disruptive because the time required for equilibration of the reference junction in the broth is significant [7]. Thus, even if the optimum pH does not change during the batch, there is an opportunity to counteract a drift in measurement by slowly trimming the pH set point. In continuous processes, it is quite common for operations to home in on a more optimum set point so as to compensate for sensor offset. However, in batch processes, the practice is to use fixed set points.

The low fluid velocities in animal cell bioreactors increase the total loop dead time and the need for optimizing the pH set point.

Load cells or mass inventories are frequently used as a standard way of adding some materials. Many believe that load cells and mass inventories are accurate standard technology, but the reality is that they are often the source of significant variability. To use load cells you must be sure that all piping and other connections to the vessel offer no vertical forces on the vessel (this includes operators who lean on the vessel when looking through the tank site glass). This is often difficult to achieve in reactors, since many steam sterilization and air/nutrient feed lines are connected to them. Also, changing ambient temperature conditions causes metal fittings and/or support structure to expand and contract and temperatures in the jacket cooling medium to change, which then changes the medium's density, hence weight, and so on. These changes can potentially change a load cell measurement unless the load cell is carefully engineered not to be affected or the change is adjusted for by correcting the model [8].

Poor control loop performance in bioreactors can often be traced back to nonrepresentative measurements or mechanical design limitations.

Totalized Coriolis mass flowmeters are used instead of load cells to achieve higher accuracy in charge measurements. These meters also provide an extremely precise density measurement for verifying the composition of charges. In the case of ethanol production, the density measurement on the feed to the fermenter provides an inference of the percentage of starch and sugar solids. Adding both a differential pressure measurement across the meter for laminar flow and a neural network produces a viscosity measurement that is an inference of starch-to-sugar conversion from the liquefaction process. Accurate mass flow measurements also facilitate an accurate mass flow ratio of feed to recycle flow at various points in the ethanol process.

A higher level of control, termed *optimization*, occurs when cell growth and product formation rates, which are indicative of intracellular reaction rates, are inferred from oxygen uptake rates (OUR) and carbon dioxide evolution rates (CER). Mass spectrometers can be used to provide online indications of OUR and CER by way of measuring the oxygen and carbon dioxide in the off-gas [1] [10]. Alternatively, the air or oxygen feed can be momentarily shut off and the rate of change in the dissolved oxygen used to infer the OUR. For ethanol production, the simple addition of a pressure measurement at the bottom of the fermenter enables us to infer CER from the broth's loss in weight. These rates can be potentially made more consistent and maximized at various points in the batch by adjusting the set point for key process outputs, such as substrate concentration, dissolved gas concentration, and pH [8] [9]. In most cases, substrate concentration is not measured, and substrate feed (a process input) is set directly.

Most of the variability in the product or effluent can be tracked down to unknown variability in the process inputs. Measurable changes in seed cells are reduced by not transferring the seed tank until appropriate criteria are met. The process inputs that vary the most and with the greatest consequences are the unknown differences in the internal composition and metabolic pathways of cells. It is important to remember that a cell is the most sophisticated reactor known, and genetic engineering is still in its early stages. Fortunately, a virtual plant can achieve its objectives using relatively simple expressions for the kinetics and mass transfer rates, with the parameters fit to experimental data and combined with equations for the material, charge, and energy balances. We describe these procedures in chapter 6 on first-principle models.

The causes of variability in the seed cells are analogous to the causes of variability in people: the current and historical environment of the culture and genetics (heredity) [8]. In the case of environment, we have to consider the entire environment that the cell culture has been exposed to since the cell culture began as a few frozen cells in a liquid nitrogen

storage container. These cells must be carefully thawed in a controlled fashion, and then go through several growth stages in different progressively larger equipment sets, typically lasting many days. During these stages the cells must be given all the appropriate nutrients and be controlled at the appropriate temperature, pH, and the like—and all of this performed in a sterile environment.

These “early” steps are typically not as well controlled (i.e., they require more manual operations) as the fermenter itself because more manual operations are entailed. So the opportunity for excursions, operator mistakes, that is, for variability, is significant [8].

There is a low, but finite, probability that a cell will mutate while dividing to become two cells. With each succeeding generation of cells, the probability increases that one or more of the cells has mutated. Note that by the time a culture is making product in a fermenter, perhaps thirty or more generations have occurred since the start of a few frozen cells. A mutated cell, theoretically, might produce the same or more product than a standard cell, but, in most cases, it produces less (or perhaps none). Companies specifically try to develop industrial processes so as to limit the number of cell generations because this minimizes concerns regarding cell mutations. The Food and Drug Administration (FDA) sometimes checks on this during its reviews. Progressive mutation is one of the reasons why bioprocesses are not typically run continuously (i.e., for very long periods of time) to maximize the probability that the culture will remain in a relatively pure state [8].

This is not to say that other bioreactor inputs cannot cause significant variability. The use of complex raw materials as nutrient sources, such as soybean meal or fish meal, is one of the most obvious significant sources of variability. This is particularly the case when a process shifts from one ocean fish harvest or soybean harvest to another. However, these kinds of changes cause sudden shifts in the productivity of all bioreactors making a product. The changes in nutrients align with any new lot of raw material being used rather than varying in a random lot-to-lot fashion, which might be associated with instrument noise or drift [8].

Process Inputs and Outputs

Figure 1-2a shows many of the process inputs and outputs that are possible for a bioreactor. Process inputs are typically quantities of reagents, substrates, nutrients, gases, enzymes, cells, or surfactants added to the vessel and cooling water circulated through the jacket. The flows of these process inputs are set by a final element, such as a pump or fan speed. Another type of process input is agitation set by motor speed. There are also unknown and often undesirable inputs such as impurities and contaminants. The process inputs determine the process outputs, such

as broth and jacket temperature, pressure, pH, dissolved gases, broth composition, and off-gas composition and flow [1] [10].

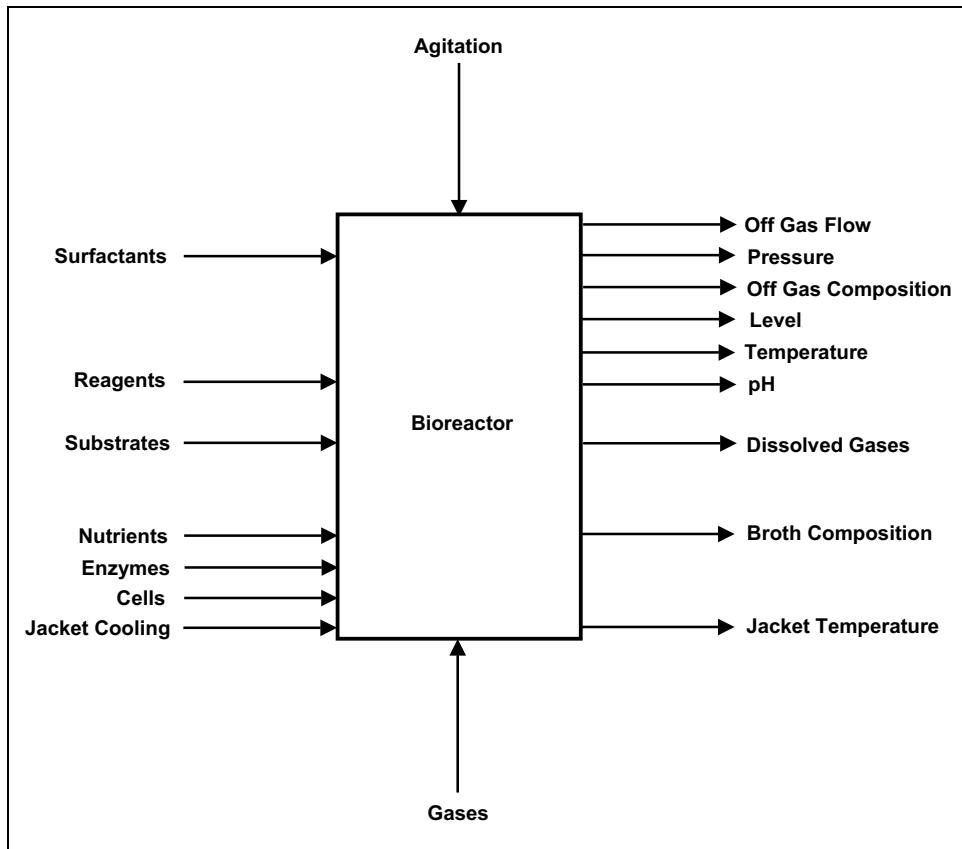


Figure 1-2a. Bioreactor Process Inputs and Outputs

Field and Laboratory Measurements

Even though the process outputs are indications of bioreactor performance, only the simpler measurements, such as temperature, pressure, pH, and differential pressure are typically installed on bioreactors. This is particularly true in the brewing and biofuel industries. The next most common measurement is dissolved oxygen for aerobic processes. Dissolved carbon dioxide electrodes have significantly improved since their appearance in the 1980s and are increasingly being used. In particular, their use is increasing for animal cell cultures when the pH is controlled by adding carbon dioxide, for the aeration and stripping of carbon dioxide in the broth by an air sparge, or for the sweeping of carbon dioxide from the head space by an inert gas [1]. Companies seeking higher levels of control add "at line" measurements of both components, such as glucose probes for substrate concentration and optical density instruments for biomass concentration.

Pharmaceutical manufacturers take advantage of sophisticated types of off-gas analyses on line. These analyses are generated by sending gas sample streams through field mass spectrometers that have a much better resolution, repeatability, reliability, and response time than other analyzers. Microprocessors and software have made possible diagnostics, computation capability, and user interfaces that greatly improve the mass spec's functionality and maintainability [1] [10].

Bioreactors do not have enough field measurements of process outputs.

The most important process output, broth composition, is measured in the laboratory. When this measurement is done in a laboratory, the time delay between the time when the sample was taken and when the analysis result is available to the control system is large and inconsistent. So too is the time interval between data points. The infrequency and variability in the timing of laboratory analysis makes it difficult to correlate process outputs with process inputs and to construct experimental models, such as partial least squares (PLS) and artificial neural network (ANN) models. This infrequency and timing variability also makes it nearly impossible to use the results for basic control directly. Inferential measurements (virtual sensors), which are discussed here and in chapter 5 on the virtual plant can provide rapid as well as more reliable measurements with less noise in broth composition, such as sugar and biomass concentration.

The laboratory measurement data of the most important process output, broth composition, is so sparse and delayed that conventional PLS and ANN modeling is difficult, making it unlikely that the lab data can be used directly for basic process control.

Experimental Models

Process models can be used to prevent abnormal situations, to detect bad batches, to experiment (exploring "what-if scenarios"), to prototype control systems, to check out configurations, to train personnel in operations, to predict process performance (batch end point, yield, and/or cycle time), and to optimize in real time.

Experimental models correlate a series of process outputs to process inputs. The three major types of experimental models used in industry (model predictive control [MPC], ANN, and PLS) are discussed in detail in chapters 4, 7, and 8, respectively. Some important fundamental conceptual differences distinguish these three models.

Artificial neural network (ANN) models predict a process output as a nonlinear function of process inputs with adjustable time delays. Partial

least squares (PLS) models predict a process output from a linear combination of a reduced set of independent variables called *latent variables* or *principle components*, which are themselves a linear combination of process inputs with adjustable time delays. The principal component analysis (PCA) eliminates correlations between process inputs. Both ANN and PLS are primarily steady-state methods reformulated for batch processes. Normally, the ANN and PLS process inputs and outputs are not incremental and do not have a time constant or process gain [11]. Model predictive control (MPC) provides a future trajectory (time response) for the change in a process output based on the linear combination of the effects of past changes in process inputs. The trajectory includes the effect of a steady state gain and time constant for self-regulating (exponential) responses or an integrating gain for integrating (ramping) responses, besides a time delay. Model identification software can be used to develop online dynamic estimators (ODE) with future trajectories for use in MPC [11].

ANN models predict nonlinear behavior, PLS and PCA models eliminate correlations between inputs, and MPC and ODE models provide the exponential and ramping time response for deviations about operating points.

Quality and Quantity of Data

To be capable of identifying a good quality model, the important inputs must have known changes that are at least five times larger than the noise associated with the input. Also, at least five changes in each input are needed. In other areas of process control, sufficient quantity and quality of data are ensured by conducting step or bump tests and pseudo random binary sequences (PRBS), in the case of MPC, or by designing experiments, in the case of PLS or ANN. It is rare that historical data can be completely relied on to generate these experimental models [3] [11].

To build good experimental models, there must be at least five changes in each important input that are each at least five times larger than the noise band.

Since each bioreactor batch may be worth hundreds of thousands to millions of dollars (for high value-added products), opportunities for introducing perturbations are severely restricted. Since each batch takes days to weeks to complete, even if perturbations could be made within the specification limits set by the plant's procedures for validating and managing change procedures, it might take months to years before enough data points were available to develop a MPC, PLS, or ANN model.

It takes months to years to get enough batches with identifiable changes in process inputs that are large enough to develop a MPC, PLS, or ANN model.

Processes that are subject to federal cGMPs, such as those that make medicines, are required to define all critical process parameters (parameters that could affect product quality) and the corresponding “proven acceptable ranges.” For fermentations that make pharmaceuticals, parameters such as pH and temperature are typical critical process parameters. In these cases, it would not be permitted to deliberately perturb the pH or temperature outside the “proven acceptable range” (which is a fairly narrow range of values for many fermentations) just for the sake of obtaining model information. This is because such perturbations would be defined as a process deviation and could cause that lot to be rejected for the purposes of making product for market [2] [8].

Golden Batch

In the early applications of PCA to batch processes, a *golden batch* for multivariate statistical process control (MSPC) was defined to be the median of the batch profiles for a fixed set of inputs. This definition is misleading because the definition is not really representative of the best possible theoretical or practical batch and it assumes fixed process inputs that may not be achieved in plant operation. The best batch profile needs to be found by theory and experimentation for any process input in order to meet the control objectives. Chapter 8 on MSPC discusses the relative merits of using multiway and model-based PCA to generate reference batch profiles.

The “golden batch” is not the median of the batch profiles but is the best batch profile achievable based on theory and experimentation for a variable set of inputs.

Dynamic First-Principle Models

An opportunity that is now emerging is the use of dynamic first-principle models in virtual plants. These models have accurate mass and energy balances and kinetics, which include the activities of the nutrients and cells in order to address many of the issues created by the lack of sufficient test and analysis data [11]. These models can provide real-time inferential measurements of broth composition without the noise and delay associated with analytical results. The models can be run faster than real time for rapid testing and for designing experiments. The results have many uses, including discovering the best batch profile for the current set of inputs, developing PCA logic for batch performance monitoring, and identifying MPC, PLS, and ANN models. The models can fill in the remaining profile of “in progress” batches for MSPC and predict key