16

INTEGRATED PLANNING, SCHEDULING, AND CONTROL IN THE PROCESS INDUSTRIES

THE COORDINATED USE of computers throughout the entire spectrum of manufacturing and business operations has been growing during the 1990s and is expected to continue during the 21st century. With the continued increases in computing power and advances in telecommunications, the use of optimization has expanded as well, including planning and scheduling, plantwide management, unit management, and data acquisition and monitoring. Coordination of manufacturing with computers has been known since the 1970s as computer-integrated manufacturing (CIM). CIM is defined as a unified network of computer hardware, software, and manufacturing systems that combine business and process functions including administration, economic analysis, scheduling, design, control, operations, interactions among suppliers, multiple plant sites, distribution sites, transportation networks, and customers. Also called *process operations*, the goal of CIM is the management and use of human, capital, material, energy, and information resources to produce desired products safely, flexibly, reliably, and cost-effectively, as rapidly as possible and in an environmentally responsible manner (often characterized as "good, fast, cheap, and clean").

In the CIM paradigm, operations are guided by extensive interchange of information that integrates sales, marketing, manufacturing, supply, and R&D data. Data and information flow in a seamless fashion among the various sectors. In addition, plant material and energy balance data are analyzed continuously, reconciled using nonlinear programming, and unmeasured variables reconstructed using parameter estimation techniques (soft sensors). General access to a common database and enterprise information are provided to managers, engineers, and operations so that optimum decisions can be made and executed in a timely and efficient manner.

In the remainder of this chapter, we address each part of the manufacturing business hierarchy, and explain how optimization and modeling are key tools that help link the components together.

16.1 PLANT OPTIMIZATION HIERARCHY

Figure 16.1 shows the relevant levels for the process industries in the optimization hierarchy for business manufacturing. At all levels the use of optimization techniques can be pervasive although specific techniques are not explicitly listed in the specific activities shown in the figure. In Figure 16.1 the key information sources for the plant decision hierarchy for operations are the enterprise data, consisting of commercial and financial information, and plant data, usually containing the values of a large number of process variables. The critical linkage between models and optimization in all of the five levels is illustrated in Figure 16.1. The first level (planning) sets production goals that meet supply and logistics constraints, and scheduling (layer 2) addresses time-varying capacity and staffing utilization decisions. The term *supply chain* refers to the links in a web of relationships involving materials acquisition, retailing (sales), distribution, transportation, and manufacturing with suppliers. Planning and scheduling usually take place over relatively long time frames and tend to be loosely coupled to the information flow and analysis that

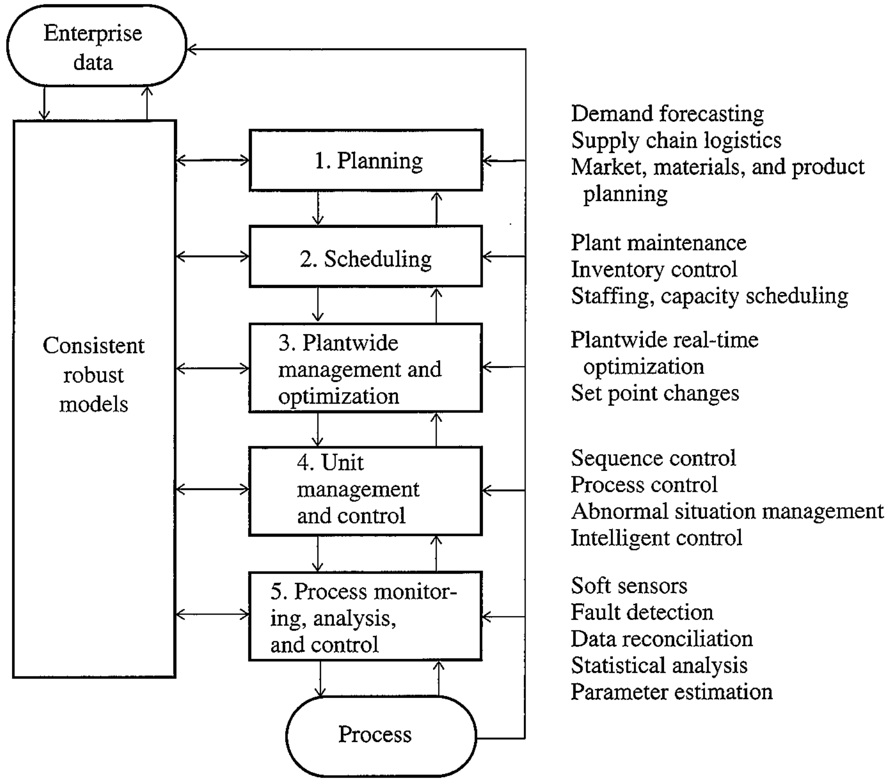

FIGURE 16.1
The five levels of integrated model-based planning, scheduling, optimization, control, and monitoring.

occur at lower levels in the hierarchy. The time scale for decision making at the highest level (planning) may be on the order of months, whereas at the lowest level (e.g., process monitoring) the interaction with the process may be in fractions of a second.

Plantwide management and optimization at level 3 coordinates the network of process units and provides cost-effective setpoints via real-time optimization. The unit management and control level includes process control [e.g., optimal tuning of proportional-integral-derivative (PID) controllers], emergency response, and diagnosis, whereas level 5 (process monitoring and analysis) provides data acquisition and online analysis and reconciliation functions as well as fault detection. Ideally, bidirectional communication occurs between levels, with higher levels setting goals for lower levels and the lower levels communicating constraints and performance information to the higher levels. Data are collected directly at all levels in the enterprise. In practice the decision flow tends to be top down, invariably resulting in mismatches between goals and their realization and the consequent

TABLE 16.1
Types of objective functions and models used in manufacturing system optimization

Optimization level	Objective function	Typical models
1. Planning	Economic	Steady state, single or multiperiod, discrete-event, material flows
2. Scheduling	Economic	Steady state, single or multiperiod, discrete-event, material flows
3. Plantwide management and optimization	Economic	Steady state, linear algebraic correlations or nonlinear simulator
4. Unit management and control		
a. Continuous process	Quadratic-noneconomic or economic	Linear or nonlinear, dynamic, empirical or physically based
b. Batch process	Economic or minimum time	Linear or nonlinear, dynamic or run-to-run, physically based or empirical
5. Process monitoring and analysis		
a. Virtual sensors	Least squares	Nonlinear, physically based, steady state, or empirical
b. Data reconciliation, parameter estimation	Least squares	Linear or nonlinear, steady state or dynamic, physical

accumulation of inventory. Other more deleterious effects include reduction of processing capacity, off-specification products, and failure to meet scheduled deliveries.

Over the past 30 years, business automation systems and plant automation systems have developed along different paths, particularly in the way data are acquired, managed, and stored. Process management and control systems normally use the same databases obtained from various online measurements of the state of the plant. Each level in Figure 16.1 may have its own manually entered database, however, some of which are very large, but web-based data interchange will facilitate standard practices in the future.

Table 16.1 lists the kinds of models and objective functions used in the CIM hierarchy. These models are used to make decisions that reduce product costs, improve product quality, or reduce time to market (or cycle time). Note that models employed can be classified as steady state or dynamic, discrete or continuous, physical or empirical, linear or nonlinear, and with single or multiple periods. The models used at different levels are not normally derived from a single model source, and as a result inconsistencies in the model can arise. The chemical processing industry is, however, moving in the direction of unifying the modeling approaches so that the models employed are consistent and robust, as implied in Figure 16.1. Objective functions can be economically based or noneconomic, such as least squares. In subsequent sections of this chapter we will demonstrate typical optimization problem formulations for each of the five levels, including decision variables, objective function, and constraints.

16.2 PLANNING AND SCHEDULING

Bryant (1993) states that *planning* is concerned with broad classes of products and the provision of adequate manufacturing capacity. In contrast, *scheduling* focuses on details of material flow, manufacturing, and production, but still may be concerned with offline planning. *Reactive scheduling* refers to real-time scheduling and the handling of unplanned changes in demands or resources. The term *enterprise resource planning* (ERP) is used today, replacing the term manufacturing resources planning (MRP); ERP may or may not explicitly include planning and scheduling, depending on the industry. Planning and scheduling are viewed as distinct levels in the manufacturing hierarchy as shown in Figure 16.1, but often a fair amount of overlap exists in the two problem statements, as discussed later on. The time scale can often be the determining factor in whether a given problem is a planning or scheduling one: planning is typified by a time horizon of months or weeks, whereas scheduling tends to be of shorter duration, that is, weeks, days, or hours, depending on the cycle time from raw materials to final product. Bryant distinguishes among system operations planning, plant operations planning, and plant scheduling, using the tasks listed in Table 16.2. At the systems operations planning level traditional multiperiod, multilocation linear programming problems must be solved, whereas at the plant operations level, nonlinear multiperiod models may be used, with variable time lengths that can be optimized as well (Lasdon and Baker, 1986).

TABLE 16.2
Planning and scheduling hierarchy

Corporate operations planning
• Allocate production requirements to plants. • Balance facility's capacity. • Optimize materials and product movements (supply chain).
Plant operations planning
• Determine production plans. • Plan inventory strategy. • Determine raw materials requirements.
Plant scheduling
• Determine run lengths. • Determine sequence of operations. • Provide inventory for production runs.

Source: Bryant (1993).

Baker (1993) outlined the planning and scheduling activities in a refinery as follows:

1. The corporate operations planning model sets target levels and prices for inter-refinery transfers, crude and product allocations to each refinery, production targets, and inventory targets for the end of each refinery model's time horizon.
2. In plant operations planning each refinery model produces target operating conditions, stream allocations, and blends across the whole refinery, which determines (a) optimal operating conditions, flows, blend recipes, and inventories; and (b) costs, cost limits, and marginal values to the scheduling and real-time optimization (RTO) models.
3. The scheduling models for each refinery convert the preceding information into detailed unit-level directives that provide day-by-day operating conditions or set points.

Supply chain management poses difficult decision-making problems because of its wide ranging temporal and geographical scales, and it calls for greater responsiveness because of changing market factors, customer requirements, and plant availability. Successful supply chain management must anticipate customer requirements, commit to customer orders, procure new materials, allocate production capacity, schedule production, and schedule delivery. According to Bryant (1993), the costs associated with supply chain issues represent about 10 percent of the sales value of domestically delivered products, and as much as 40 percent internationally. Managing the supply chain effectively involves not only the manufacturers, but also their trading partners: customers, suppliers, warehousers, terminal operators, and transportation carriers (air, rail, water, land).

In most supply chains each warehouse is typically controlled according to some local law such as a safety stock level or replenishment rule. This local control can cause buildup of inventory at a specific point in the system and thus propagate disturbances over the time frame of days to months (which is analogous to disturbances in the range of minutes or hours that occur at the production control level). Short-term changes that can upset the system include those that are "self-inflicted" (price changes, promotions, etc.) or effects of weather or other cyclical consumer patterns. Accurate demand forecasting is critical to keeping the supply chain network functioning close to its optimum when the produce-to-inventory approach is used.

16.2.1 Planning

Figure 16.2 shows a simplified and idealized version of the components involved in the planning step, that is, the components of the supply chain. S possible suppliers provide raw materials to each of the M manufacturing plants. These plants manufacture a given product that may be stored or warehoused in W facilities (or may not be stored at all), and these in turn are delivered to C different customers. The nature of the problem depends on whether the products are made to order or made

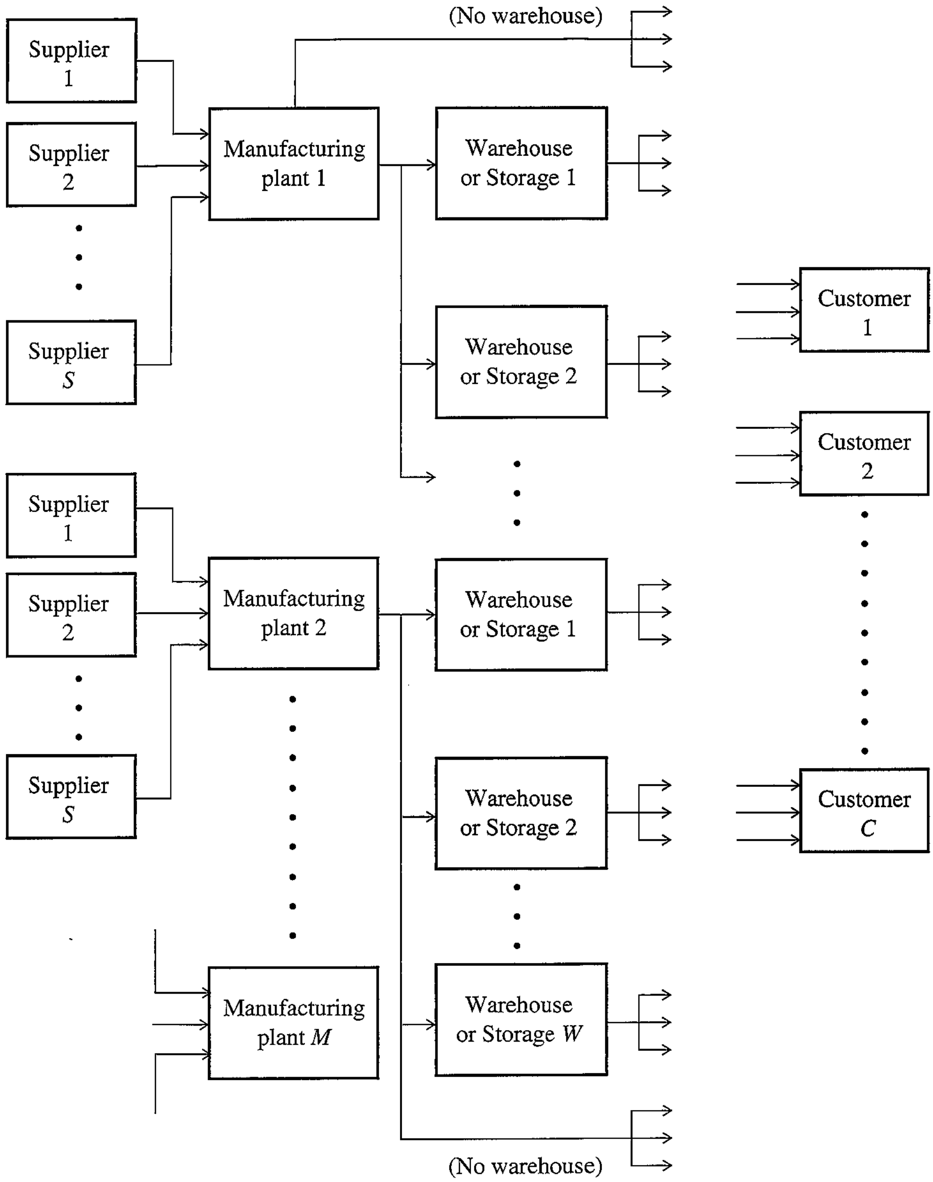

FIGURE 16.2
Supply chain in a manufacturing system.

to inventory; made to order fulfills a specific customer order, whereas made to inventory is oriented to the requirements of the general market demand. Figure 16.2 is similar to a linear allocation process of Chapter 7, with material balance conditions satisfied between suppliers, factories, warehouses, and customers (equality constraints). Inequality constraints would include individual line capacities in each manufacturing plant, total factory capacity, warehouse storage limits, supplier limits, and customer demand. Cost factors include variable manufacturing costs, cost of warehousing, supplier prices, transportation costs (between each sector), and variable customer pricing, which may be volume and quality-dependent. A practical problem may involve as many as 100,000 variables and can be solved using mixed-integer linear programming (MILP); see Chapter 9.

EXAMPLE 16.1 REFINERY PLANNING AND SCHEDULING

Consider a very simple version of a refinery blending and production problem, which is often formulated and solved in an algebraic modeling language such as GAMS (see Chapters 7 and 9). Figure E16.1 is a schematic of feedstocks and products for the refinery. Table E16.1 lists the information pertaining to the expected yields of the four types of crudes when processed by the refinery. Note that the product distribution from the refinery is quite different for the four crudes. The entire multiunit refinery is aggregated into two processes: a fuel chain and a lube chain. Table E16.1 also lists the forecasted upper limits on the established markets for the various products in terms of the allowed maximum weekly production. The processing costs and other data were taken from Karimi (1992).

The problem is to allocate optimally the crudes between the two processes, subject to the supply and demand constraints, so that profits per week are maximized. The objective function and all constraints are linear, yielding a linear programming problem (LP). To set up the LP you must (1) formulate the objective function and (2) formulate the constraints for the refinery operation. You can see from Figure E16.1 that nine variables are involved, namely, the flow rates of each of the crude oils and the four products.

Solution. We want to decide how much of crudes 1, 2, and 3 should be used in the fuel process, and how much of crude 4 should be allocated to the fuel and the lube processes so as to maximize the weekly profit. One decision variable exists for the amount (kbbl/wk) of each crude 1, 2, and 3 used in the fuel process. Two variables exist for the amount (kbbl/wk) of crude 4: one for the amount of crude 4 allocated to the fuel process and the other for the amount allocated to the lube process. Denote the variables by x_c ($c = 1$ to 5), where x_1 through x_3 represent the amounts of crudes 1 through 3, x_4 represents the crude 4 sent to the fuel process, and x_5 represents the crude 4 sent to the lube process. Because the crude supplies are limited, the x_c will be constrained by

$$\begin{aligned} x_1 &\le S_1 \\ x_2 &\le S_2 \\ x_3 &\le S_3 \\ x_4 + x_5 &\le S_4 \end{aligned} \qquad (a)$$

where S_c is the maximum supply (kbbl/wk) of crude c listed in Table E16.1

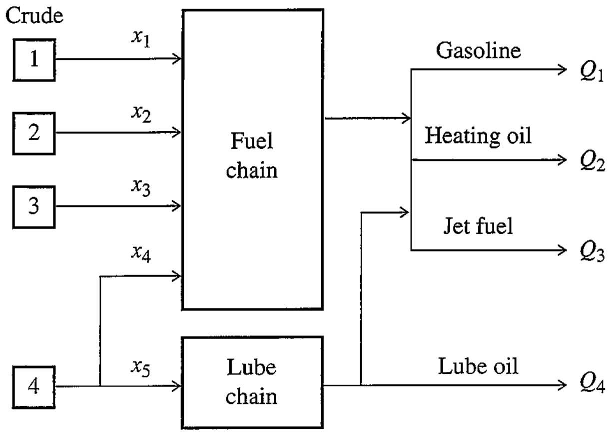

FIGURE E16.1
Processing operation schematic.

TABLE E16.1
Refinery data

	Product yields (bbl/bbl crude)						
	Fuel chain				Lube chain		
Crudes	1 (x_1)	2 (x_2)	3 (x_3)	4 (x_4)	4 (x_5)	Product value [selling price ($/bbl)]	Maximum demand (10^3 bbl/wk)
Products							
Gasoline (P_1)	0.6	0.5	0.3	0.4	0.4	45.00	170
Heating oil (P_2)	0.2	0.2	0.3	0.3	0.1	30.00	85
Jet fuel (P_3)	0.1	0.2	0.3	0.2	0.2	15.00	85
Lube oil (P_4)	0.0	0.0	0.0	0.0	0.2	60.00	20
Operating losses	0.1	0.1	0.1	0.1	0.1	—	—
Crude cost ($/bbl)	15.00	15.00	15.00	25.00	25.00		
Operating cost ($/bbl)	5.00	8.50	7.50	3.00	2.50		
Available crude supply (10^3 bbl/wk)	100	100	100	200			

Next, we want to find the amounts of different products produced for the given usage x_c of the crudes. Let Q_p ($p = 1$ to 4) refer to the gasoline, heating oil, jet fuel, and lube oil, respectively. Define Q_p as the amount (kbbl) of product p produced, and let a_{pi} denote the yield of product p from crude x (in bbl/bbl of crude); ($a_{23} = 0.3$, $a_{35} = 0.2$, etc.) Thus, using the a_{pc} from Table E16.1,

$$Q_p = a_{p1}x_1 + a_{p2}x_2 + a_{p3}x_3 + a_{p4}x_4 + a_{p5}x_5, \quad p = 1, \ldots, 4 \tag{b}$$

Let D_p be the maximum demand for product $p(D_1 = 170$, etc.). The maximum demands D_p provide the upper bounds on Q_p.

$$Q_p \leq D_p, \quad p = 1, \dots, 4 \tag{c}$$

Finally, we will formulate the objective function. Using the production amounts Q_p and the crude selection x_c, we can calculate the profit as total income from product sales minus the total production cost. If $v_p(p = 1$ to $4)$ is the value of product p, then total income (k\$) from product sales is $v_1Q_1 + v_2Q_2 + v_3Q_3 + v_4Q_4$. The production cost consists of the costs of crudes and the operating costs. Let $C_c(c = 1$ to $5)$ denote the sum of crude and operating costs (\$/bbl) for crude usage x_c (e.g., $C_1 = \$20/\text{bbl}$). Then the total production cost is $\sum_{c=1}^{5} C_c x_c$. Therefore, the complete problem statement is

$$\text{Maximize:} \quad \sum_{p=1}^{4} v_p Q_p - \sum_{c=1}^{5} C_c x_c$$

$$\begin{aligned} \text{Subject to:} \quad & x_1 \leq S_1 \\ & x_2 \leq S_2 \\ & x_3 \leq S_3 \\ & x_4 + x_5 \leq S_4 \end{aligned}$$

$$Q_p \leq D_p \; (p = 1, \dots, 4) \tag{a}$$

$$Q_p = a_{p1}x_1 + a_{p2}x_2 + a_{p3}x_3 + a_{p4}x_4 + a_{p5}x_5, \quad p = 1, \dots, 4 \tag{b}$$

$$Q_p \geq 0 \quad (p = 1, \dots, 4)$$

$$x_c \geq 0 \quad (c = 1, \dots, 5) \tag{c}$$

The problem involves nine optimization variables (x_c, $c = 1$ to 5; Q_p, $p = 1$ to 4) in the preceding formulation. All are continuous variables. The objective function is a linear function of these variables, and so are Equations (a) and (b), hence the problem is a linear programming problem and has a globally optimal solution.

Results. The optimal solution can be obtained using GAMS (Karimi, 1992); the optimum flows are 100, 100, 66.667, and 100 kbbl/wk, respectively, of crudes 1, 2, 3, and 4 and 170, 70, 70, and 20 kbbl/wk, respectively, of gasoline, heating oil, jet fuel, and lube oil are produced. All of crude 4 is used in the lube chain. The maximum profit obtained is 3400 k\$/wk.

As discussed by Karimi (1992), the results for this problem can be interpreted by considering the profit per kilobarrel for each crude. For 1 kbbl/wk of crude 1, we can produce 0.6 kbbl/wk of gasoline, 0.2 kbbl/wk of heating oil and 0.1 kbbl/wk of jet fuel, with production cost of 20 k\$/kbbl/wk and value of the products of $45 * 0.6 + 30 * 0.2 + 15 * 0.1 = 34.5$ k\$. Thus, for 1 kbbl/wk of crude 1, a profit of 14.5 k\$ results. A similar analysis for other crudes yields 8.0 k\$, 4.5 k\$, 2 k\$, and 8.5 k\$, respectively, for crude variables 2, 3, 4, and 5; the priority for the crude options should be 1, 5, 2, 3, and 4. Note that all of crude 1 is used in the optimal solution. Using 100 kbbl/wk of crude 1 produces 60 kbbl/wk of gasoline, 20 kbbl/wk of heating oil, and 10 kbbl/wk of jet fuel. Because this does not exceed the demands of any of the products, the next most profitable crude (crude 4) can be used in the lube process. Because demand for lube oil cannot be exceeded, only 100 kbbl/wk of crude 4 can be used in the lube process. Next we can use crude 2, because it does not produce lube oil and is the next most profitable crude. If all of crude 2 (100 kbbl/wk) is processed, the production amounts become 150, 50, 50, and 20 kbbl/wk, respectively, but more products can still be manufactured. The maximum amount of crude 3 that can be used without exceeding any of the product demands is 66.667 kbbl/wk, when the demand of gasoline is equaled. Finally, crude 4 cannot be consumed in the fuel process, because it also produces gasoline and it is not economical to produce any more gasoline.

Most international oil companies that operate multiple refineries analyze the refinery optimization problem over several time periods (e.g., 3 months). This is because many crudes must be purchased at least 3 months in advance due to transportation requirements (e.g., the need to use tankers to transport oil from the Middle East). These crudes also have different grades and properties, which must be factored into the product slate for the refinery. So the multitime period consideration is driven more by supply and demand than by inventory limits (which are typically less than 5 days). The LP models may be run on a weekly basis to handle such items as equipment changes and maintenance, short-term supply issues (and delays in shipments due to weather problems or unloading difficulties), and changes in demand (4 weeks within a 1-month period). Product properties such as the Reid vapor pressure must be changed between summer and winter months to meet environmental restrictions on gasoline properties. See Pike (1986) for a detailed LP refinery example that treats quality specifications and physical properties by using product blending, a dimension not included in Example 16.1 but one that is relevant for companies with varied crude supplies and product requirements.

16.2.2 Scheduling

Information processing in production scheduling is essentially the same as in planning. Both plants and individual process equipment take orders and make products. For a plant, the customer is usually external, but for a process (or "work cell" in discrete manufacturing parlance), the order comes from inside the plant or factory. In a plant, the final product can be sold to an external customer; for a process, the product delivered is an intermediate or partially finished product that goes on to the next stage of processing (internal customer).

Two philosophies are used to solve production scheduling problems (Puigjaner and Espura, 1998):

1. The top-down approach, which defines appropriate hierarchical coordination mechanisms between the different decision levels and decision structures at each level. These structures force constraints on lower operating levels and require heuristic decision rules for each task. Although this approach reduces the size and complexity of scheduling problems, it potentially introduces coordination problems.

TABLE 16.3
Characteristics of batch scheduling and planning problems

DETERMINE	GIVEN
What Product amounts: lot sizes, batch sizes	**Product requirements** Horizon, demands, starting and ending inventories
When Timing of specific operations, run lengths	**Operational steps** Precedence order Resource utilization
Where Sites, units, equipment items	**Production facilities** Types, capacities
How Resource types and amounts	**Resource limitations** Types, amounts, rates

Source: Pekny and Reklaritis (1998).

2. The bottom-up approach, which develops detailed plant simulation and optimization models, optimizes them, and translates the results from the simulations and optimization into practical operating heuristics. This approach often leads to large models with many variables and equations that are difficult to solve quickly using rigorous optimization algorithms.

Table 16.3 categorizes the typical problem statement for the manufacturing scheduling and planning problem. In a batch campaign or run, comprising smaller runs called lots, several batches of product may be produced using the same recipe. To optimize the production process, you need to determine

1. The recipe that satisfies product quality requirements.
2. The production rates needed to fulfill the timing requirements, including any precedence constraints.
3. Operating variables for plant equipment that are subject to constraints.
4. Availability of raw material inventories.
5. Availability of product storage.
6. The run schedule.
7. Penalties on completing a production step too soon or too late.

EXAMPLE 16.2 MULTIPRODUCT BATCH PLANT SCHEDULING

Batch operations such as drying, mixing, distillation, and reaction are widely used in producing food, pharmaceuticals, and specialty products (e.g., polymers). Scheduling of operations as described in Table 16.3 is crucial in such plants. A principal feature of batch plants (Ku and Karimi, 1987) is the production of multiple products using the same set of equipment. Good industrial case studies of plant scheduling include those by Bunch et al. (1998), McDonald (1998), and Schulz et al. (1998). For example, Schulz et al. described a polymer plant that involved four process steps (preparation, reaction, mixing, and finishing) using different equipment in each step. When products are similar in nature, they require the same processing steps and hence pass through the same series of processing units; often the batches are produced sequentially. Such plants are called multiproduct plants. Because of different processing time requirements, the total time required to produce a set of batches (also called the makespan or cycle time) depends on the sequence in which they are produced. To maximize plant productivity, the batches should be produced in a sequence that minimizes the makespan. The plant schedule corresponding to such a sequence can then be represented graphically in the form of a Gantt chart (see the following discussion and Figure E16.2b). The Gantt chart provides a timetable of plant operations showing which products are produced by which units and at what times. Chapter 10 discusses a single-unit sequencing problem.

In this example we consider four products ($p1, p2, p3, p4$) that are to be produced as a series of batches in a multiproduct plant consisting of three batch reactors in series (Ku and Karimi, 1992); see Figure E16.2a. The processing times for each batch reactor and each product are given in Table E16.2. Assume that no intermediate storage is available between the processing units. If a product finishes its processing on unit k and unit $k + 1$ is not free because it is still processing a previous product, then the completed product must be kept in unit k, until unit $k + 1$ becomes free. As an example, product $p1$ must be held in unit 1 until unit 2 finishes processing $p3$. When a product finishes processing on the last unit, it is sent immediately to product storage. Assume that the times required to transfer products from one unit to another are negligible compared with the processing times.

The problem for this example is to determine the time sequence for producing the four products so as to minimize the makespan. Assume that all the units are initially

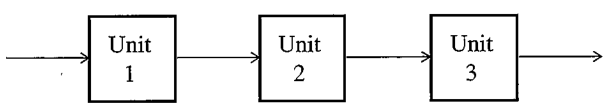

FIGURE E16.2a
Multiproduct plant.

TABLE E16.2
Processing times (h) of products

	Products			
Units	$p1$	$p2$	$p3$	$p4$
1	3.5	4.0	3.5	12.0
2	4.3	5.5	7.5	3.5
3	8.7	3.5	6.0	8.0

empty (initialized) at time zero and the manufacture of any product can be delayed an arbitrary amount of time by holding it in the previous unit.

Solution. Let N be the number of products and M be the number of units in the plant. Let $C_{j,k}$ (called completion time) be the "clock" time at which the jth product in the sequence leaves unit k after completion of its processing, and let $\tau_{j,k}$ be the time required to process the jth product in the sequence on unit k (See Table E16.2). The first product goes into unit 1 at time zero, so $C_{1,0} = 0$. The index j in $\tau_{j,k}$ and $C_{j,k}$ denotes the position of a product in the sequence. Hence $C_{N,M}$ is the time at which the last product leaves the last unit and is the makespan to be minimized. Next, we derive the set of constraints (Ku and Karimi, 1988; 1990) that interrelate the $C_{j,k}$. First, the jth product in the sequence cannot leave unit k until it is processed, and in order to be processed on unit k, it must have left unit $k - 1$. Therefore the clock time at which it leaves unit k (i.e., $C_{j,k}$) must be equal to or after the time at which it leaves unit $k - 1$ plus the processing time in k. Thus the first set of constraints in the formulation is

$$C_{j,k} \geq C_{j,k-1} + \tau_{j,k} \quad j = 1, \dots, N \quad k = 2, \dots, M \tag{a}$$

Similarly, the jth product cannot leave unit k until product $(j - 1)$ has been processed and transferred:

$$C_{j,k} \geq C_{j-1,k} + \tau_{j,k} \quad j = 1, \dots, N \quad k = 1, \dots, M \tag{b}$$

Set $C_{0,k} = 0$. Finally the jth product in the sequence cannot leave unit k until the downstream unit $k + 1$ is free [i.e., product $(j - 1)$ has left]. Therefore

$$C_{j,k} \geq C_{j-1,k+1} \quad j = 1, \dots, N \quad k = 1, \dots, M - 1 \tag{c}$$

Although Equations (a)–(c) represent the complete set of constraints, some of them are redundant. From Equation (a) $C_{j,k} \geq C_{j,k-1} + \tau_{i,k}$ for $k \geq 2$. But from Equation (c), $C_{j,k-1} \geq C_{j-1,k}$, hence $C_{j,k} \geq C_{j-1,k} + \tau_{i,k}$ for $k = 2, M$. In essence, Equations (a) and (c) imply Equations (b) for $k = 2, M$, so Equations (b) for $k = 2, M$ are redundant.

Having derived the constraints for completion times, we next determine the sequence of operations. In contrast to the $C_{j,k}$, the decision variables here are discrete (binary). Define $X_{i,j}$ as follows. $X_{i,j} = 1$ if product i (product with label pi) is in slot j of the sequence, otherwise it is zero. So $X_{3,2} = 1$ means that product $p3$ is second in the production sequence, and $X_{3,2} = 0$ means that it is not in the second position. The overall integer constraint is

$$X_{1,j} + X_{2,j} + X_{3,j} + X_{4,j} + \cdots + X_{N,j} = 1 \quad j = 1, \dots, \tag{d}$$

Similarly every product should occupy only one slot in the sequence:

$$X_{i,1} + X_{i,2} + X_{i,3} + X_{i,4} + \cdots + X_{i,N} = 1 \quad i = 1, \dots, N \tag{e}$$

The $X_{i,j}$ that satisfy Equations (d) and (e) always give a meaningful sequence. Now we must determine the clock times $t_{i,k}$ for any given set of $X_{i,j}$. If product pi is in slot j, then $t_{j,k}$ must be $\tau_{i,k}$ and $X_{i,j} = 1$ and $X_{i,1} = X_{i,2} = \cdots = X_{i,j-1} = X_{i,j+1} = \cdots = X_{i,N} = 0$, therefore we can use $X_{i,j}$ to pick the right processing time representing $t_{j,k}$ by imposing the constraint.

$$\tau_{j,k} = X_{1,j}t_{i,k} + X_{2,j}t_{2,k} + X_{3,j}t_{3,k} + \cdots + X_{N,j}t_{N,k} \qquad j = 1, \dots, N \qquad k = 1, \dots, M \tag{f}$$

To reduce the number of constraints, we substitute $\tau_{j,k}$ from Equation (f) into Equations (a) and (b) to obtain the following formulation (Ku and Karimi, 1988).

Minimize: C_{NM}

Subject to: Equations (c), (d), (e) and

$$C_{i,k} \geq C_{i,k-1} + \sum_{j=1}^{N} X_{j,i} t_{j,k} \quad i = 1, \ldots, N \quad k = 2, \ldots, M \tag{g}$$

$$C_{i,1} \geq C_{i-1,1} + \sum_{j=1}^{N} X_{j,i} t_{j,i} \quad i = 1, \ldots, N \tag{h}$$

$C_{i,k} \geq 0$ and $X_{i,j}$ binary

Because the preceding formulation involves binary ($X_{i,j}$) as well as continuous variables ($C_{i,k}$) and has no nonlinear functions, it is a mixed-integer linear programming (MILP) problem and can be solved using the GAMS MIP solver.

Solving for the optimal sequence using Table E16.2, we obtain $X_{1,1} = X_{2,4} = X_{3,2} = X_{4,3} = 1$. This means that $p1$ is in the first position in the optimal production sequence, $p2$ in the fourth, $p3$ in the second, and $p4$ in the third. In other words, the optimal sequence is in the order $p1$-$p3$-$p4$-$p2$. In contrast to the $X_{i,j}$, we must be careful in interpreting the $C_{i,k}$ from the GAMS output, because $C_{j,k}$ really means the time at which the jth product in the sequence (and not product pi) leaves unit k. Therefore $C_{2,3} = 23.3$ means that the second product (i.e., $p3$) leaves unit 3 at 23.3 h. Interpreting the others in this way, the schedule corresponding to this production sequence is conveniently displayed in form of a Gantt chart in Figure E16.2b, which shows the status of the units at different times. For instance, unit 1 is processing $p1$ during [0, 3.5] h. When $p1$ leaves unit 1 at $t = 3.5$ h, it starts processing $p3$. It processes $p3$ during [3.5, 7] h. But as seen from the chart, it is unable to discharge $p3$ to unit 2, because unit 2 is still processing $p1$. So unit 1 holds $p3$ during [7, 7.8] h. When unit 2 discharges $p3$

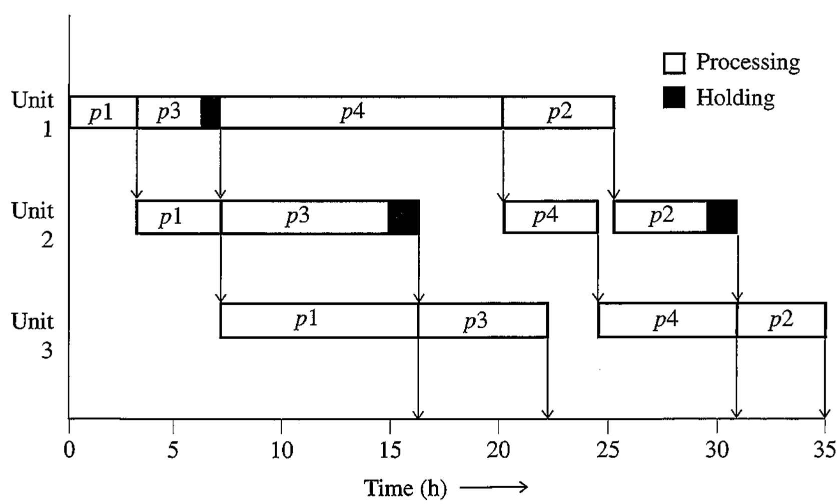

FIGURE E16.2b
Gantt chart for the optimal multiproduct plant schedule.

to unit 3 at 16.5 h, unit 1 is still processing $p4$, therefore unit 2 remains idle during [16.5, 19.8] h. It is common in batch plants to have units blocked due to busy downstream units or units waiting for upstream units to finish. This happens because the processing times vary from unit to unit and from product to product, reducing the time utilization of units in a batch plant. The finished batches of $p1$, $p3$, $p4$, and $p2$ are completed at times 16.5 h, 23.3 h, 31.3 h, and 34.8 h. The minimum makespan in 34.8 h.

This problem can also be solved by a search method (see Chapter 10). Because the order of products cannot be changed once they start through the sequence of units, we need only determine the order in which the products are processed. This is the same problem as considered in Section 10.5.2, to illustrate the workings of tabu search. Using the notation of that section, let

$$\mathbf{P} = (p(1), p(2), \dots, p(N))$$

be a permutation or sequence in which to process the jobs, where $p(j)$ is the index of the product in position j of the sequence. To evaluate the makespan of a sequence, we proceed as in Equations (a)–(c) of the mixed-integer programming version of the problem. Let $C_{j,k}$ be the completion time of product $p(j)$ on unit k. If product $p(j)$ does not have to wait for product $p(j - 1)$ to finish its processing on unit k, then

$$C_{j,k} = C_{j,k-1} + t_{p(j),k} \qquad (i)$$

If it does have to wait, then

$$C_{j,k} = C_{j-1,k} + t_{p(j),k} \qquad (j)$$

Hence $C_{j,k}$ is the larger of these two values:

$$C_{j,k} = \max(C_{j-1,k} + t_{p(j),k}, C_{j,k-1} + t_{p(j),k}) \qquad (k)$$

This equation is solved first for $C_{1,k}$ for $k = 1, \dots, M$, then for $C_{2,k}$ for $k = 1, 2, \dots, M$, and so on. The objective function is simply the completion time of the last job:

$$f(\mathbf{P}) = C_{N,M} \qquad (l)$$

In a four-product problem, there are only $4! = 24$ possible sequences, so you can easily write a simple FORTRAN or C program to evaluate the makespan for an arbitrary sequence, and then call it 24 times and choose the sequence with the smallest makespan. For larger values of N, one can apply the tabu search algorithm described in Section 10.5.2. Other search procedures (e.g., evolutionary algorithms or simulated annealing), can also be developed for this problem. Of course, these algorithms do not guarantee that an optimal solution will be found. On the other hand, the time required to solve the mixed-integer programming formulation grows rapidly with N, so that approach eventually becomes impractical. This illustrates that you may be able to develop a simple but effective search method yourself, and eliminate the need for MILP optimization software.

The classical solution to a scheduling problem assumes that the required information is known at the time the schedule is generated and that this a priori scheduling remains fixed for a planning period and is implemented on the plant equipment. Although this methodology does not compensate for the many external disturbances and internal disruptions that occur in a real plant, it is still the strategy most commonly found in industrial practice. Demand fluctuations, process deviations, and equipment failure all result in schedule infeasibilities that become apparent during the implementation of the schedule. To remedy this situation, frequent rescheduling becomes necessary.

In the rolling horizon rescheduling approach (Baker, 1993), a multiperiod solution is obtained, but only the first period is implemented. After one period has elapsed, we observe the existing inventories, create new demand forecasts, and solve a new multiperiod problem. This procedure tries to compensate for the fixed nature of the planning model. However, as has been pointed out by Pekny and Reklaitis (1998), schedules generated in this fashion generally result in frequent resequencing and reassignment of equipment and resources, which may induce further changes in successive schedules rather than smoothing out the production output. An alternative approach uses a master schedule for planning followed by a reactive scheduling strategy to accommodate changes by readjusting the master schedule in a least cost or least change way.

The terms *able to promise* or *available to promise* (ATP) indicate whether a given customer, product, volume, date, or time request can be met for a potential order. ATP requests might be filled from inventory, unallocated planned production, or spare capacity (assuming additional production). When the production scheduler is content with the current plan, made up of firm orders and forecast orders, the forecast orders are removed but the planned production is left intact. This produces inventory profiles in the model that represent ATP from inventory and from unallocated planned production (Baker, 1993; Smith, 1998).

An important simulation tool used in solving production planning and scheduling problems is the *discrete event dynamic system* (DEDS), which gives a detailed picture of the material flows through the production process. Software for simulating such systems are called discrete event simulators. In many cases, rules or expert systems are used to incorporate the experience of scheduling and planning personnel in lieu of a purely optimization-based approach to scheduling (Bryant, 1993). Expert systems are valuable to assess the effects of changes in suppliers, to locate bottlenecks in the system, and to ascertain when and where to introduce new orders. These expert systems are used in reactive scheduling when fast decisions need to be made, and there is no time to generate another optimized production schedule.

16.3 PLANTWIDE MANAGEMENT AND OPTIMIZATION

At the plantwide management and optimization level (see Figure 16.1), engineers strive for enhancements in the operation of the equipment once it is installed in order to realize the most production, the greatest profit, the minimum cost, the least energy usage, and so on. In plant operations, benefits arise from improved plant performance, such as improved yields of valuable products (or reduced yields of contaminants), better product quality, reduced energy consumption, higher processing rates, and longer times between shut downs. Optimization can also lead to reduced maintenance costs, less equipment wear, and better staff utilization. Optimization can take place plantwide or in combinations of units.

The application of real-time optimization (RTO) in chemical plants has been carried out since the 1960s. Originally a large mainframe computer was used to optimize process setpoints, which were then sent to analog controllers for implementation. In the 1970s this approach, called supervisory control, was incorporated into computer control systems with a distributed microprocessor architecture called a *distributed control system*, or DCS (Seborg et al., 1989). In the DCS both supervisory control and regulatory (feedback) control were implemented using digital computers. Because computer power has increased by a factor of 10^6 over the past 30 years, it is now feasible to solve meaningful optimization problems using advanced tools such as linear or nonlinear programming in real time, meaning faster than the time between setpoint changes.

In RTO (level 3), the setpoints for the process operating conditions are optimized daily, hourly, or even every minute, depending on the time scale of the process and the economic incentives to make changes. Optimization of plant operations determines the setpoints for each unit at the temperatures, pressures, and flow rates that are the best in some sense. For example, the selection of the percentage of excess air in a process heater is quite critical and involves a balance on the fuel–air ratio to ensure complete combustion and at the same time maximize use of the heating potential of the fuel. Examples of periodic optimization in a plant are minimizing steam consumption or cooling water consumption, optimizing the reflux ratio in a distillation column, blending of refinery products to achieve desirable physical properties, or economically allocating raw materials. Many plant maintenance systems have links to plant databases to enable them to track the operating status of the production equipment and to schedule calibration and maintenance. Real-time data from the plant also may be collected by management information systems for various business functions.

The objective function in an economic model in RTO involves the costs of raw materials, values of products, and costs of production as functions of operating conditions, projected sales or interdepartmental transfer prices, and so on.

Both the operating and economic models typically include constraints on

(a) *Operating Conditions*: Temperatures and pressures must be within certain limits.
(b) *Feed and Production Rates*: A feed pump has a fixed capacity; sales are limited by market projections.
(c) *Storage and Warehousing Capacities*: Storage tanks cannot be overfilled during periods of low demand.
(d) *Product Impurities*: A product may contain no more than the maximum amount of some contaminant or impurity.

In addition, safety or environmental constraints might be added, such as a temperature limit or an upper limit on a toxic species. Several steps are necessary for implementation of RTO, including determining the plant steady-state operating conditions, gathering and validating data, updating of model parameters (if necessary) to match current operations, calculating the new (optimized) setpoints, and implementing these setpoints. An RTO system completes all data transfer, optimization calculations, and setpoint implementations before unit conditions change and require a new optimum to be calculated.

A number of RTO problems characteristic of level 3 in Figure 16.1 have been presented in earlier chapters of this book:

1. Reflux ratio in distillation (Example 12.2).
2. Olefin manufacture (Example 14.1).
3. Ammonia synthesis (Example 14.2).
4. Hydrocarbon refrigeration (Example 15.2).

The last example is particularly noteworthy because it represents the current state of the art in utilizing fundamental process models in RTO.

Another activity in RTO is determining the values of certain empirical parameters in process models from the process data after ensuring that the process is at steady state. Measured variables including flow rates, temperatures, compositions, and pressures can be used to estimate model parameters such as heat transfer coefficients, reaction rate coefficients, catalyst activity, and heat exchanger fouling factors. Usually only a few such parameters are estimated online, and then optimization is carried out using the updated parameters in the model. Marlin and Hrymak (1997) and Forbes et al. (1994) recommend that the updated parameters be observable, represent actual changes in the plant, and significantly influence the location of the optimum; also the optimum of the model should be coincident with that of the true process. One factor in modeling that requires close attention is the accurate representation of the process constraints, because the optimum operating conditions usually lie at the intersection of several constraints. When RTO is combined with model predictive regulatory control (see Section 16.4), then correct (optimal) moves of the manipulated variables can be determined using models with accurate constraints.

Marlin and Hrymak (1997) reviewed a number of industrial applications of RTO, mostly in the petrochemical area. They reported that in practice a maximum change in plant operating variables is allowable with each RTO step. If the computed optimum falls outside these limits, you must implement any changes over several steps, each one using an RTO cycle. Typically, more manipulated variables than controlled variables exist, so some degrees of freedom exist to carry out both economic optimization as well as establish priorities in adjusting manipulated variables while simultaneously carrying out feedback control.

16.4 UNIT MANAGEMENT AND CONTROL

Because of greater integration of plant equipment, tighter quality specifications, and more emphasis on maximum profitability while maintaining safe operating conditions, implementation of advanced multivariable process control is increasing. The distributed control system (DCS) architecture for computer control mentioned in the previous section normally uses feedback control based on a proportional integral derivative (PID) controller at the implementation level for regulatory control (Seborg et al., 1989). Although in principle you can select the three design parameters for PID control in an individual control loop using an optimization technique discussed in Chapter 6 (based on minimizing the sum of squares of the error from setpoint), this design method is not the normal approach currently taken for level 4 in Figure 16.1 (unit management and control). In industrial practice today, advanced multivariable control strategies are being applied using a mathematical programming approach, which is the main topic of this section.

Model predictive control (MPC) refers to a class of control techniques in which a process model is used to predict the future values of the process outputs, and these predictions are used in computing the best control strategy. The most powerful MPC techniques are based on optimization of a quadratic objective function involving the error between the setpoints and predicted outputs. MPC is especially well suited for difficult multiple-input/multiple-output (MIMO) control problems, in which significant interactions exist between the manipulated inputs and the controlled outputs. In addition, MPC can easily accommodate inequality constraints on the input and output variables, such as upper and lower limits, or rate-of-change limits. The operating goal is to keep the process variables within their limits while moving the process to an economic optimum. The success of model predictive control in solving large multivariable industrial control problems is impressive, perhaps even reaching the status of a "killer" application. Control of units with as many as 60 inputs and 40 outputs is already established in industrial practice. Since the 1970s more than a thousand applications of MPC techniques have been used in oil refineries and petrochemical plants around the world. Thus, MPC has had a substantial influence and is currently the method of choice for difficult multivariable control problems in these industries (Camacho and Bordons, 1999).

A key feature of MPC is that future process behavior is predicted using a dynamic model and the available measurements. The controller outputs are calculated so as to minimize the difference between the predicted process response and the desired response. At each sampling instant the control calculations are repeated and the predictions updated based on current measurements, which is a moving horizon approach. Garcia et al. (1989), Richalet (1993), and Qin and Badgwell (1997) have provided surveys of the MPC approach.

Because empirical dynamic models are generally used, they are only valid over the range of conditions considered during the original plant tests, but MPC can be adapted to optimize plant performance. In this case the control strategy is updated periodically to compensate for changes in process conditions, constraints, or performance criteria. Here the MPC calculations need to be done more frequently (e.g., solving an LP or QP problem at each sampling instant) and thus may require an increased amount of computer resources.

16.4.1 Formulating the MPC Optimization Problem

In MPC a dynamic model is used to predict the future output over the prediction horizon based on a set of control changes. The desired output is generated as a setpoint that may vary as a function of time; the prediction error is the difference between the setpoint trajectory and the model prediction. A model predictive controller is based on minimizing a quadratic objective function over a specific time horizon based on the sum of the square of the prediction errors plus a penalty related to the square of the changes in the control variable(s). Inequality constraints on the input and output variables can be included in the optimization calculation. At each sampling instant, values of the manipulated variables and controlled variables for the next m time steps are calculated; m is the number of control "moves," and its selection is discussed later. At each sampling instant, only the first control move (of the m moves that were calculated) is actually implemented. Then, the prediction and control calculations are repeated at the next sampling instant, based on the currently measured state of the process.

In principle, any type of process model can be used to predict future values of the controlled outputs. For example, one can use a physical model based on first principles (e.g., mass and energy balances), a linear model (e.g., transfer function, step response model, or state space-model), or a nonlinear model (e.g., neural nets). Because most industrial applications of MPC have relied on linear dynamic models, later on we derive the MPC equations for a single-input/single-output (SISO) model. The SISO model, however, can be easily generalized to the MIMO models that are used in industrial applications (Lee et al., 1994). One model that can be used in MPC is called the step response model, which relates a single controlled variable y with a single manipulated variable u (based on previous changes in u) as follows:

$$\hat{y}(k) = \sum_{i=1}^{N} S_i \Delta u(k - i) + y(0) \tag{16.1}$$

where $\hat{y}(k)$ is the predicted value of $y(k)$ at the k-sampling instant ($k = 1, 2, \ldots$), $\Delta u(k - i)$ is the change in the manipulated input at time $k - i\,[\Delta u(k - i) = u(k - i) - u(k - i - 1)]$, N is the number of terms in the step response model (usually less than 50), and the N model parameters S_{ix} are referred to as the step response coefficients. The initial value $y(0)$ is assumed to be known. Other model forms in MPC can involve fewer parameters and can be expressed using state space form (Lee et al., 1994), which is now more frequently used in commercial software packages.

Figure 16.3 shows a hypothetical step response for an industrial process generated by a step change in the manipulated variable u. The model is developed by performing a step change in $u(k)$ and recording the response $y(k)$ until it essentially reaches steady state. In theory, the S_i can be determined from a single-step response but in practice a number of step tests are required to compensate for unanticipated disturbances, process nonlinearities, and noisy measurements. The step response coefficients S_i can be estimated by applying linear regression to the values of the output variable at each sampling instant. Usually the final or steady-state value $y(ss)$ is the last sampled value of y, and the number of data points is selected to be larger than N, the number of terms in the model.

We now develop a mathematical statement for model predictive control with a quadratic objective function for each sampling instant k and linear process model in Equation 16.1:

$$\min f = \sum_{i=1}^{p} w_i e^2(k + i) + \lambda \sum_{i=1}^{m} \Delta u^2(k + i - 1) \tag{16.2}$$

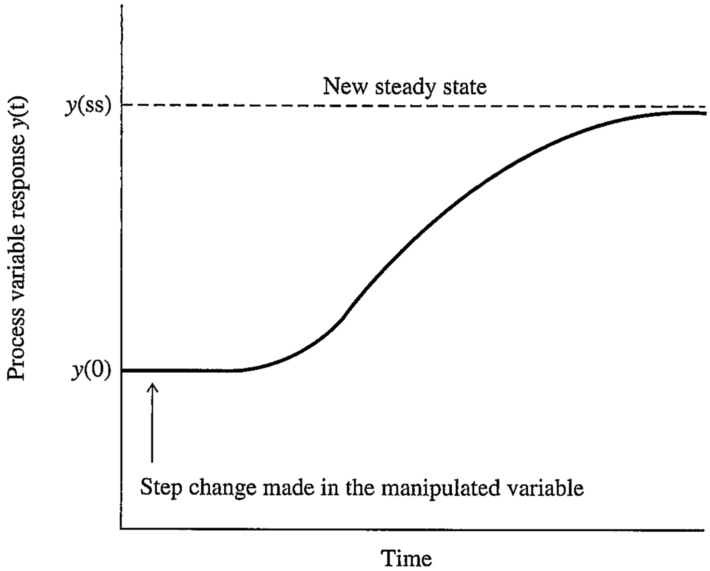

FIGURE 16.3
Typical step response for an industrial process. A time delay may occur between the time that the manipulated variable is changed and the time that the process response occurs.

where $e(k+i)$ denotes the predicted error at time $(k+i)$, $i = 1, \ldots, p$,

$$e(k+i) = r(k+i) - \hat{y}(k+i) \tag{16.3}$$

$r(k+i)$ is the reference value or setpoint at time $k+i$, and $\Delta u(k)$ denotes the vector of current and future control moves over the next m sampling instants:

$$\Delta u(k) = [\Delta u(k), \Delta u(k+1), \ldots, \Delta u(k+m-1)]^T \tag{16.4}$$

To minimize f, you balance the error between the setpoint and the predicted response against the size of the control moves. Equation 16.2 contains design parameters that can be used to tune the controller, that is, you vary the parameters until the desired shape of the response that tracks the setpoint trajectory is achieved (Seborg et al., 1989). The "move suppression" factor λ penalizes large control moves, but the weighting factors w_i allow the predicted errors to be weighted differently at each time step, if desired. Typically you select a value of m (number of control moves) that is smaller than the prediction horizon p, so the control variables are held constant over the remainder of the prediction horizon.

Inequality constraints on future inputs or their rates of change are widely used in the MPC calculations. For example, if both upper and lower limits on u and Δu are required, the constraints could be expressed as

$$B^l \leq u(k+i) \leq B^u, \qquad \text{for } i = 1, 2, \ldots, m \tag{16.5}$$

$$C^l \leq \Delta u(k+i) \leq C^u, \quad \text{for } i = 1, 2, \ldots, m \tag{16.6}$$

where the B^l, C^l, and B^u, C^u are lower and upper bounds, respectively. Note that $u(k+i)$ is determined by whatever values $\Delta u(k+i)$ assume. Constraints on the predicted outputs are sometimes included as well:

$$D^l \leq \hat{y}(k+i) \leq D^u, \quad \text{for } i = 1, 2, \ldots, p \tag{16.7}$$

The minimization of the quadratic performance index in Equation (16.2), subject to the constraints in Equations (16.5–16.7) and the step response model such as Equation (16.1), forms a standard quadratic programming (QP) problem, described in Chapter 8. If the quadratic terms in Equation (16.2) are replaced by linear terms, a linear programming program (LP) problem results that can also be solved using standard methods. The MPC formulation for SISO control problems described earlier can easily be extended to MIMO problems and to other types of models and objective functions (Lee et al., 1994). Tuning the controller is carried out by adjusting the following parameters:

- The weighting factor w.
- The move suppression factor λ.
- Bounds for the inputs and input moves.
- The input horizon (m) and output horizon (p).

See the review by Qin and Badgwell (1997) for details on commercial MPC packages.

EXAMPLE 16.3 MODEL PREDICTIVE CONTROL OF A CHEMICAL REACTOR

To carry out changes in the desired operating conditions a chemical reactor is to be controlled using MPC. The reactor is treated as a SISO system; the heat addition rate is the input, and reactor outlet concentration is the output. To design the controller, the system is subjected to a step change in the input, and the output is measured using a constant sampling interval of 1.0 min. Table E16.3 lists the values of the measured output (the response data have been normalized to have a final steady-state value of 1.0). The step response data follow the pattern shown in Figure 16.3. We will use Equation 16.1 to match the step response, with N equal to 70. Once the model coefficients of the response are determined, we can use a QP solver to find the response for a specific setpoint change given the horizons $m = 2$, $p = 4$ for the following three cases:

1. Unconstrained $u(k)$, $\lambda = 0$, $w = 1$
2. $40 \leq u(k) \leq 40$, $\lambda = 0$, $w = 1$
3. Unconstrained $u(k)$; λ is varied using a one-dimensional search (external to the MPC program) to find a good response that satisfies the input constraints in step 2.

Solution. For a given setpoint change you want a smooth, reasonably rapid rise to the new operating point with a small amount of overshoot before settling to the desired

TABLE E16.3
Step response for $\Delta t = 1$

Time	Step response	Time	Step response
1	0.000	36	0.940
2	0.009	37	0.946
3	0.033	38	0.951
4	0.067	39	0.956
5	0.108	40	0.960
6	0.155	41	0.964
7	0.203	42	0.967
8	0.253	43	0.970
9	0.303	44	0.973
10	0.352	45	0.976
11	0.399	46	0.978
12	0.445	47	0.980
13	0.488	48	0.981
14	0.529	49	0.984
15	0.568	50	0.985
16	0.603	51	0.987
17	0.637	52	0.988
18	0.668	53	0.989
19	0.697	54	0.990
20	0.723	55	0.991
21	0.748	56	0.992
22	0.770	57	0.992
23	0.791	58	0.993
24	0.809	59	0.994
25	0.827	60	0.994
26	0.842	61	0.995
27	0.857	62	0.995
28	0.870	63	0.996
29	0.882	64	0.996
30	0.892	65	0.997
31	0.903	66	0.997
32	0.912	67	0.997
33	0.920	68	0.997
34	0.928	69	0.998
35	0.934	70	0.998

operating point. In addition, the changes in the input variable (e.g., valve position for heat transfer medium) should not be too extreme during the transition. Although we do not place a hard limit on the changes in the input, this could easily be done. The step response model for $N = 70$ is simply the values of y for $k = 1$ to 70.

For this example, the controller design was carried out using the MATLAB Model Predictive Control toolbox, which includes a QP solver. Three cases were considered in the preceding problem statement.

1. The MPC controller that minimizes the variance of the output (minimum variance controller) during a setpoint change corresponds to the controller setting $w = 1$, $\lambda = 0$, and no bounds on the input. The response for this controller design for $m = 2$ and $p = 4$ is given in Figure E16.3 by the solid line.

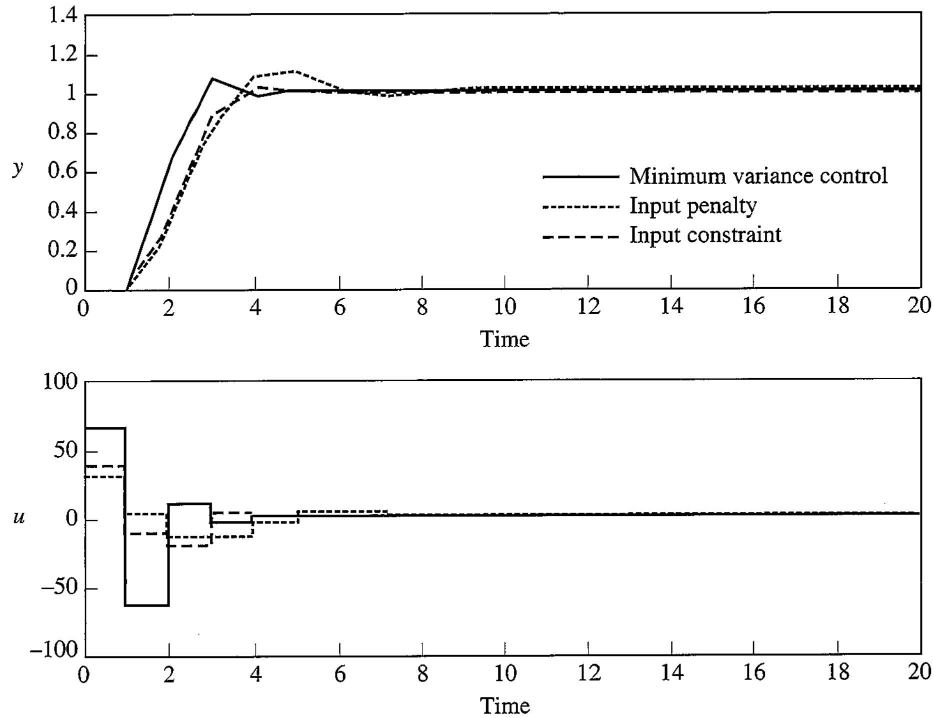

FIGURE E16.3
Comparison of the system behavior using three different model predictive controllers (a) minimum variance, (b) input constraint, (c) input penalty.

2. The input for most chemical processes is normally constrained, (e.g., a valve ranges between 0 and 100 percent open). An unconstrained minimum variance controller might not be able to achieve the desired input trajectory for the response. The controller design should take the process input constraints into account. The results of a simulated setpoint change for such a controller with bounds of -40 and 40 for the input and controller parameters $w = 1$ and $\lambda = 0$ is given by the dashed line in Figure E16.3.
3. An alternative method to limit the control action for a controller is to increase the value of the move suppression factor λ, penalizing the change in the input. The system response for small values of λ is close to the unconstrained minimum variance controller as expected, but it violates the constraints. With increasing values of the move suppression factor, however, the second term in Equation (16.2) becomes more important in the objective function, and control changes can correspondingly be limited to the range $-40 \leq u(k) \leq 40$. The dotted line in Figure E16.3 corresponds to a system with the controller setting $w = 1$, $\lambda = 0.01$, and no bounds on the input. Note that the response is much slower than in the direct constraint approach used in case 2.

The control actions in Figure E16.3 are influenced by the choice of the input and output horizon. For this example, all of the controllers had an input horizon of 2 and an output horizon of 4. In addition to w and λ, the two parameters m and p can be adjusted to improve the response. A selection of shorter horizons will result in more aggressive controllers.

Implementation issues

A critical factor in the successful application of any optimization technique is the availability of a suitable dynamic model. As mentioned previously, in typical MPC applications an empirical model is identified from data acquired during extensive plant tests. The experiments generally consist of a series of step tests, in which the manipulated variables are adjusted one at a time, and the tests require a period of 1–3 weeks. Details concerning the procedures used in the plant tests and subsequent model identification are usually considered to be proprietary information. The scaling and conditioning of plant data for use in model identification and control calculations can be key factors in the success of the application.

Integration of MPC and real-time optimization

Significant potential benefits can be realized by using a combination of MPC and RTO of setpoints that was discussed in Section 16.3. At the present time, most commercial MPC packages integrate the two methodologies in a configuration such as the one shown in Figure 16.4. The MPC calculations are imbedded in the prediction and controller blocks and are carried out quite often (e.g., every 1–10 min). The prediction block predicts the future trajectory of all controlled variables, and the controller achieves the desired response while keeping the process within limits.

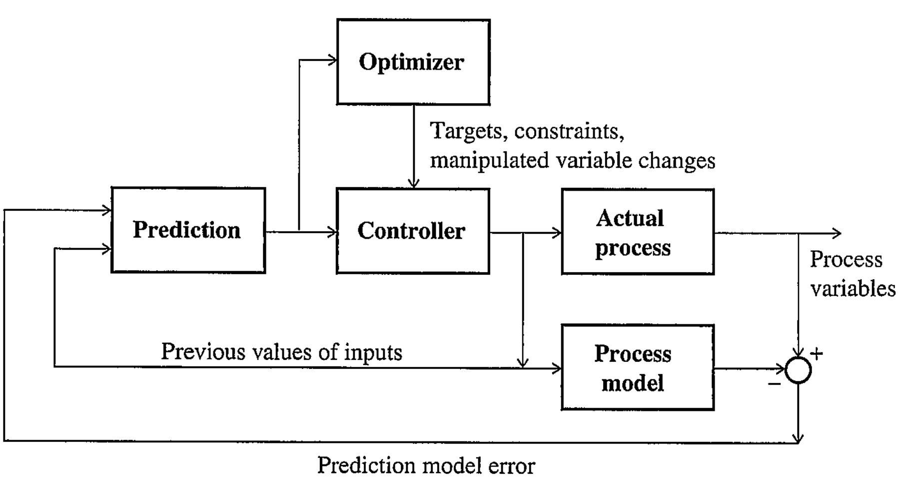

FIGURE 16.4
Diagram showing the combination of real-time optimization and model predictive control in a computer control system.

The targets for the MPC calculations are generated by solving a steady-state optimization problem (LP or QP) based on a linear process model, which also finds the best path to achieve the new targets (Backx et al., 2000). These calculations may be performed as often as the MPC calculations. The targets and constraints for the LP or QP optimization can be generated from a nonlinear process model using a nonlinear optimization technique. If the optimum occurs at a vertex of constraints and the objective function is convex, successive updates of a linearized model will find the same optimum as the nonlinear model. These calculations tend to be performed less frequently (e.g., every 1–24 h) due to the complexity of the calculations and the process models.

16.5 PROCESS MONITORING AND ANALYSIS

Measured process data inherently contain inaccurate information because the measurements are obtained with imperfect instruments. When flawed information is used for estimation of process variables and process control, the state of the system can be misrepresented and the resulting control performance is poor, leading to suboptimal and even unsafe process operation. Data reconciliation means the adjustment of process data measurements in order to force the data to agree in some sense with a model so that the estimates are better than the data. *Better* is usually defined as the optimal solution to a constrained least squares or maximum likelihood objective function. It is important to understand what is wrong with the values obtained by measurement and why they must be adjusted (Romagnoli and Sanchez, 1999). Data reconciliation can make the process data more useful for decision making and control by smoothing, eliminating outliers, and adjusting for bias and drift, thereby leading to better quality control, detection of faulty instrumentation, detection of process leaks, and increased profits. Computer-integrated manufacturing systems provide plant engineers direct access to extensive plant data as they are recorded. Automation of the data reconciliation computations is necessary to make use of the large amount of information available.

Suppose that the relationship between a measurement of a variable and its true value can be represented by

$$y_m = y + e \tag{16.8}$$

where y_m = measured value
y = true value
e = error

Measurements can contain any of several types of errors: (1) small random errors, (2) systematic biases and drift, or (3) gross errors. Small random errors are zero-mean and are often assumed to be normally distributed (Gaussian). Systematic biases occur when measurement devices provide consistently erroneous values, either high or low. In this case, the expected value of e is not zero. Bias may arise from sources such as incorrect calibration of the measurement device, sensor degradation, damage to the electronics, and so on. The third type of measurement

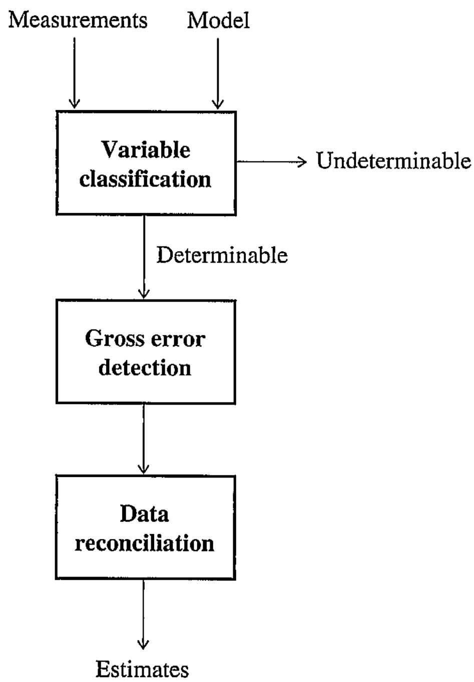

FIGURE 16.5
Steps for data improvement.

error is gross error and is usually caused by large, short-term, nonrandom events. Gross errors can be subdivided into measurement-related errors, such as malfunctioning sensors, and process-related errors, such as process leaks.

Typically, process data are improved using spatial, or functional, redundancies in the process model. Measurements are spatially redundant if more than enough data exist to completely define the process model at any instant, that is, the system is overdetermined and requires a solution by least squares fitting. Similarly, data improvement can be performed using temporal redundancies. Measurements are temporally redundant if past measurement values are available and can be used for estimation purposes. Dynamic models composed of algebraic and differential equations provide both spatial and temporal redundancy.

A simplified view of measurement data improvement techniques can be divided into three basic steps as shown in Figure 16.5. The first step, variable classification, involves determining which variables are observable or unobservable and which are redundant or underdetermined. Several authors have published algorithms for variable classification (Crowe, 1986; Stanley and Mah, 1981; Mah, 1990). Those that are undeterminable are not available for improvement. Next, all gross errors are identified and removed. Several methods proposed for gross error detection have been evaluated by Mah (1990), Rollins et al. (1996) and Tong and Crowe (1997). Data reconciliation concentrates on removing the remaining small, random measurement errors from the data. A key assumption frequently made during the reconciliation step is that the errors are normally distributed, but gross errors severely violate that assumption. If a measurement containing a gross error were allowed into the reconciliation scheme, the resulting estimates of the values of the variable would contain a portion of the gross error distributed among some or perhaps all the estimates (referred to as "smearing"). In practice, gross error detection and elimination are usually performed iteratively along with the final step—data reconciliation.

Historically, treatment of measurement noise has been addressed through two distinct avenues. For steady-state data and processes, Kuehn and Davidson (1961) presented the seminal paper describing the data reconciliation problem based on least squares optimization. For dynamic data and processes, Kalman filtering (Gelb, 1974) has been successfully used to recursively smooth measurement data and estimate parameters. Both techniques were developed for linear systems and weighted least squares objective functions.

The steady-state linear model data reconciliation problem can be stated as

$$\min f = \tfrac{1}{2}(\hat{y} - y)^T V^{-1}(\hat{y} - y) \tag{16.9}$$

subject to the model constraints

$$A\hat{y} - b = 0 \tag{16.10}$$

where V = variance–covariance matrix (usually diagonal)
y_i = measurement of variable i
$\hat{y}_i$ = reconciled estimate of variable i
A = matrix of linear constraints
b = vector of right-hand side terms in linear constraints

The optimal solution to this problem is

$$\hat{y}^* = [I - VA^T(AVA^T)^{-1}A]y + VA^T(AVA^T)^{-1}b \tag{16.11}$$

If the model includes nonlinear constraints, the problem can be solved using nonlinear programming (Chapter 8).

Several researchers [e.g., Tjoa and Biegler (1992) and Robertson et al. (1996)] have demonstrated advantages of using nonlinear programming (NLP) techniques over such traditional data reconciliation methods as successive linearization for steady-state or dynamic processes. Through the inclusion of variable bounds and a more robust treatment of the nonlinear algebraic constraints, improved reconciliation performance can be realized.

Extended Kalman filtering has been a popular method used in the literature to solve the dynamic data reconciliation problem (Muske and Edgar, 1998). As an alternative, the nonlinear dynamic data reconciliation problem with a weighted least squares objective function can be expressed as a moving horizon problem (Liebman et al., 1992), similar to that used for model predictive control discussed earlier.

The nonlinear objective function (usually quadratic) is

$$\min_{\hat{y}(t)} f(y(t), \hat{y}(t)) \tag{16.12}$$

which is subject to the dynamic model

$$h\left(\frac{d\hat{y}(t)}{dt}, \hat{y}(t)\right) = 0 \tag{16.13}$$

and inequality constraints

$$g(\hat{y}(t)) \geq 0 \tag{16.14}$$

This problem can be solved using a combined optimization and constraint model solution strategy (Muske and Edgar, 1998) by converting the differential equations to algebraic constraints using orthogonal collocation or some other model discretization approach.

EXAMPLE 16.4 STEADY-STATE MATERIAL BALANCE RECONCILIATION

Consider the process flowsheet shown in Figure E16.4, which was used by Rollins and Davis (1993) in investigations of gross error detection. The seven stream numbers are identified in Figure E16.4. The overall material balance can be expressed using the constraint matrix $\mathbf{A}\hat{y} = 0$, where $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix}$$

As a simple case, reconcile a single data set for the stream flows as follows:

$$\mathbf{y} = \begin{bmatrix} 49.5 \\ 81.5 \\ 85.3 \\ 10.1 \\ 72.9 \\ 25.7 \\ 50.7 \end{bmatrix}$$

Use the variançe–covariance matrix below as a measure of the variability (and reliability) of the stream measurements:

$$\mathbf{V} = \begin{bmatrix} 1.5625 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4.5156 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4.5156 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0625 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.5156 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.3906 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.3906 \end{bmatrix}$$

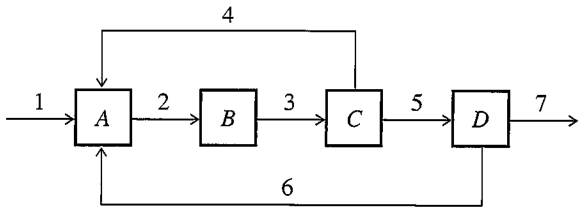

FIGURE E16.4
Recycle process network.

TABLE E16.4
Data reconciliation results

Stream number	True value (kg/min)	Measured value (kg/min)	Reconciled value (kg/min)
1	50.0	49.5	50.0
2	85.0	81.5	85.2
3	85.0	85.3	85.2
4	10.0	10.1	10.0
5	75.0	72.9	75.2
6	25.0	25.7	25.2
7	50.0	50.7	50.0

Solution. The reconciled results in Table E16.4 are obtained by solving the optimization problem with the process model as the only set of constraints. Because all constraints are linear, an analytical solution exists to the problem, as given in Equation 16.11. This results in an 89.6% reduction in the sum of the absolute error. Note that all reconciled values are positive and hence feasible. It is not unusual for some reconciled flow rates to go negative, in which case it is necessary to solve the problem using a constrained minimization code such as QP.

REFERENCES

Backx, T.; O. Bosgra; and W. Marguardt. "Integration of Model Predictive Control and Optimization of Processes." *ADCHEM Proceedings,* pp. 249–259, Pisa, Italy (2000).

Baker, T. E. "An Integrated Approach to Planning and Scheduling." In *Foundations of Computer Aided Process Operations (FOCAPO),* D. W. T. Rippin; J. C. Hale; and J. F. Davis, eds. CACHE Corporation, Austin, TX (1993), pp. 237–252.

Bryant, G. F. "Developments in Supply Chain Management Control Systems Design." In *Foundations of Computer Aided Process Operations (FOCAPO),* D. W. T. Rippin; J. C. Hale; and J. F. Davis, eds. CACHE Corporation, Austin, TX (1993), pp. 317–340.

Bunch, P. R.; D. L. Watson; and J. F. Pekny. "Improving Batch Manufacturing Process Operations Using Mathematical Programming Based Models." *FOCAPO Conference Proceedings, AIChE Symp Ser* 320, **94:** 204–209 (1998).

Camacho, E. F.; and C. Bordons. *Model Predictive Control.* Springer-Verlag, New York (1999).

Crowe, C. M. "Reconciliation of Process Flow Rates by Matrix Projection, Part II: The Nonlinear Case." *AIChE J* **32**(4): 616–623 (1986).

Forbes, F.; T. Marlin; and J. MacGregor. "Model Selection Criteria for Economics-Based Optimizing Control." *Comput Chem Eng* **18:** 497–510 (1994).

Garcia, D. E.; C. E. Prett; and M. Morari. "Model Predictive Control: Theory and Practice—A Survey." *Automatica,* **25:** 335–348 (1989).

Gelb, A., ed. *Applied Optimal Estimation.* The M.I.T. Press, Cambridge, MA (1974).

Karimi, I. A. *Refinery Scheduling.* CACHE Process Design Case Studies, I. Grossmann and M. Morari, eds., CACHE Corporation, Austin, TX (1992).

Ku, H. M.; and I. A. Karimi. "Multiproduct Batch Plant Scheduling." In *CACHE Process Design Case Studies,* I. Grossmann; and M. Morari, eds. CACHE Corporation, Austin, TX (1992).

Ku, H. M.; and I. A. Karimi. "Completion Time Algorithms for Serial Multiproduct Batch Processes with Shared Storage." *Comput Chem Eng* **14:** 1, 49–59 (1990).

Ku, H. M.; and I. A. Karimi. "Scheduling in Serial Multiproduct Batch Processes with Finite Interstate Storage: A Mixed Integer Linear Program Formulation." *Ind Eng Chem Res* **27:** 10, 1840 (1988).

Ku, H. M.; D. Rajagopalan; and I. A. Karimi. "Scheduling in Batch Processes." *Chem Eng Prog* **83:** 8, 35 (1987).

Kuehn, D. R.; and H. Davidson. "Computer Control." *Chem Eng Prog* **57**(6): 44–47 (1961).

Lasdon, L. S.; and T. E. Baker. "The Integration of Planning, Scheduling, and Process Control." *Chemical Process Control,* III. T. J. McAvoy; and M. Morari; eds. Elsevier, Amsterdam, Netherlands, (1986), pp. 579–620.

Lee, J. H.; M. Morari; and C. E. Garcia. "State Space Interpretation of Model Predictive Control." *Automatica* **30**(4): 707–717 (1994).

Liebman, M. J.; T. F. Edgar; and L. S. Lasdon. "Efficient Data Reconciliation and Estimation for Dynamic Processes Using Nonlinear Programming Techniques." *Comput Chem Eng* **16**(10/11): 963–986 (1992).

Mah, R. S. H. *Chemical Process Structures and Information Flows.* Butterworth Publishers, Stoneham, MA (1990).

Marlin, T. E.; and A. N. Hrymak. "Real-Time Operations Optimization of Continuous Processes." In *Chemical Process Control V. AIChE Symp Ser* 316, **93:** 156–164 (1997).

McDonald, C. M. "Synthesizing Enterprise-Wide Optimization with Global Information Technologies." *FOCAPO Conference Proceedings, AIChE Symp Ser* 320, **94:** 62–74 (1998).

Muske, K.; and T. F. Edgar. Chapter 6, "Nonlinear State Estimation." In *Nonlinear Process Control,* M. A. Henson; and D. E. Seborg, eds. Prentice-Hall, Englewood Cliffs, NJ (1998).

Pekny, J. F.; and G. V. Reklaitis. "Towards the Convergence of Theory and Practice: A Technology Guide for Scheduling/Planning Methodology." *FOCAPO Conference Proceedings, AIChE Symp Ser* **94:** 91–111 (1998).

Pike, R. W. *Optimization for Engineering Systems.* Van Nostrand Reinhold, New York (1986).

Puigjaner, L.; and A. Espura. "Prospects for Integrated Management and Control of Total Sites in the Batch Manufacturing Industry." *Comput Chem Eng* **22**(1–2): 87–107 (1998).

Qin, J.; and T. A. Badgwell. "An Overview of Industrial Model Predictive Control Technology." In *Chemical Process Control V, AIChE Symp Ser* 316, **93:** 232–256 (1997).

Richalet J. "Industrial Applications of Model Based Predictive Control." *Automatica* **29:** 1251–1274 (1993).

Robertson, D.; J. H. Lee; and J. B. Rawlings. "A Moving Horizon-based Approach for Least Squares State Estimation." *AIChE J* **42**(8): 2209–2224 (1996).

Rollins, D. K.; and J. F. Davis. "Gross Error Detection When Variance–Covariance Matrices are Unknown." *AIChE J* **39**(8): 1335–1341 (1993).

Rollins, D. K.; Y. Cheng; and S. Devanathan. "Intelligent Selection of Tests to Enhance Gross Error Identification." *Comput Chem Eng* **20**(5): 517–530 (1996).

Romagnoli, J. A.; and M. C. Sanchez. *Data Processing and Reconciliation for Chemical Process Operation.* Academic Press, New York (1999).

Schulz, C.; S. Engell; and R. Rudof. "Scheduling of a Multiproduct Polymer Plant." *FOCAPO Conference Proceedings, AIChE Symp Ser* 320, **94:** 224–230 (1998).

Seborg, D. E.; T. F. Edgar; and D. A. Mellichamp. *Process Dynamics and Control.* Wiley, New York (1989).

Smith, W. K. *Time Out.* Wiley, New York (1998).

Stanley, G. M.; and R. S. H. Mah. "Observability and Redundancy in Process Data." *Chem Eng Sci* **36:** 259–272 (1981).

Tjoa, I. B.; and L. T. Biegler. "Reduced Successive Quadratic Programming Strategy for Errors-in-Variables Estimation." *Comput Chem Eng* **16**(6): 523–533 (1992).

Tong, H.; and C. M. Crowe. "Detecting Persistent Gross Errors by Sequential Analysis of Principal Components." *AIChE J* **43**(5): 1242–1249 (1997).

SUPPLEMENTARY REFERENCES

Boddington, C. E. *Planning, Scheduling, and Control Integration in the Process Industries.* McGraw-Hill, New York (1995).

Gooding, W. B.; Pekny, J. F.; and P. S. McCroskey. "Enumerative Approaches to Parallel Flowshop Scheduling Via Problem Transformation." *Comput Chem Eng* **18**(10): 909–928 (1994).

Kondili, E.; Pantelides, C. C.; and R. W. H. Sargent. "A General Algorithm for Short-term Scheduling of Batch Operations—I. MILP Formulation." *Comput Chem Eng* **17**(2): 211–228 (1993).

Lee J. H.; and B. Cooley. "Recent Advances in Model Predictive Control and Other Related Areas." Chemical Process Control—V Proceedings, *AIChE Symp Ser* 316, **93:** 201–216 (1997).

Lee, Y. G.; and M. F. Malone. "Batch Process Planning for Waste Minimization." *Ind Eng Chem Res* **39**(6): 2035–2044 (2000).

Macchieto, S. "Bridging the Gap—Integration of Design, Operations Scheduling, and Control." FOCAPO-93 Proceedings, pp. 207–231, CACHE, Austin, TX (1994).

Morari, M.; J. H. Lee; and C. E. Garcia. *Model Predictive Control.* Prentice-Hall, Englewood Cliffs, NJ (in press).

Pantelides, C. C. "Unified Frameworks for Optimal Process Planning and Scheduling." In *Foundations of Computer Aided Process Operations.* D. W. T. Rippin; J. C. Hale; and J. F. Davis; eds., pp. 253–274, CACHE, Austin, TX (1994).

Papageorgaki, S.; and G. V. Reklaitis. "Optimal Design of Multipurpose Batch Plants—I. Problem Formulation." *Ind Eng Chem Res* **29**(10): 2054–2062 (1990).

Pekny, J. F.; and D. L. Miller. "Exact Solution of the No-Wait Flowshop Scheduling Problem with a Comparison to Heuristic Methods." *Comput Chem Eng* **15**(11): 741–748 (1991).

Perkins, J. D. "Plant-wide Optimization: Opportunities and Challenges." FOCAPO-98 Proceedings. *AIChE Symp Ser* 320, **94:** 15–26 (1998).

Shah, N. "Single and Multisite Planning and Scheduling—Current Status and Future Challenges." FOCAPO-98 Proceedings. *AIChE Symp Ser* 320, **94:** 75–90 (1998).

Shah, N.; C. C. Pantelides; and R. W. H. Sargent. "A General Algorithm for Short-Term Scheduling of Batch Operations. II. Computational Issues." *Comput Chem Eng* **17**(2): 229–244 (1993).

Shobrys, D. E.; and D. C. White. "Planning, Scheduling, and Control Systems: Why Can They Not Work Together." *Comput Chem Eng* **24:** 163–173 (2000).

Vision 2020 website: www.chem.purdue.edu/v2020.

Wright, S. J. "Applying New Optimization Algorithms to Model Predictive Control." Chemical Process Control—V Proceedings. *AIChE Symp Ser* 316, **93:** 147–155 (1994).