The Use of a Plant Capacity Model for Production Scheduling and Operations Analysis

BY

Caleb A. Dailey

B.S.E. IN MECHANICAL ENGINEERING
PRINCETON UNIVERSITY AT PRINCETON, NJ (1992)

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT AND THE DEPARTMENT OF
MECHANICAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREES OF

MASTER OF SCIENCE IN MANAGEMENT
AND
MASTER OF SCIENCE IN MECHANICAL ENGINEERING

IN CONJUNCTION WITH THE
LEADERS FOR MANUFACTURING PROGRAM
AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY
MAY, 1998

SIGNATURE OF AUTHOR_____
Sloan School of Management
Department of Mechanical Engineering
› MAY 1, 1998

CERTIFIED BY_____
DONALD ROSENFIELD
Senior Lecturer, Sloan School of Management

CERTIFIED BY_____
KENNETH RUSSELL
Professor of Materials Science and Engineering

CERTIFIED BY_____
STANLEY GERSHWIN
Senior Research Scientist, Department of Mechanical Engineering

ACCEPTED BY_____
ANTHONY PATERA
Acting Chairman, Department Committee
Department of Mechanical Engineering

ACCEPTED BY_____
LAWRENCE ABELN
Director of the Masters Program
Sloan School of Management

The Use of a Plant Capacity Model for Production Scheduling and Operations Analysis

BY

Caleb A. Dailey

B.S.E. IN MECHANICAL ENGINEERING
PRINCETON UNIVERSITY AT PRINCETON, NJ (1992)

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT AND THE DEPARTMENT OF
MECHANICAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREES OF

MASTER OF SCIENCE IN MANAGEMENT
AND
MASTER OF SCIENCE IN MECHANICAL ENGINEERING

Abstract

In a rapidly changing business environment, qualitative analyses are often not sufficient to make good business decisions. With growing globalization and increased merger & acquisition activity, managers are faced with more complexity than ever before. This environment was the context for a study into the development and implementation of a quantitative operations model.

This thesis describes a deterministic plant capacity model, including its structure, development, uses and implementation. It discusses the use of the model to guide production scheduling. The thesis presents a number of applications of the model for operations analysis, including product costing, multi-plant production location and capacity planning. Many examples are given regarding how the model can guide the user to make the right decision even when it is counter-intuitive.

The model uses station utilization as a measure of how efficiently a resource is being utilized. It relies on standard costing for its cost analysis. Despite well-founded criticisms of these types of analyses, from Eliyahu Goldratt and others, the thesis asserts that when carefully applied they can be both accurate and useful. It discusses a number of factors that must be considered to use these methods of analysis.

The thesis discusses many of the challenges that one encounters in implementing such a system. It asserts that complementary changes are often vital to successful implementation. These include improved communication practices, modifications to ordering practices and changes to work prioritization rules.

Thesis Advisors:
Dr. Donald Rosenfield, Senior Lecturer, Sloan School of Management
Dr. Kenneth Russell, Professor of Material Science and Engineering

Thesis Reader:
Dr. Stanley Gershwin, Senior Research Scientist, Mechanical Engineering

ACKNOWLEDGEMENT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

In a static business environment, managers can successfully rely on intuition and experience to make decisions. When the situation changes, however, these same managers require information based on data. This thesis documents a case where the managers of a medium-sized company were rapidly pushed into a new business environment confronted with decisions they had never made before. The outcome was a set of analytic models for use in production scheduling and in cost and capacity analysis, augmented by organization changes. These changes are based on many traditional operations management concepts, such as lean manufacturing, theory of constraints (TOC) and material requirements planning (MRP), tailored to the specific needs of the company.

1.1 Company Background

1.1.1 General Information

This thesis is based on a six and a half-month internship at Reading Tube Corporation. Reading Tube is a copper tube manufacturing company based in Reading, Pennsylvania. As of December 1997, Reading Tube had approximately 500 total employees at the company headquarters and four production facilities. 1996 revenues for Reading Tube were approximately $200 million on 110 million pounds of finished tubing.[1] This volume corresponds to about 20% market share for U. S. copper water tube.

---

[1] At the author's discretion, any information that could be considered proprietary to Reading Tube Corporation has be disguised. This may include, but is not limited to, strategies, costs, capacities and operating practices.

1.1.2 Products

Reading Tube produces many different types of copper tubing for many different applications. Below is a list of the major product categories, which comprise nearly all of Reading Tube's sales.

1.1.2.1 Water tube

The company's largest market is tubing for plumbing applications, shown in Figure 1.1. These products, collectively called water tube, are sold variety of diameters (1/4 inch – 6 inch), wall thicknesses and lengths (10 foot – 100 foot). Water tube is primarily sold to building contractors and construction firms, or to end users through direct retail outlets.
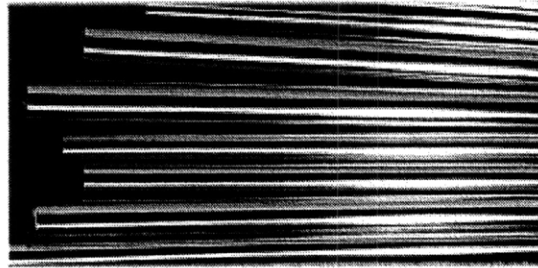
Figure 1.1 Water tube

1.1.2.2 Refrigeration coils

Another large market is that for refrigeration coils, which are used in refrigeration systems in buildings. These products, shown in Figure 1.2, are also made in a large range in sizes, ranging from diameters of 1/8" to 3" and lengths of 50 feet and 100 feet.
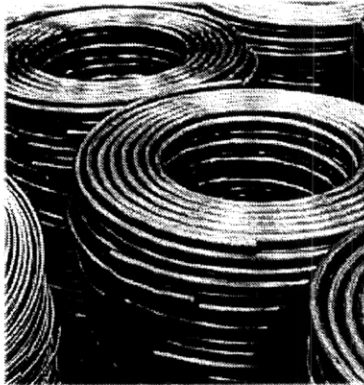
Figure 1.2 Refrigeration coils

1.1.2.3 Level wound coils

Copper tubing is also used in the manufacturing of air conditioners and other commercial products. These manufacturers of these products are generally supplied with level wound coils, which are tightly-wrapped coils ranging between 150 and 450 lbs. in weight. Examples are shown in Figure 1.3.
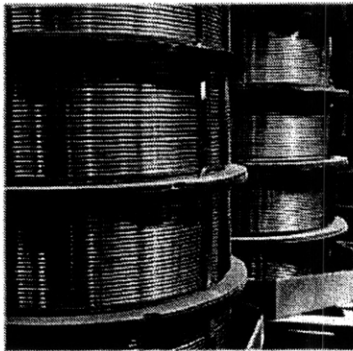
Figure 1.3 Level wound coils

1.1.2.4 Redraw

Another product is partially-processed tube, called redraw, that is sold to other copper tube manufacturers. Final processing is often performed on specialized finishing

equipment. Redraw is shipped as unfinished coils, 500 to 1000 lbs., with two to four draw passes.

1.1.3 The Copper Tubing Industry

Nearly all of the copper tube sold in for the United States market is produced by five companies. These companies have all existed for many years with few changes to the overall structure of the market. At times, one of the companies has tried to significantly increase their market share. The resulting price wars damaged all of the companies in the industry. As a result, the companies are averse to aggressive market moves, and therefore the structure of the market does not change quickly.

Water tube sales are generally linked to new building construction and therefore are highly cyclical. Markets for other products, such as level-wound coils, do not exhibit the same degree of cyclicality. In addition, in recent years, the water tube portion of the market has come under pressure from lower-cost substitute products, especially PVC tubing. Most of the growth in the copper tubing industry is anticipated to come from higher value-added products, such as finned or grooved tube, that increase energy efficiencies in heat exchange applications like air conditioners. These products often require highly-specialized manufacturing processes and a high grade of incoming copper material.

1.1.4 Recent Company History

Until September 1996, Reading Tube Corporation was a privately-held, independent company. At that time, Cambridge-Lee Industries, an industrial metals sales and distribution firm, acquired Reading Tube. Cambridge-Lee's parent company is a Mexican conglomerate called IUSA. IUSA has operations in a number of industries, including industrial metals, telecommunications and textiles. In addition, IUSA has two copper tube manufacturing mills in Pasteje, Mexico that produce similar products to Reading Tube.

1.1.5 Competitive Situation

At the time the internship began, Reading Tube was in a state of transition. The scope of the company's operations had grown from a single site to two sites, each having considerably different cost structures and production capabilities. The company's customer base had also grown significantly, opening up potential new markets and products.

Reading Tube's management needed to develop new long-term strategies to respond to the many new possibilities it faced. To aid in this process, reliable information on the company's costs and capacities was required to aid in decisions regarding how to proceed in the future.

1.1.5.1 Reading Needs

The primary need for the Reading facility was to control costs. Its unit labor costs were greater and its tooling and equipment was generally older than those the Mexican facilities. In addition, management believed that certain products could not be efficiently produced at Reading. The company needed to understand which products these were, and how to meet sales demand for these product, by either shifting production to Mexico, purchasing from third-party producer, or exiting these markets.

1.1.5.2 Pasteje Needs

With low labor costs and relatively new equipment, Pasteje's primary need was to understand and grow its production capacity. To maximize its overall throughput, Pasteje needed to understand its bottleneck resources and the effect that a changing product mix had on the bottleneck. Capacity expansions were possible, but since most involved relatively large expenditures, the investment case needed to be clearly understood.

1.1.5.3 Supply Chain Needs

Reading Tube – including both Reading and Pasteje – needed a plan to manage its new supply chain. Costs associated with the supply chain include both transportation costs and inventory holding costs. The base copper in copper tubing can add one to three times the value-added cost of processing the copper, significantly escalating inventory holding costs. An additional cost is the risk of holding a commodity like copper that can greatly fluctuate in value. Futures hedging can partially alleviate this risk, but it cannot perfectly match market fluctuations. The combination of these factors made it important to develop a strategy to limit the amount of total inventory that the company held in finished goods and in work in process (WIP).

1.2 Copper Tube Manufacturing

A number of general steps must be followed to manufacture copper tube. Below is the process used to produce high-volume seamless copper tube products like water tube, refrigeration coils and level wound coils.

1.2.1 Refining

The process begins with scrap copper or copper ore. The first step in the manufacturing process is to remove any impurities to produce a charge of copper to nearly 100% pure. Impurities include both copper oxides or other metals introduced from the incoming material.

1.2.2 Casting

After the copper charge has been refined, the copper is cast into long cylinders, called logs. The logs are then cut to standard lengths, called billets.

1.2.3 Extrusion

The extrusion process is shown in Figure 1.4 below. In this process, the billet is first heated to near the melting point of copper. A mandrel then pierces the billet from behind, creating the hole in the tube. Then a ram presses the billet between the mandrel and a stationary die, extruding a constant section. Finally, the extrusion is quenched to bring it back to ambient temperature for further processing.
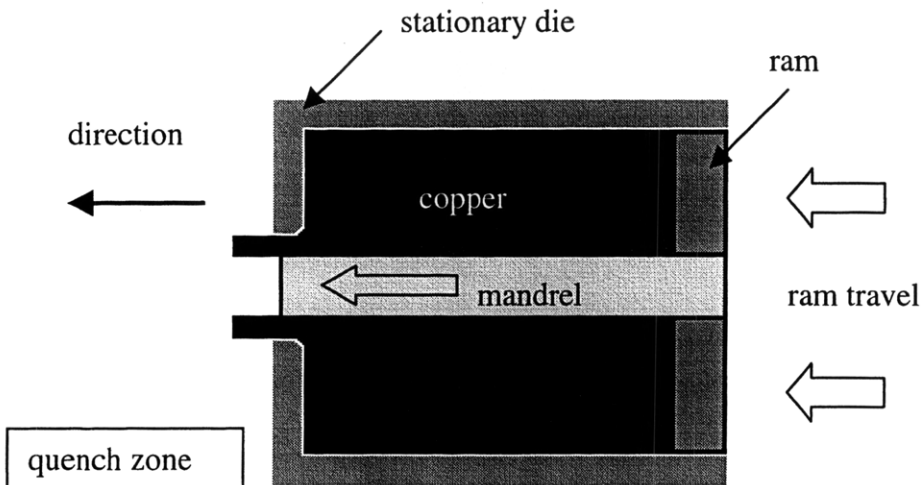
Figure 1.4 Tube Extrusion Process

1.2.4 Drawing

In the drawing process, the tube is cold drawn to the desired outer diameter and wall thickness dimensions. Depending on the size of the initial extrusion and the final tube size, multiple draw passes are required. Most products produced at Reading Plant 4 require four to eight draws. The list of processes to draw each tube size is called the draw schedule.

The tube drawing process is shown in Figure 1.5 below. In the drawing process, the draw eqipment pulls the tube between a stationary die, which controls the outer diameter, and a floating mandrel, which controls the inner diameter. First, an operator places the floating mandrel inside the tube. The operator points the tube to allow it to pass through the stationary die. Then the machine pulls it through the stationary die by its point. Once the process is complete, a shear cuts off the point and the tube is ready for further processing.
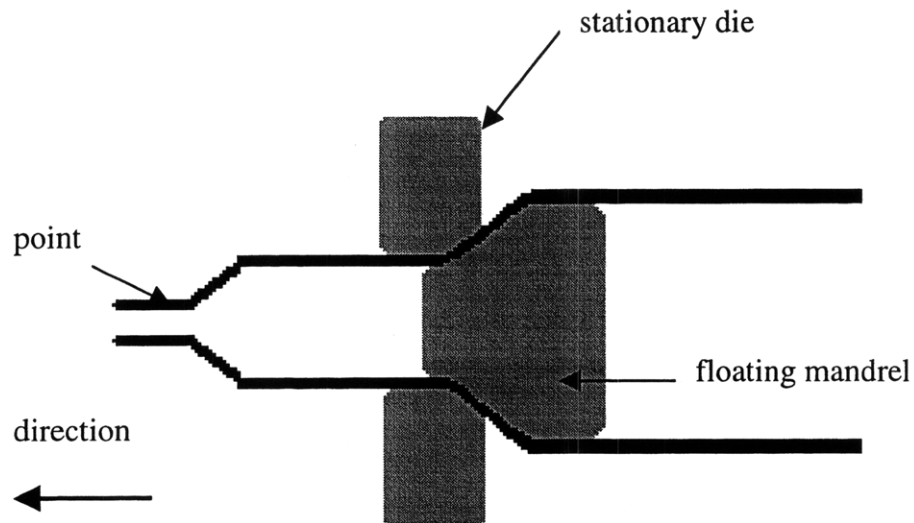
Figure 1.5 Tube Drawing Process

1.2.5 Finishing

In the finishing stage, a number of operations are performed on the tube to produce the final product. These operations include:

- Straightening or coiling
- Cutting to length
- Marking
- Bar coding
- Quality measurements and sorting.

Due to the variety of tasks, a large variety of stations are used for the finishing operations.

1.2.6 Annealing

The drawing and finishing operations work harden the tube to a hard, relatively brittle state. Certain applications require manipulation of the copper tube. In order to produce flexible tubing, the product is annealed by heating it at a specific combination of time and temperature. Annealing increases the grain size of the copper, thereby decreasing its modulus of elasticity.

1.3 Reading Plant 4

The facility that was the focus of the internship was Reading Plant 4. Compared to Reading Tube other facilities, Plant 4 is the largest in terms of employees and production volume and the most varied based on its production capabilities, and is consequently the most complicated in its operations.

1.3.1 Order Types

Water tube and refrigeration coils are generally made to restock inventory in warehouses. As a result, the company has the ability to use inventory to buffer production from demand and gain flexibility in what products to run at a given time. Level wound coils and redraw, which normally represent less that one-fourth of Reading's total volume, are generally made to fill orders with specific delivery dates. Orders are generally received well in advance of the required delivery date, allowing for flexibility in when to produce the product.

1.3.2 Process Flow

A process flow diagram for Plant 4 is shown in Figure 1.6 on the next page. Billets are refined and cast in a nearby refinery, and are initially held in a large billet rack. The billets are pulled to the extrusion press and then held in a smaller work in process buffer. The heaviest draw passes are performed on one of two drop blocks, which are identical. All remaining draw passes are performed on one of five spinners, which are nearly identical. Finishing is performed at one of six stations, all having unique product

capabilities, such as power and size, and operating characteristics, such as staffing requirements and throughput rates. Some products can be finished on only one line, some may be produced on a number of stations. Little or no inventory is generally held in front of the spinners or the finishing lines. Straight-length products (water tube) is directly transferred to shipping racks. Coils from the finishing stations, as well as large diameter coils produced in Reading Plant 3, are placed in a queue in front of the anneal furnace. At this station, the products are annealed and packaged. Finally, all products are shipped either to warehouses or directly to customers.
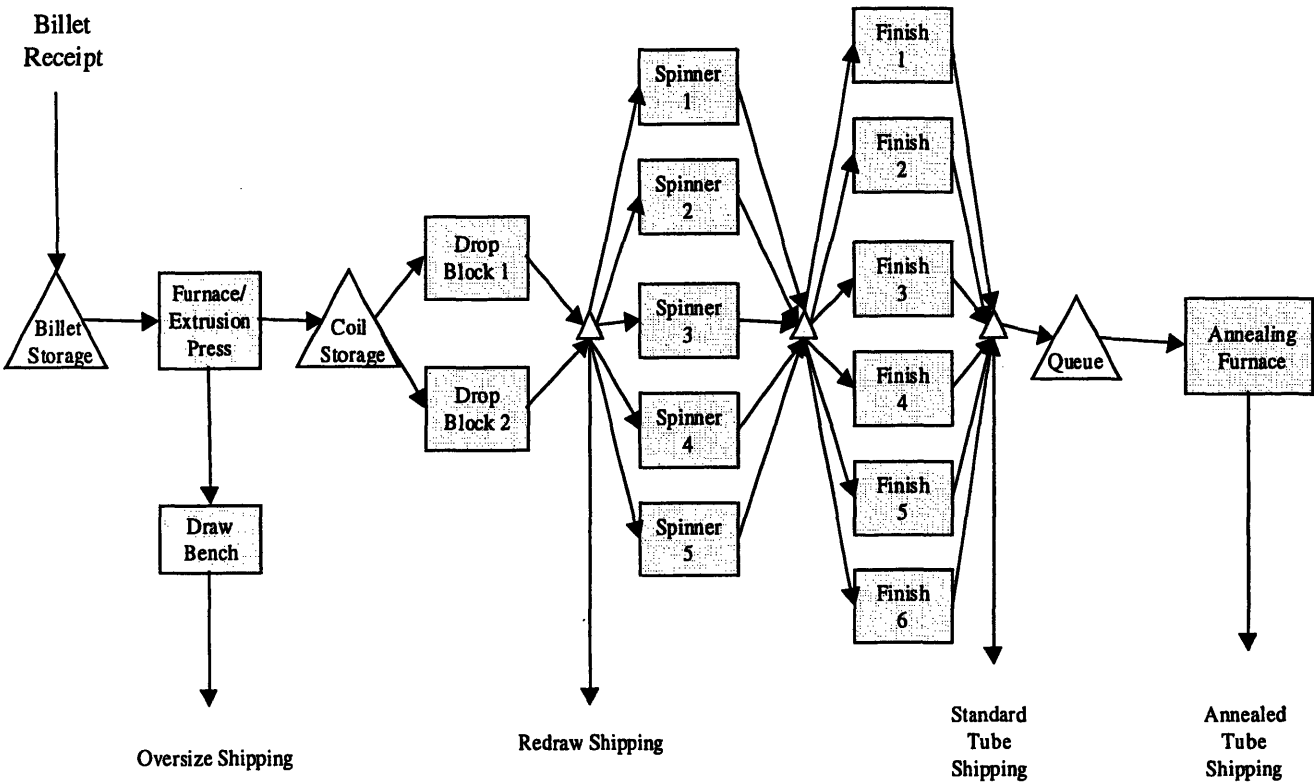
Figure 1.6 Plant 4 Process Flow Diagram

1.3.3 Process Constraints

The production of the plant is generally constrained by sales. As discussed before, in the copper tubing industry it is common to limit production to maintain high margins.

Within the plant, production is constrained at different stations, or multiple stations, depending on the specific operating conditions. Specifically, the bottleneck location is affected by product mix, staffing decisions and unexpected breakdowns. In most conditions, however, the throughput bottleneck is in the press and drop block stations. The annealing furnace can also be a production bottleneck, depending on the product mix. Bottlenecks are discussed in more detail in Section 4.

1.4 Planning and Scheduling Systems

A number of basic operations methodologies exist for planning and scheduling operations. These systems also generally provide a framework to analyze production capacities and costs. Because these needs are similar to those found at Reading Tube, each of the major systems is discussed for the purpose of comparing and contrasting to the system ultimately implemented.

1.4.1 Material Requirements Planning

Materials Requirement Planning (MRP) systems gained popularity in the 1970's. In a MRP system, each product is broken down into the component pieces and operations that will be required to produce the final assembly. Based on this breakdown, a master production schedule is created which allocates ordering of materials and sequencing of production. A number of sequencing rules exist for which product at a particular station should be given priority. As work is completed, it is placed in the inventory buffer of the product's subsequent process. This is generally done without regard for the subsequent process' buffer. As a result, MRP systems are often called push systems. [Nahmias, 1995]

1.4.2 Just in Time

The Just in Time (JIT) production system, sometimes called the *kanban* system, was developed at Toyota in the 1960's. In a JIT system, work is only performed when a subsequent station "calls" for it. For this reason, JIT systems are called pull systems. In many systems, communication between stations is accomplished by sending a card, called a *kanban*, to the previous station to request another unit of production. Inventory is therefore controlled by controlling the number of *kanban* cards in the system. [Nahmias, 1995]

1.4.3 Drum-Buffer-Rope

The Drum-Buffer-Rope (DBR) production methodology was developed by Eliyahu Goldratt. The system is founded on Goldratt's work in the Theory of Constraints (TOC), which proposes that every system has a bottleneck which limits the throughput of the entire system. Increasing the capacity of the bottleneck will increase the capacity of the system up to the point where another bottleneck arises; increasing the capacity of a non-bottleneck operation will not add to the capacity of the system. [Goldratt and Cox, 1984]

DBR is a pull system before the bottleneck and a push system after it. Unlike *kanban*, which treats every station equally, DBR focuses work flow on the bottleneck of the system. An objective of the DBR system is to hold a sufficient amount of WIP in front of the bottleneck operation to protect it from starvation. No inventory is justified in front of other stations, since their processing rate will not affect the rate of the system as a whole. When the bottleneck station completes one unit, a signal is sent (Drum) to the work release control to start work on another unit. As a result, a consistent amount of work in process inventory is held in the system prior to the bottleneck (Rope). Much of this work is held in inventory directly in front of the bottleneck (Buffer). Work beyond the bottleneck station is completed as rapidly as possible. [Alcalde, 1997] [Goldratt and Cox, 1984]

1.5 Thesis Outline

The basic objectives for this thesis and relevant background information have been presented in Chapter 1. Chapter 2 discusses the plant capacity model that was developed. Topics include the structure of the model, the process used to develop the model and the two specific interfaces that were developed: one for scheduling and one for cost and capacity analysis. Chapter 3 discusses the use of the model for scheduling. Chapter 4 discusses the use of the model for analyzing cost and capacity. After both Chapter 3 and Chapter 4, there are critical analyses of the systems that were developed, including advantages, disadvantages and comparisons to other solutions. Chapter 5 is focused on implementation issues. Chapter 6 presents an overall analysis of the project, including a brief discussion of future extensions of the project.

2 PLANT CAPACITY MODEL

The plant capacity model developed for Reading Plant 4 is the general tool for use in both production scheduling and analysis of the plant costs and capacities. This chapter will discuss the structure of the model (Section 2.1), the parameters used in the model (2.2) and how they were obtained (2.3), the two different interfaces that were developed (2.4) and how the model was validated (2.5).

2.1 Model Structure

The capacity model used in the internship has the ability to calculate both the throughput capacity of a system and to costs associated with that system.

The model operates using a number of linked spreadsheets in Microsoft Excel. Depending on the work environment, this type of model could work in a stand-alone program or in a database format.

2.1.1 Capacity

The primary output of the capacity model are the times that are required at each station to produce an order. From this information, users can determine which stations are bottlenecks and which have excess capacity. The information is also the basis for calculating production costs. To develop the required time, the model applies the order requirements to a set of parameters relating to the operations of the facility: throughput rates, downtimes, routings, etc. The basic structure is shown in Figure 2.1.
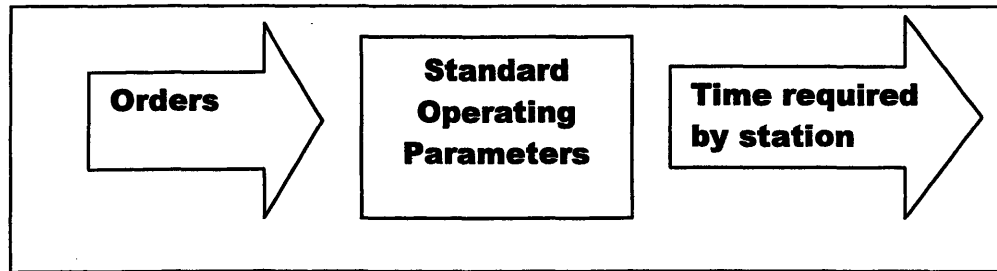
Figure 2.1 Time Calculation

The standard operating parameters are discussed in Section 2.2. The inputs and outputs to the model were specialized for two different applications discussed in Section 2.4.

2.1.2 Cost

Based on the expected times to produce an order, the model also calculates expected production costs. These costs are based on the variable cost rates for each station per unit of time. The total cost includes fixed costs that are independent of production volumes.
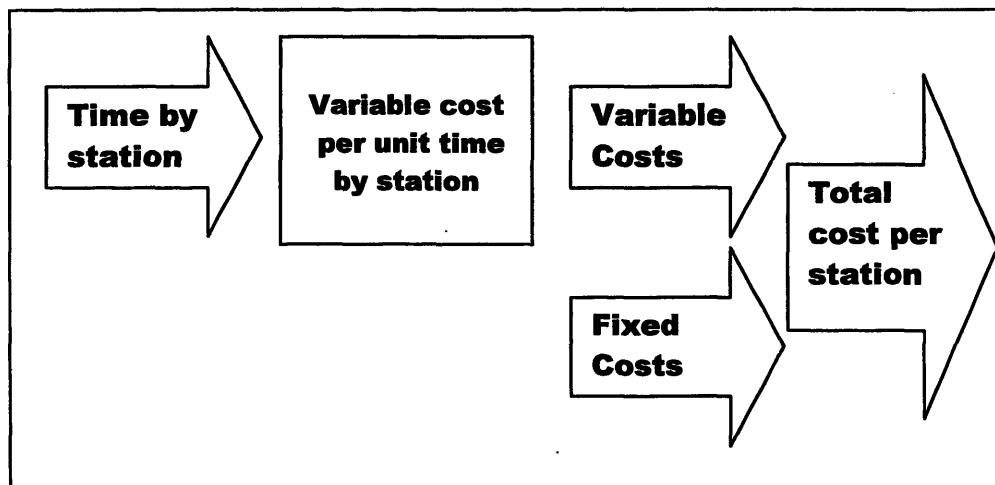
Figure 2.2 Cost Calculation

Algebraically, given the station times (t_i), variable costs rates per station (VC_i) and fixed costs (FC), the expected cost to produce an order is represented by the formula:

$$Cost = \sum_i \left(time_i * VC_i \right) + FC$$

Where applicable, additional costs such as raw materials, packaging and shipping must be added to produce the expected total cost to the producer.

2.2 Standard Operating Parameters

The model uses a number of standard operating parameters to calculate both capacity and cost, which are explained below. To limit complexity, the model treats all of the variables as being deterministic (not stochastic) in nature.

2.2.1 Capacity Parameters

The capacity model uses four standard operating parameters. These are:

- Throughput rates
- Downtimes
- Product routings
- Yields

Each is explained below. In general, the user can modify the standard parameters to represent changes to the operating conditions. This can be done to update the model to actual changes or to analyze the consequences of potential changes. In practice, however, the only parameters that should be regularly modified are the product routings (within the limitations given below).

2.2.1.1 Throughput rates

For each product at each station, the model uses specific processing rate measured in weight per unit time. Since the processes at Plant 4 are linear in nature, no efficiencies of scale are assumed.

2.2.1.2 Downtimes

Each station is assumed to have a specific downtime. Downtime is the result of a number of factors, such as:

- Material starvation
- Breakdowns
- Loading/unloading delays

The model assumes that downtimes are independent of orders or staffing levels. The user can vary the downtime parameters, however, to represent changing conditions.

2.2.1.3 Product Routings

The only processing flexibility at Plant 4 is between finishing stations. To represent this, the model has embedded in it the possible stations that are capable of producing each product, ranked by which is the most suitable. The user has the flexibility to shift production to a lower-priority station, but not to a station that is designated as incapable.

2.2.1.4 Yields

Product losses occur at nearly every step of the manufacturing process. These losses are a combination of process losses, those that are a direct result of the manufacturing process, and non-process losses, those from all other sources. Examples of process losses are:

- Outside scalping of the billet at the extrusion press
- Points that are clamped and then cut off in the drawing process
- Leftover ends at the finishing operation that are not of salable length

Examples of non-process losses are:

- Dimensional or metallurgical rejects
- Tubes which break during drawing, called breakers, that cannot be repaired
- Material handling damage

Process losses are reasonably consistent from product to product, since all follow nearly the same basic process. Non-process losses, however, can vary greatly based on the

product quality specifications and the details of the draw schedule. As a result, each product has its own yield, expressed as the ratio of incoming material to finished product.

Nearly all of the scrap can be salvaged, remelted and recast at the refinery. Therefore, the significance of scrap is not a loss of material; it is a loss of value-added work. By calculating yields, the capacity model captures this effect. For example, if a product has a 5% yield loss at each of the drop blocks, the spinners and the finishing lines, to produce 100 lbs. of product a total of:

$$100/(1-.05)^3 = 117 \text{ lb.}$$

must enter the drop blocks. Accordingly, the amount of time that must be allocated at the drop blocks must increase by 17%.

2.2.2 Cost Parameters

Many companies classify their costs as either variable or fixed. By definition, variable costs are dependent on production volumes, while fixed costs are independent of production volumes. Some costs which are not linearly related to volumes, called semi-variable costs, can be broken into fixed and variable components. [Zimmerman, 1993] The cost model treated all costs as either fixed or variable, a decision that is discussed in more detail in Chapter 4.5.2.

2.2.2.1 Variable Costs

Variable costs are those costs that are directly associated with the operation of a station. When the station is not in operation, all variable costs are zero. This is represented by costs directly associated with operating the station. In addition, certain indirect costs are allocated to the station based on some measure like direct labor hours. Direct labor is one of the most commonly used variable costs. [Zimmerman, 1993] In this analysis, other costs categorized as variable are:

- Indirect labor (material handling, support, etc.)
- Utilities

- General plant expenses
- Maintenance expenses

2.2.2.2 Fixed Costs

Fixed costs do not change based on production volumes. Depreciation is one of the most commonly agreed-upon fixed costs. [Zimmerman, 1993] Other costs that the model treats as fixed include:

- Supervisory labor
- Salaried Quality and Engineering labor
- Property taxes

2.3 Data Search

As in many manufacturing facilities, production data at Reading Tube's Plant 4 were found to be less than complete. Traditionally, scheduling was performed using a number of general rules of thumb. Examples include:

- The drop blocks process a batch every two hours.
- The annealing furnace processes 12 tons every shift.
- Every product yields three tons per spinner round (seven coils).

The schedulers knew that these rules were not optimal, but major problems could generally be avoided since generally there was reserve capacity in the plant and safety stock in finished goods inventories. For example, if over a few days the annealing furnace only processed eight tons per shift (instead of 12), extra shifts could be added on a weekend to compensate. If one product's yields were poor, the plant could short a shipment and make up the balance in the next batch.

To provide the most benefit to users, the capacity model required better data than was available. For example, every product does not have the same yield, nor does every station have the same processing rate. In addition, more accuracy was needed on all of the throughput rates. Therefore, a considerable amount of time was spent developing operating parameters that could confidently be used in the capacity model.

One decision made early in the internship was to use historical data as much as possible. The current rules of thumb lacked the detail and accuracy needed for use in the model. Another possibility was to use theoretical production rates. For example, one could calculate the time it should take to process a load of material on the spinners based on the number of passes required, the draw schedule (weight per foot), processing speeds, setup times and transfer times. Comparisons between the theoretical rates and historical rates often showed considerable gaps in performance. While plant personnel could generally explain these gaps using their understanding how the products actually ran, it was apparent that no simple calculation could adequately represent the production rates. The use of historical rates therefore assured better accuracy and helped build credibility in the operating parameters.

Another decision was to disaggregate the operating parameters as much as possible. For example, historical records showed that the press produced 20 good coils per hour. This value was broken into three component parameters:

- 30 total coils per production hour
- 25% downtime
- 11% scrap rate

The primary reason this was done was to help users better understand their system. Improvement can be accomplished by working on any one of the three parameters or on all three together. Users can determine the net result of a tradeoff (*i.e.* increased throughput, increased downtime). A side benefit was that these parameters allowed the model to calculate information like overall plant yield and incoming billet requirements.

2.4 User Interfaces

The capacity model was designed with two different user interfaces. One is the scheduling model, which is primarily intended for day-to-day production decisions. The second is the operations analysis model, which is intended for higher level decision-

making. Custom interfaces were designed to make the programs more convenient for the user by adding functions that are tailored to a specific need. Both operate with the same standard operating parameters, however, which is important to assure data consistency between the two functions.

2.4.1   Scheduling Model

The target users for the scheduling model are the plant personnel that actually make day-to-day decisions about staffing and prioritization. The model is used for the following purposes:

- Set weekly staffing levels for each station.
- Determine bottleneck stations.
- Provide information to guide supervisors on what products should be given priority and what are the optimal product routings.
- Determine the amount of material that must be pulled into the system to meet order requirements without overproducing.
- Track actual production relative to orders.

A schematic diagram of the scheduling is shown in Figure A.1 (Appendix A). The inputs to the model are the weekly production orders, consisting of both make-to-stock (MTS) water tube products and make-to-order (MTO) commercial products. For the water tube products, the orders consist of backordered products and orders for the current week.

The primary output is the expected utilization of each station, which is the expected required time for a station divided by the expected available time. The model also has a number of customized reports for each supervisory area: drop block/spinners, annealing and shipping. The use of the scheduling model is discussed in Section 3.

2.4.2   Operations Analysis Model

Compared to users of the scheduling model, the target users of the operations analysis model have a broader scope and a longer time frame. The model is used for the following purposes:

- Determine long-term capacity constraints
- Predict plant costs
- Determine costs for individual products

A block diagram of the operations analysis model is shown in Figure A.2. The inputs to the model are a set of standard product categories, which are applied to a historic mix of that type of product. Alternatively, the user can analyze the cost and capacity impact of an individual product.

The operations analysis model has two output screens: one for capacity and one for cost. The capacity analysis screen is similar to that for scheduling. It calculates the utilization rates of each station based on user inputs for the staffing. The objective of the cost analysis screen is to predict expected total product costs broken down by each component. First the calculated times are multiplied by standard variable cost rates. These variable costs are added to standard fixed costs to give the expected total cost for each station. These costs are then summed to give the expected cost for the total plant. Costs for incoming material from the refinery and for general overhead are also added to give the total product cost. Use of the operations analysis model is discussed in detail in Section 4.

2.5 Validation

Before the model could be used, validation was required to build confidence and credibility. This was accomplished by comparing the actual times and costs for a month to the times and costs predicted based on that month's actual production. Since no month's production is exactly to standard, it was not expected that the two would perfectly match. The validation results, shown below in Table 2.1, were judged acceptable to begin use of the model in scheduling and for studies on cost and capacity.

	% difference
Overall Cost (x1000)	5.2%
Overall Plant Yield	1.1%
Operation hours excluding downtime	
Press	2.0%
Blocks	-0.4%
Spinners	-0.4%
21/22 finish lines	9.8%
Conran/Reusch finish lines	-19.6%
LW	-21.1%
Operation hours including downtime	
Press	12.6%
Blocks	1.4%
Spinners	-3.6%
21/22 finish lines	7.8%
Conran/Reusch finish lines	-19.3%
LW	-13.7%

Table 2.1 Validation Results

3 SCHEDULING APPLICATIONS

At its most basic level, plant scheduling involves two factors: delivery and cost. To achieve delivery and cost objectives concurrently, plant management must effectively manage a number of tasks, including:

- Manage the utilization of assets to maintain sufficient capacity without excessive over-capacity

- Prioritize work processing at the decision-points in the operations to assure schedule completion while maintaining a balanced workflow.

- Plan for staffing, incoming material requirements and maintenance.

- Track order processing and completion.

One of the significant uses of the scheduling model is to provide information to the plant management to meet the goals listed above. This chapter discusses the use of the scheduling model (Section 3.1) and the advantages and disadvantages of this using this type of system (3.2).

3.1 Use of the Scheduling Model

The scheduling process takes place in three stages at Reading Tube. Before the week, schedulers create the operations plan. During the week, schedulers track progress and make adjustments as necessary. When the week is over, the performance is evaluated. The capacity model affects each of these functions, as described below.

3.1.1 Weekly Planning

Before the start of the week, the user of the system inputs the weekly order. The order consists of MTS water tube items, which are listed by item in a table, and MTO commercial products, which are listed in large lookup table based on the item number. The user also enters the expected backorder, which is added to the order to produce a preliminary listing of products that must be produced in the following week.

Based on this weekly order, the system determines how many expected hours are required at each station. The user enters the planned number of shifts and staffing of stations, including any expected overtime. The model calculates the expected number of hours required for each station, as well as the expected number of hours available. See Figure 3.1. Often the initial schedule does not create a feasible schedule due to insufficient capacity or severe imbalances in utilization, as in the left graph. The user has the ability to shift product routings to try to correct these conditions. If this cannot correct the problem, the plant management can negotiate with sales and distribution to change the order to create a feasible schedule. The result is the right graph. Based on the agreed-upon schedule, management creates the work schedules for direct labor and also determine the need for indirect labor and maintenance.
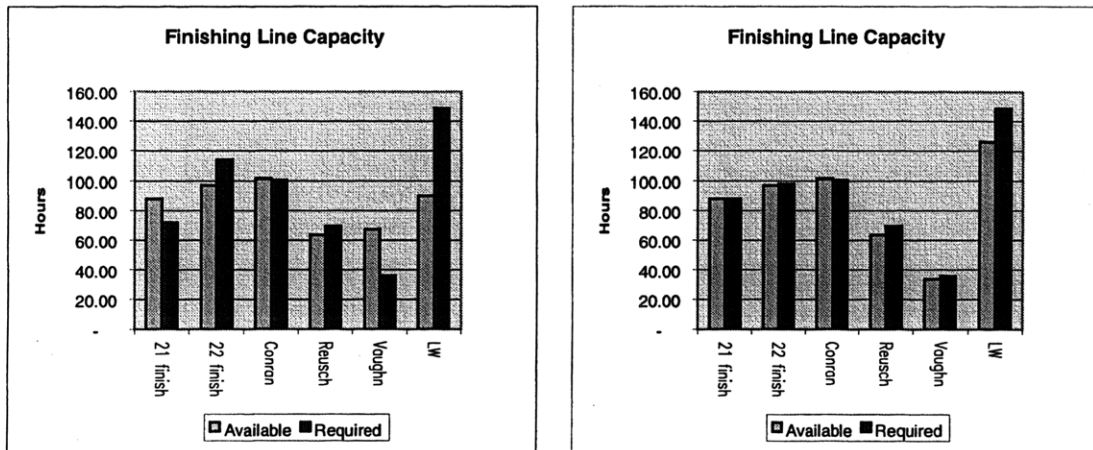
Figure 3.1 Before and After Capacity Graphs

At the beginning of the week, the actual backorder is input to the system. Unless there are any corrections, the backorder is directly carried over from the remaining items of the previous week. For each backorder, the scheduler enters the number of weeks late for use in prioritization. In addition, if there are any high-priority items, these are given a special designation. The final task is to update the work in process. Most of this work is automated using an Excel macro.

3.1.1 Daily Operation

During the week, the users of the system enter the number of units that are finished, annealed and shipped after each shift. The program automatically calculates the work in process in front of annealing and the finished goods inventory. As data are input, the model re-calculates the required hours and available hours, so that if unforeseen events take place, the users can adjust staffing or routings to compensate.

The model creates daily reports specifically designed for the supervisors of particular departments. For example, the annealing area report has the following columns for each item:
* priority status (backordered or high priority?)
* total annealing WIP units
* total annealing WIP weight
* the expected time to process

The reports also show total annealing by pounds and expected hours. Similar reports are created daily for supervisors in the blocks/spinners and shipping areas.

3.1.2 Reporting

At the end of the week, the scheduling model automatically calculates the delivery performance for the plant. Calculations are made for the fill rate for backordered items and the fill rate for the entire order. Another report shows the volume of each product ordered, shipped and in process. These reports are disseminated to production, sales and distribution at the beginning of the following week.

3.2 Discussion

The system developed over the internship for scheduling Plant 4 helps meet many of the scheduling needs outlined in the introduction to the chapter. The system calculates expected processing times and material requirements. Based on this information, direct and indirect labor can be scheduled. In addition, the system aids in order tracking and

encourages better on-time performance by calculating weekly order fill rate. One important criticism relative to this system comes the work of Eliyahu Goldratt, who stresses the important of the production bottleneck over all of the other stations. [Goldratt & Cox, 1984] The criticisms are addressed at the end of this section.

3.2.1 Better and More Accessible Information

The use of the capacity model for scheduling yielded a number of benefits over the existing system. First was an increase in accuracy in planning and scheduling. Even though many of the inputs to the capacity model came directly from the plant schedulers, the model could perform exact calculations based on the specific products ordered, while before general rules of thumb were required. Increased accuracy allows the plant to decrease waste associated with overstaffing, and improve order fill performance associated with understaffing.

A second benefit was that the system automates many functions that were previously done manually. The system automatically tracks inventories at the annealing furnace and in finished goods. In addition, the system provides an easy method to track production order and report out results. It communicates backorder positions to be used in the new prioritization system. All of these functions allow the plant managers to spend more of their time managing and less doing routine non-value-added work.

The ability to calculate order fill rates is another strength of the system. Fulfilling customer orders is an important function that production must perform. It is vital to reducing inventories and becoming more responsive to customer needs. While the calculation is simple (filled orders/total order), without the spreadsheet it is so tedious that is would not be done.

3.2.2 The Link to Distribution

While the scheduling system improves the plant scheduling process in many ways described above, fundamentally the communication process is unchanged. Many of these are opportunities for future improvements that could not be included in the scope of the internship.

The most significant point that the system does not address is the information gap between distribution and production. Orders are only placed once per week. The system looks like:
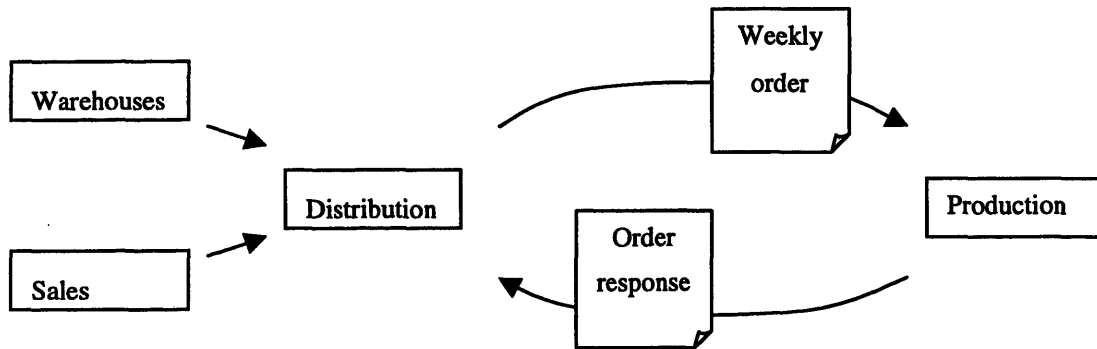
Figure 3.2 Weekly Order Communication Process

During the week, often major changes can occur in sales (i.e. large new orders), distribution (i.e. shipping problems) or production (i.e. breakdowns). To respond to these changing conditions, the only options are to (1) change the order or (2) wait until the next week's order. The first option causes operational inefficiencies and (perhaps) political difficulties, since the agreed-upon plan will need to be changed. The second option forces the system to have sub-optimal response times, which must be countered by holding higher inventory levels.

An improvement is to tie the needs of distribution to production scheduling more closely. In this case, orders would be made more frequently than once a week. The optimal

system is a true just in time (JIT) pull system, with orders updated real-time based on the needs of the sales and distribution. Such a system would look like:
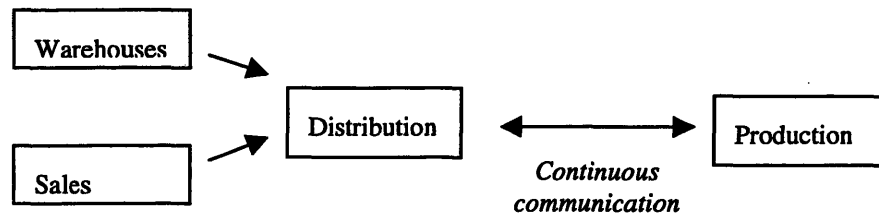
Figure 3.3 Continuous Communication Ordering

Order responsiveness increases and finished goods inventories can be dramatically cut. A move to this type of system requires a number of complementary changes to succeed, however, including:

- Improved information systems to facilitate communication
- Reduced changeover times/costs to allow for small batch production
- Multi-skilled workers and flexible machinery to allow for more responsiveness in production

These are longer-term issues that could not be executed in the time frame of the internship.

3.2.3   Criticisms from Goldratt

Many of the considerations made in this scheduling system seem to be in disagreement with the theory of constraints as described in Eliyahu Goldratt's book *The Goal* [1984]. Consider the following statements:

> "Some percentage of a non-bottleneck's time should be idle."
>
> "An hour saved at a non-bottleneck is a mirage."
>
> "A plant in which everyone is working all the time is very *in*efficient."
>
> [Goldratt & Cox, 1984]

In preparing the schedule for Plant 4, part of the scheduling process is to calculate efficiencies for each station. Based on this information, the scheduler tries to staff each station to full utilization – that is, that every station is staffed for the exact amount of time that should be required. While this methodology does conflict with Goldratt's theories, the conflict is justified based on specific operating conditions in Plant 4. In practice, most of the underlying principles of *The Goal* still apply.

One reason is that Plant 4 is not generally run at its maximum throughput capacity. The bottleneck is market demand. Since capacity is not an issue, the primary objective is cost reduction. By reducing the number of shifts that are worked, actual cost savings can be realized. In many cases, Goldratt reminds us, headcount reductions are not possible and there are no real savings. At Reading Tube, however, the situation allowed for actual headcount reductions and their corresponding savings.

The second reason is that in most cases the layout of Plant 4 protects the bottleneck from either blockage or starvation. The inventory buffer in front of the bottleneck protects it from starvation. Inventory after the bottleneck does not cause blockage (except in extreme cases). Overtime can work off inventory either in front of or after the bottleneck station. In this environment, Goldratt agrees that as long as the bottleneck is not compromised, efficiencies outside of the bottleneck should be maximized.

While Goldratt's principles may not be perfectly suited to Plant 4, they still have great value. Supervisors need to be aware that idle time at non-bottleneck stations is not necessarily cause for concern. Alternatively, work rules for giving the bottleneck priority for maintenance and repair must stress the significance of lost of production capacity at the bottleneck. Both of these concepts, which come from Goldratt, can only be accomplished by a combination of training and management support. One example of how these concepts manifested themselves is how non-bottleneck stations were staffed.

In practice, 10% to 20% excess capacity is scheduled for non-bottleneck stations. This protects the bottleneck from upstream or downstream problems.

4 COST AND CAPACITY ANALYSIS

A detailed understanding of a plant's capacities and costs is a vital component to developing long- and short-term operations strategies. Given this information, a company can:

- Focus capital investments or continuous improvement efforts to those areas of greatest leverage for increasing capacity and/or decreasing costs.
- Make Sales aware of products that have the highest contribution and those that have negative contribution.
- Develop realistic implementation plans to achieve strategic goals such as decreased costs, increased profits, or product diversification.
- Allocate production between multiple facilities to achieve the lowest possible manufacturing costs.

The capacity model was used at Reading Tube to meet these needs. This chapter examines four specific analyses: capacity planning and budgeting (Section 4.1), operational improvements (4.2), product costing (4.3) and multi-plant product locating (4.4). Section 4.5 discusses the strengths and weaknesses of this methodology.

4.1 Capacity Planning and Budgeting

The capacity model was used to evaluate the throughput capacity of each Reading Tube manufacturing plants. As mentioned before, copper tubing is generally sold by product categories, such as water tube, level wound coil and redraw. Standard product mixes exist for each category. To analyze the total capacity of each facility, projected sales volumes of each category are applied to these standard mixes as inputs to the capacity model. The model then predicts where capacity shortfalls are probable given the specified operating parameters. By applying the calculated times to the standard costs, an expected cost can also be obtained to produce the mix. At Reading Tube, these results were used to project costs and staffing requirements for each station in planning the annual budget. The individual products costs became the basis of a new manufacturing cost variance system.

The model can also demonstrate how a production bottleneck can shift when improvements are made in one station. For a simplified example, see Figure 4.1.

Hypothetical Production Capacity

Station	Capacity

Press: # 1

Drop Blocks: # 1, # 2, # 3

Spinners: # 1, # 2, # 3, # 4

tons/hour: 0 2 4 6 8 10 12 14
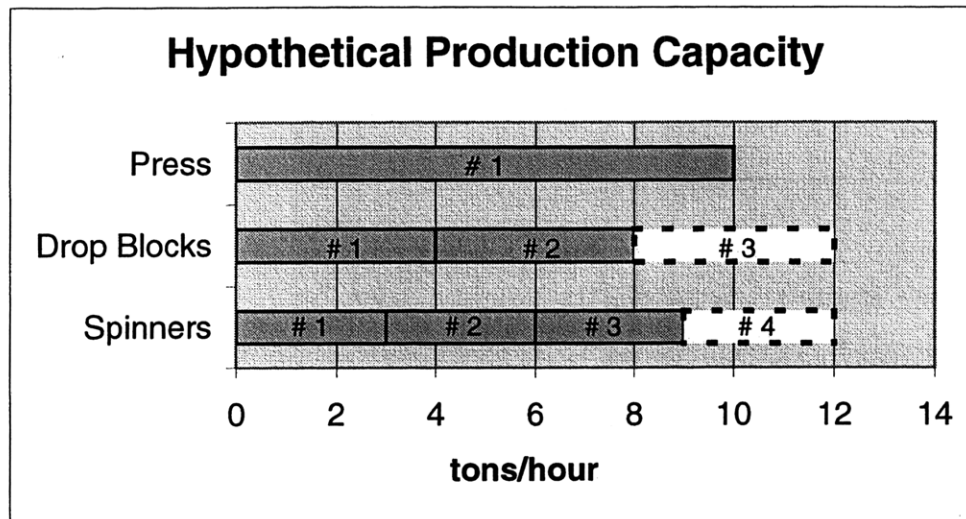
Figure 4.1 Capacity Planning Example

In this example, a hypothetical facility has one press, two drop blocks and three spinners. In the current configuration, the drop blocks are the production bottleneck, and the total throughput of the system is limited to the eight tons per hour that the drop blocks can produce. An investment in an additional drop block only increases the total throughput to nine tons per hour, however, since bottleneck shifts to the spinners. Adding another spinner only increases throughput to ten tons per hour because the press becomes the bottleneck. Without understanding the impact on the entire system, an analysis of one investment – such as one additional drop block – can produce a flawed result.

4.2 Analysis of Operational Improvements

The model also has the capability to comprehend the effects that changes in the standard operating parameters have on capacity and cost. Often these analyses involve a tradeoff between a gain in one area and a loss in another. The model can also show how small but

well-placed investments could have great leverage on lowering costs or improving capacity.

At Reading Tube, this analysis consistently led to focus on yield improvement. Because scrap copper can be nearly completely salvaged, yield improvements do not significantly reduce material costs. As a result, the cost impact of yield losses is not always obvious. When work in process is scrapped, however, a considerable amount of value-added work is lost. In addition to the cost savings, yield losses that are downstream from the bottleneck also directly detract from the production throughput capability of the facility.

Consider the following example, detailed in Table 4.2 on the next page. A finish line has a 10% scrap rate. If the machine speed were cut by 25%, many of these losses could be eliminated, so the scrap rate would drop to 5%. One way to analyze this is to only consider the finish line. Based on this simple analysis (Table 4.2, Scenario 1) this change cannot be justified. When the additional costs of melting, casting and drawing the scrapped material is included in the analysis, slowing the machines is clearly the lowest cost way to operate (Table 4.2, Scenario 2). The breakeven yield improvement is easy to calculate. Assuming the throughput capacity exists to operate the finishing lines at the lower rate (which also can be analyzed), the change should be made.

Scenario 1: Consider Only Finishing Costs

        Assume: Finish line costs $1/minute to operate

Current   Processing at 40 lb./minute, 10% losses

           Effective throughput = 36 lb./min

           Cost = ($1/min.)/(36 lb./min.) = $0.028/lb.

Proposed Processing at 30 lb./minute, 5% losses

           Effective throughput = 28.5 lb./min

           Cost = ($1/min.)/(28.5 lb./min.) = $0.035/lb.

Net Loss = $0.035 - $0.028 = $0.007/lb.

Scenario 2: Finishing and Incoming Material Costs

        Assume: Material cost for stock at the finish line: $0.40/lb.

Current   Cost of 1 finish lbs. with 10% losses = $0.4/0.9 = $0.444/lb.

           Total Cost = $0.444 + $0.028 = $0.472/lb.

Proposed Cost of 1 finish lb. with 5% losses = $0.4/0.95 = $0.421/lb.

           Total Cost = $0.421 + $0.035 = $0.456/lb.

Net Savings = 0.472 - 0.456 = $0.016/lb.

Table 4.1 Operational Improvement Comparison

4.3 Analysis of Product Contribution

The system was used to calculate the variable cost to produce each product. The net contribution of each product line is calculated by subtracting the variable cost from the sales price. This information was used at Reading Tube to understand which categories of products were most profitable. For example, one class of products was found to have a negative net contribution. Consequently, the company significantly reduced its exposure to this market. The decrease in these volumes was offset by increase in another class of products that was calculated to have a relatively high net contribution.

4.4 Multi-Plant Product Allocation

A total of four copper tubing plants (two in Reading, Pennsylvania and two in Pasteje, Mexico) are under Reading Tube's operational control. Significant overlap exists in the production capabilities that each of the plants can produce. In addition, the plants have significantly different cost structures. Two methods were used to give information about the lowest-cost method to allocate production: a linear program and a rank-order system.

4.4.1 Linear Program Solution

One methodology used to solve this problem is a linear program (LP). The parameters for the linear program are listed in Table 4.2. It was solved using Frontline Systems "Super Solver" in Microsoft Excel. There were approximately 60 water tube products, which correspond to the Copper Development Association's (CDA) standard product mix. In addition, product categories were included for products in other markets, such as level wound coil and redraw (see Section 1.1.2). To simplify calculations, products that had only one capable site were removed as decision variables from that LP. The impact of these products on station and plant capacities was included, however. The final optimization had approximately 400 decision variables and 80 constraints.

Constraints for the LP included demand (all products had to be produced) and capacity constraints. Two of the plants were constrained at the press, so that they could produce nearly any mix of products for which they were capable. Placing an upper bound on the

Type	Name	Formula
Variables	Volume of product j at plant k for market l	V_{jkl}
	Market demand of product j in market l	D_{jl}
	Time to produce product j at station i of plant k	t_{ijk}
	Variable cost to ship prod j from plant k to market l	$C_{ship, jkl}$
	Variable cost to produce product j at plant k	C_{jk}
Decision Variable	Volume of product j at plant k for market l	V_{jkl}
Constraints	Demand	$\sum_l D_j \leq \sum_k V_{jl}$
	Station time	$t_{avail, i} \geq \sum_j V_k * t_{ijk}$
	Plant maximum volume	$\sum_k V_k \leq V_{max, k}$
	Plant minimum volume	$\sum_k V_k \geq V_{min, k}$
	Non-negativity	$V_{jkl} \geq 0$
Objective function	Minimize total cost	$\sum_{jkl} (C_{jk} + C_{ship, jkl}) * V_{jkl}$

Table 4.2 Linear Program Parameters

total plant volume captured this constraint. In the other two plants, there were production bottlenecks within the plant. These were captured by the total time to produce. A final constraint that was used in some scenarios was a minimum volume for each plant. This was done to force certain levels of utilization to each plant -- despite the negative cost implications.

The minimization encompassed variable costs from both production and shipping. Production costs included labor, utilities, and expenses as discussed in Section 2.2.2. Unit shipping costs were estimated by dividing the total per shipment by the total volume per shipment. Note that the cost to ship from Reading to the US market is non-zero, as is the cost from Pasteje to the Mexican market. All fixed costs were excluded from the analysis.

The model could potentially be expanded to encompass product sales associated with changing market demand. In this situation, the objective function would be:

$$\text{Maximize:} \quad \sum_{jkl} P_{jl} - (C_{jk} + C_{ship, jkl}) * V_{jkl}$$

Due to the nature of the copper tubing market, however, the decision was made to focus on cost minimization given a sales mix. The reason is that the sales price, P_{jl}, is a function of the volume produced and the relationship is not linear. Certain volumes cause only minor changes in price, and then at certain threshold volumes a pricing war could result with dramatic consequences. By using a cost minimization over multiple sales scenarios instead, the user can develop a plausible production plan.

4.4.2 Rank Order Solution

While the linear program solution accurately comprehends all the cost and capacity issues in product placement, its use is difficult to operationalize. With hundreds of decision variables and numerous constraints, running the optimization on a weekly basis was not feasible.

A more user-friendly system is to create a rank-ordered list of products. At the top of the list are the products most suited to production at Reading and at the bottom were the ones most suited to Pasteje. With this list established, product is allocated to a plant by moving down the list until a target volume is reached.

The simplest list ranks the products by cost differential: $(C_{pas} - C_{rdg})$. Since this does not consider production capacities, the solution is accurate only when there are no production bottlenecks. At Reading Tube, however, in many scenarios the system one relevant bottleneck. In these cases, a refined solution is a rank-order list of products by cost differential adjusted for the bottleneck's time, that is:

$$\left(C_{pas} - C_{rdg}\right) \Big/ \left(t_{prod}/t_{avg}\right)$$

For the specific situation with one pre-defined bottleneck, this solution yields exactly the same results as the linear program. This is because a linear program with only one constraint, like this one, can be explicitly solved. For situations with multiple bottlenecks or for situations where the bottleneck is unclear, the system does not yield accurate results. In such cases, the linear program optimization is recommended.

4.5 Discussion

As described above, there are a number of potential applications for the capacity model. The use of the system at Reading Tube, and some general comments about its use, are given below. Since cost accounting has come under criticism recently [Johnson and Kaplan, 1991], there is a discussion about how cost accounting was used, including some pitfalls to avoid when using standard costs.

4.5.1 Use of the System

The primary benefit of the system was its ability to provide good information for a number of managerial decisions. The model provides a quantitative framework for

analysis that is immune to emotion or favor. As with any analysis, the information provided by the model is only as good as the data upon which the model is based. The information is also based on a number of assumptions relative to the operating environment and the market structure. When the analysis is challenged, it was productive to focus the challenge on the quality of the incoming data and the assumptions that were made. If the data were inaccurate or the assumptions flawed, these aspects of the model could generally be adjusted with minimal work. This process was helpful to build confidence and consensus in accepting the results of the model.

An important issue to consider is consistency between the complexity of a company's operations and the sophistication of the company's analytic tools. In a relatively simple system, the intuition of a skilled manager can often sufficiently capture enough of the situation to make a good decision. As complexity increases, however, the analytic tools much also be improved. Reading Tube faced such an increase in complexity, with the potential for new markets and changes in facilities. The model developed on the internship and described here suit the many of the needs of the new system. A more complex environment would certainly suggest the need for a more sophisticated model than that developed on the internship.

4.5.2 Cost Accouting Criticisms

Much of the field of cost accounting has come under fire in recent years. Johnson and Kaplan make the statement, "Contemporary cost accounting and management control systems are no longer providing accurate signals about the efficiency and profitability of internally managed transactions." [Johnson and Kaplan, 1991] Eliyahu Goldratt goes so far as to state, "If it comes from cost accounting it must be wrong." [Goldratt & Cox, 1984] Since this model makes extensive use of cost accounting, it is necessary to explain how cost accounting has been integrated into the model.

Proper assignment of costs as either fixed or variable is vital to making good decisions with cost accounting data. The time frame of the decision must be considered when making this assignment, since many costs are fixed in the short-term but variable if the

time frame is longer. Many of the applications for this model are long-term decisions, such as capital expenditure or product location. The assignment of costs should, therefore, correlate to this time frame. As a result, many costs that managers traditionally consider to be fixed, such as maintenance and supervisory labor, are treated as variable costs when making these decisions.

An inherent challenge to using a cost accounting model like the one described here is that one cannot easily and precisely match the time frame of the cost source to the time frame of the decision. Logistically, this presents difficulty since only those who have knowledge of both the operations and the accounting (few) can correctly match the operational time frame to the accounting time frame. We were able to control this during the internship because the company was small and the applications were relatively limited, but in a larger, more open company, this control against misuse of data will be difficult to implement.

The subjectivity involved with matching the time frame of the decision and the cost accounting means that the results the results are not highly accurate, with perhaps only two significant digits. This contrasts to traditional cost accounting systems that purport to have accuracy to five significant digits. [Johnson and Kaplan, 1991] Most managers will prefer two digits that can be trusted to five that are not reliable. A challenge in implementing this system is to properly convey the limitations of such a system, so that people do not use the model to discriminate between options in the case when it does not have useful accuracy.

Goldratt's concerns focus on how cost accounting treats bottlenecks. He states, "The actual cost of a bottleneck is the total expense of the system divided by the number of hours the bottleneck produces." [Goldratt & Cox, 1984] Under this methodology, since the bottleneck drives the output of the system, the only relevant processing rates are those at the bottleneck. This method was not used during the internship, for reasons explained below.

To understand how Goldratt's method compares with that used in the internship, consider the following example. At certain process has three stations: A, B and C. Each is staffed by one operator making $10 per hour. These are the only relevant costs to the system. Station A can produce 5 units per hour, while Stations B and C can produce 10 units per hour. The total cost of operating this system will be $30 per hour. Output is limited by Station A, the bottleneck, to 5 units per hour. Changing the throughput rates at Stations B or C will not increase production or decrease costs, and therefore their costs are essentially fixed. It is correct to state that the cost of the bottleneck is equal to the cost of the entire system, $30 per hour.

To compare, consider the same system with the ability to staff Stations B and C only to the level to which they are required: one-half hour for every hour that Station A operates. One way this could be achieved by holding inventory between the stations. The system works with only two operators, with the second operator moving between Stations B and C, first processing a batch through Station B, then through Station C, then back to B and so forth. Alternatively, Station A could operate one shift on its own, building inventory, then all three stations would work a second shift, with Stations B and C working off the backlog. In either case, the cost of this system is only $20 per hour. Furthermore, the costs of the Stations B and C are no longer fixed relative to the throughput of the bottleneck. Therefore, it would not be correct to state that the cost of the bottleneck is equal to that of the entire system ($20 per hour); instead, the cost of each station is $10 per hour.

While it was not used on this internship, an interesting methodology uses linear regression to help reduce some of the judgement associated with distinguishing fixed and variable costs. Total cost is plotted on the y-axis and working time is plotted on the x-axis. In this case, the y-intercept of the regression is the fixed cost component, while the slope of the line is the variable cost rate. [Strimling, 1996] The main difficulty in implementing such a system is that of data. Historical data is full of many undocumented events so that the confidence intervals on the two values are too wide for practical use. In addition, it is generally not feasible to perform a controlled experiment to obtain more

accurate data. Instead of the Strimling system, the costs used were based on general classifications of costs into either fixed or variable categories.

5 IMPLEMENTATION

To successfully introduction this system, it was important to understand Reading Tube's culture and business practices. This chapter discusses the keys to implementing of the scheduling system, which includes both employee support for the system (Section 5.1) and complementary changes to the business practices (5.2).

5.1 Employee Support

Throughout the internship, it was important to develop and maintain organizational support for the new system. This was addressed at both the management and shop floor levels.

5.1.1 Management Support

The coordinated support of Reading Tube's managers was critical to the success of the project. Before the project was officially underway, all of the relevant functional managers were consulted and a consensus was reached regarding the objectives for the internship and the resources were required for success. Interestingly, much of the support for the project came peripheral functions like Sales and Finance, who desired more accurate information from Production to better plan for sales and capital requirements. Management's commitment to the project was clearly communicated throughout the organization through both memos and meetings.

Communication between functional areas was also an important factor to first implement and then manage the scheduling system. To address this need, regular weekly meetings were established to bring together functional leaders from Sales, Distribution and each of the plants. The managers share information and discuss issues that require coordination or support. Examples of specific issues that are regularly discussed include:

- The need for overtime or expediting

- Prioritization of production or shipping
- Order placement between different plants
- Inventory levels in the plant and in the warehouses

Because the meetings are held in the middle of the week, the group also reviews the production statistics from the previous week, the status of the current week and the plans for the following week.

5.1.2 Shop Floor Support

Before the implementation of the capacity model, scheduling was led by a few of the senior plant supervisors. Their decisions were based off general rules of thumb for each station, modified to the specific conditions that they knew of in the plant. An important implementation issue was the role that these employees would have after the introduction of the scheduling system. Systems exist that provide information to allow the human schedulers to make the scheduling decisions (a prescriptive model), as do systems that automatically direct instructions to each work station, removing the human schedulers from the process (a proscriptive model). After many conversations regarding this issue, we determined that the objective for the internship was to develop a prescriptive model. The primary reasons for this decision were:

- The complexity of the operations. The human schedulers take into account many subtle issues when they plan production, ones that could not be realistically comprehended given the tools available for this project. For example, if a spinner station is causing breakers on certain products the supervisors adjust the flow to compensate. Similarly, if a spinner operator has a back strain the supervisors try to divert products that require heavy lifting from that station. Given the goal of an optimal production schedule, these issues are equally as important as purely quantitative issues like station utilization and supply chain inventories.
- Human and cultural issues in the plant. A proscriptive system was viewed as a threat to the supervisors who had traditionally scheduled the plant. On the other hand, a prescriptive system was viewed as a tool to allow these supervisors to do their jobs better. Everyone knew that the support of these people was vital to successfully developing and implementing any scheduling system.

While the ultimate goal may be a sophisticated proscriptive system, a consensus was reached that this goal is best achieved in steps, not all at once. A computer-based prescriptive model was an essential step to understand and eventually overcome both the complexity and cultural issues discussed above.

5.2 Complementary Changes

Good information is only one part making good decisions. Managers react to the business system in which they work, including how they are measured, who they trust and respect, and what their personal goals are. The internship encompassed creating a business environment in which mangers make the right decision for the company when given good information. Three specific issues that were addressed are changes to the product ordering system, the prioritization methodology and the performance measurement system.

5.2.1 Product Ordering

Historically, Distribution used a number of different ways to order water tube from Plant 4, including:
- Biweekly orders
- Weekly orders
- Daily priority lists based on inventory level reorder points

A tradeoff that must be made in selecting a system is that between operational efficiency and order responsiveness. Two-week orders are good for efficiencies because they enable larger batch sizes and greater scheduling flexibility. The drop in responsiveness, however, must be countered by either increased finished goods inventories or increased expediting, which both increase costs. Daily priority lists have the potential to increase responsiveness, but this can only be accomplished by decreased operational efficiency. In addition, the plant schedulers still have flexibility to choose which products to produce because the lists do not consider the plant's capacity. Since managers know that small batches and difficult-to-make items hurt production efficiencies, historically expediting was still necessary to produce these items.

To analyze what type of system to use, two factors must be considered: feasibility and cost-effectiveness. Feasibility is the capability of the ordering system to meet the customer needs. Given enough finished goods inventory, any system is feasible. To produce using a pull system (just-in-time), however, manufacturing response time must be considered. The relevant comparison involves comparing the expected lead time by the end customer, t_{lead}, to the amount of time to respond to the order, t_{res}. The response time is composed of time to manufacture (which includes order processing and scheduling) and time to ship. If

$$t_{lead} \geq t_{res}$$

then the system is feasible to ship just-in-time. On the other hand, if

$$t_{lead} < t_{res}$$

the system cannot ship just-in-time. A finished goods inventory buffer is required. In a stochastic world, some buffer is still required, but the distinction above is a strong indicator of how to move towards a lean system. [Womack and Jones, 1996]

The second issue is cost-effectiveness. For this analysis, traditionally the economic order quantity (EOQ) model is used. This model strikes a balance between inventory holding costs (h) and setup costs (K). In this model, the optimal order amount is given by the formula:

$$Q^* = \sqrt{\frac{2K\lambda}{h}}$$

in which λ is the demand rate. The optimal lead time is then given by Little's law:

$$T^* = Q^*/\lambda.$$

[Nahmias, 1993] Although the formula is simple, traditionally many of the variables have been improperly defined. For example, the setup cost should only reflect the incremental costs incurred by a setup. If the setup uses excess capacity, no additional

costs are incurred and the setup cost is zero. [Goldratt and Cox, 1984] As a result, the EOQ and the response time are zero – the theoretical extension of a JIT system.

Karmarker [1987] uses queueing models to show the relationship between capacity losses associated with setup time and inventory holding costs. As mentioned above, most direct costs associated with setups do not result in a negative cash flow and are therefore not included. In this simplified model, the optimal batch size can be shown to be:

$$Q^o = \frac{2\lambda\tau}{1 - \lambda/P}$$

in which τ is the setup time per batch and P is the machine processing rate. Using a M/M/1 queueing model, the average time is:

$$T(Q^o) = \frac{2\tau\left(1 + \lambda/P\right)}{\left(1 - \lambda/P\right)^2}$$

Note that setup costs do not explicitly enter this formula; setup times do. Additionally, this formula shows the effect of machine utilization (λ/P) on the optimal batch size. As utilization approaches one, both the optimal batch size and the corresponding lead time grow to infinity; as utilization approaches zero, the optimal lead time is equal to just two times the setup time. [Karmarker, 1987]

Another common mistake in the use of the EOQ model regards holding costs, which much somehow quantitatively capture not only the cost of tied-up working capital, but also costs of storage, obsolescence and undiscovered defects. Many proponents of lean manufacturing assert that these costs lead to an implicitly higher cost of holding inventory, therefore justifying smaller batch sizes. [Womack *et al*, 1990]

The complexity and subjectivity of this analysis makes it generally unreasonable to talk of a true "optimal" system for production ordering and scheduling. [Karmarker, 1974] One consistent theme throughout this research is a trend towards smaller batch sizes. As a result of this analysis, the company made a move from a two-week to a one-week order

system. The new system is closer to a pull system, reducing the need for finished goods inventory to meet demand drops. While the number of setups increases, many of these are performed using excess operator time so the net cost impact is lower than expected. In addition, a minimum order quantity was established for all orders so that low-volume products are produced only in efficient quantities. While most of the high-volume products are produced every week, the lower-volume lines are produced less frequently.

5.2.2 Prioritization Methodology

In Plant 4 there are three major inventory buffers (see Figure 1.3): one in front of the press, one in front of the annealing furnace and one of finished goods awaiting shipment. At each of these buffers the supervisor must decide how to use his or her resources. At the start of the internship, the *ad hoc* rules that were used for determining what items should be given priority were:

1. Expedited items based on the needs of Distribution and Sales
2. Larger lot sizes and easy-to-process items
3. Smaller lot sizes and difficult-to-process items

The overriding focus was local efficiency in operations. Most major problems could be avoided by expediting priority items. Expediting, however, often caused large inefficiencies in operations due to unplanned machine setups and improper staffing to run certain products.

The system dynamics model shown in Figure 5.1 demonstrates the situation. Management stresses efficient production, generally be measuring the number pounds of production per hour. This action allows quick gains in cost when the most efficient products are produced. Eventually, imbalances in inventory cause expediting, which increases costs. The result is a long-term situation characterized by chronic expediting and overall under-performance.
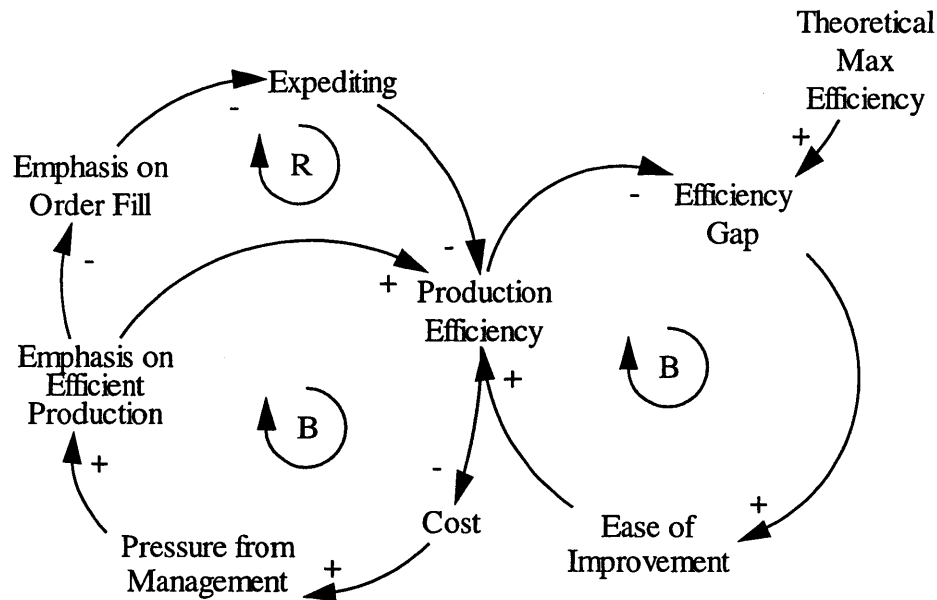
Figure 5.1 System Dynamics Model of Initial Prioritization Rules

To compensate for these issues, the scheduling system that was introduced used the following rules:

1. Expedited items based on the needs of distribution and sales

2. Backordered items, starting with those that are the most overdue

3. Items in the current week's order

4. All other items

The critical change in practice was to identify and prioritize the backordered items. The scheduling model automatically communicates this information to the appropriate supervisor. By shifting the focus to completing the backorder first, expediting was significantly reduced and overall efficiencies were improved. A small amount of expediting, due to significant unforeseen events in sales and distribution, cannot be eliminated. The majority, due to improper prioritization, can be eliminated using the system described above.

The new system dynamics model in Figure 5.2 shows the situation with the new prioritization system. In this case, management's primary emphasis is order fill rate. This system will improve to the point that the theoretical physical limitations on the system hold it back.
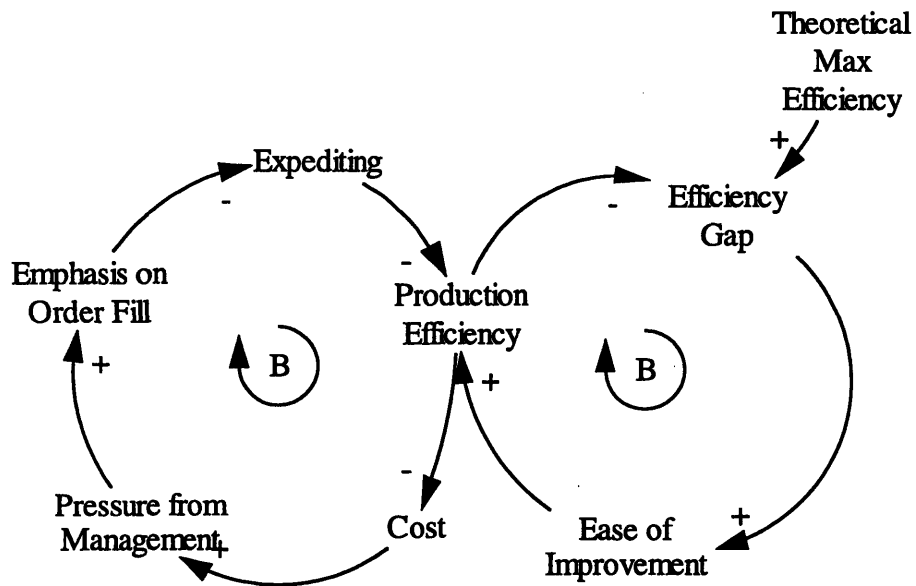
Figure 5.2 System Dynamics Model of Modified Prioritization Rules

5.2.3   Performance Measurement[2]

While upper management support is important, most operational improvements are actually made on the shop floor. The machine operators and first-line supervisors must be aware (1) what issues are important to management and (2) how they are doing relative to these issues. At Reading Tube Plant 4, plant personnel now track a number of important performance statistics on a weekly basis. Plant-wide metrics that are tracked include:

- Overall volume produced

- Plant yield

- Order fill rates

In addition, a number of measurements show the performance of specific stations:

- Station downtimes

- Station throughputs

- Station scrap rates

The production statistics are posted on the company's computer network so that anyone in the company can review them. In addition, the plant manager makes a point to review these numbers in detail at his weekly staff meetings.

This process allowed the supervisors to lead important changes to improve overall performance. Figure 5.3 shows the actual data from six charts, including four-week moving averages to highlight trends. The point of showing the graphs is to demonstrate different trends. In some metrics improvements were made (#1, 3, 5), while in others there has been little or no progress (#2, 4). One showed worse results (#6).

These graphs show that in some cases improvement may be relatively easy, while in others it may be difficult. Both physical barriers or human issues can limit improvement. Regardless, without monitoring any sustained improvement is difficult to achieve.
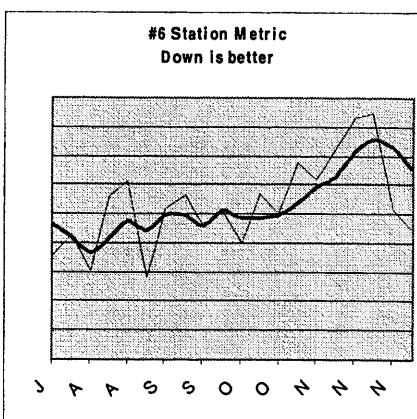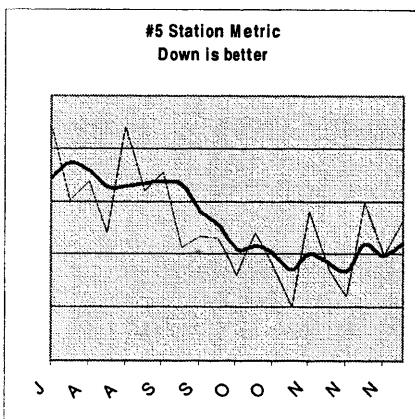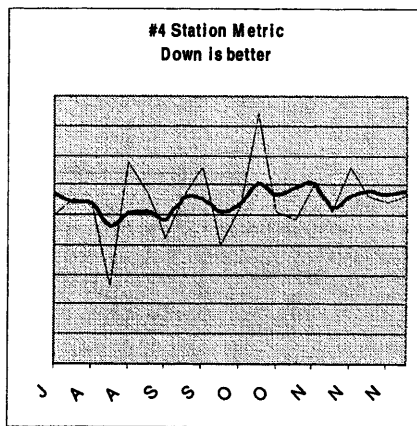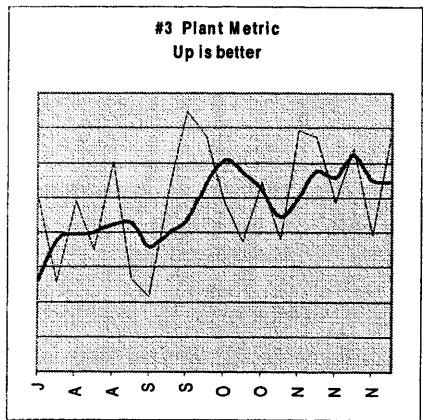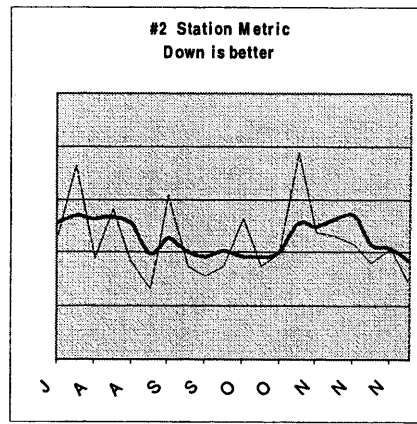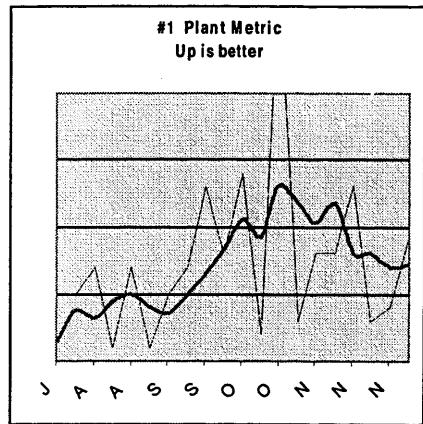
Figure 5.3 Performance Measurement Examples

6 CONCLUSIONS

The transition from qualitative to quantitative decision-making is a difficult one. Someone must develop and validate the quantitative models and train the new users. Often business practices must be modified to support the new system. The developers of such a system take on this work because they believe that the company can enjoy financial gains with the new system. The system described in this paper has the potential for such gains by assisting its users to decrease costs and to increase revenue.

The model assists Reading Tube's employees to cut costs in a number of areas. The plant can more accurately staff the plant to reduce labor costs. The company can reduce its inventory levels, and corresponding holding costs, by shipping orders on time more consistently. The company can identify and eliminate unprofitable product lines.

The model can also guide management about how to intelligently grow the company. Users can analyze and compare capital expenditure alternatives. The costs of new products and new markets can be estimated to determine if expansion is prudent.

There are a number of challenges in using a model like the one described in this paper. While there is a general desire to expand the use of the model, extreme care must be taken that users clearly understand the assumptions made by the model, or else misleading results can occur. These analyses are inherently complex and numerous assumptions are necessary. A recommendation for future work is to clarify these assumptions into an error-proofed user-friendly system.

7 REFERENCES

- Alcalde, Jeffrey. "The Design and Implementation of a Synchronous Manufacturing System in a Job-Shop Environment". Massachusetts Institute of Technology Master's Thesis. 1997.

- Goldratt, Eliyahu and Jeff Cox. *The Goal.* Croton-on-Hudson, New York: North River Press, Inc. 1984.

- Johnson, H. Thomas and Robert S. Kaplan. *Relevance Lost: The Rise and Fall of Management Accounting.* Boston, Massachusetts: Harvard Business School Press. 1991.

- Karmarker, Uday S., "Multilocation Distribution Systems". Technical Report 117, Operations Research Center, Massachusetts Institute of Technology. 1974.

- Karmarket, Uday S., "Lot Sizes, Lead Times and In-Process Inventories". *Management Science* Volume 33, Number 3. March 1987, 409-418.

- Nahmias, Steven. *Production and Operations Analysis.* Burr Ridge, Illinois: Richard C. Irwin, Inc. 1993.

- Strimling, Jonathan. "A Constraint Based Optimization of Manufacturing and Sales in the Copper Tubing Industry". Massachusetts Institute of Technology Master's Thesis. 1996.

- Womack, James P. and Jones, Daniel T. *Lean Thinking.* New York, New York: Simon & Schuster. 1996.

- Womack, James P.; Jones, Daniel T. and Roos, Daniel. *The Machine that Changed the World.* New York, New York: Rawson Associates. 1990.

- Zimmerman, Jerold L. *Accounting for Decision Making and Control.* Chicago, Illinois: Richard C. Irwin, Inc. 1995.

APPENDIX A: SCHEMATIC DIAGRAMS OF INPUT & OUTPUT MODELS

Order Tracking

Backorder
Weekly order   *Product*
Shipped volume   *yields*
Finished volume

bold = undated every day/shift
italic = standard parameter
normal = calculation or link

Remaining Coils
Remaining Units

Press, Block & Spinners

Remaining   *Process*
Coils   *rates*

Process times

Finish Lines

Remaining   *Routings*
units to
finish

Finish rates

Process times

Annealing

Remaining   **Anneal**
units to   **ing**
anneal   **WIP**

Anneal
rates

Process times

Machine Assumptions

 # Shifts remaining
Hours/shift
Downtimes
Scrap rates

Capacity Summary

Work
time   Downtime

Available   Process
time   times

Utilization

Figure A.1 Scheduling Model Schematic

Order Volumes

Volumes of Product Categories

Standard Mixes

Product yields

Coils to Process
Units to Process

bold = user input
italic = standard parameter
normal = calculation or

Press, Block & Spinners

Coils to Process

Process rates

Process times

Finish Lines

Units to Process

Routings

Finish rates

Process times

Annealing

Units to Process

Anneal rates

Process times

Machine Assumptions

 # Shifts
Hours/shift
Downtimes
Scrap rates

Capacity Summary

Work time

Downtime

Process times

Available time

Utilization

Cost Summary

Process times

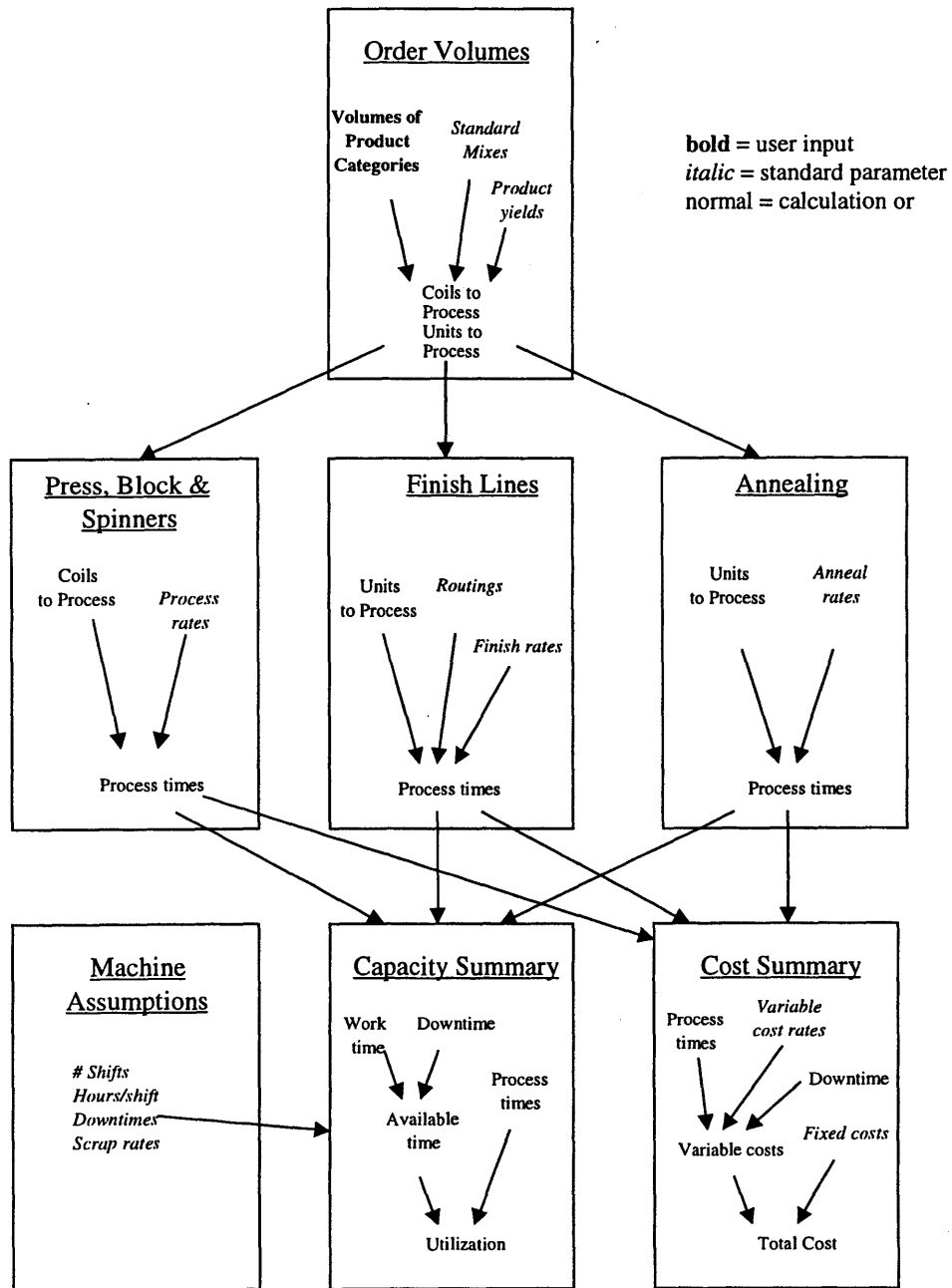*Variable cost rates*

Downtime

Variable costs

Fixed costs

Total Cost

Figure A.2 Operations Analysis Model Schematic

∂133−8