

# Measurement and control of business processes

Stephen G. Powell,<sup>a\*</sup> Markus Schwaninger<sup>b</sup> and Chris Trimble<sup>a</sup>

Stephen G. Powell is Professor of Management Science at the Amos Tuck School of Business Administration, Dartmouth College, USA. He has a Ph.D. in Engineering-Economic Systems from Stanford University. At Tuck he teaches management science, business process design, and modeling for managers. His research involves the design and management of production processes in manufacturing and services.

## Abstract

It is widely accepted that any well-designed organizational process includes a control mechanism through which management decides which aspects of the performance of the process are to be measured and how these measurements are to be used to change the level of resources utilized in the process. Little is known, however, about the best ways to design such a control mechanism for typical business processes. Our goal in this research is to identify control mechanisms for business processes that are effective in different types of environments. In this article we present a system dynamics model of a typical service-sector business process, such as is used in processing administrative paperwork in insurance, banking, and so on. These processes are subject to random, time-varying, and non-postponable demands for service. They are also subject to randomness in processing times, as well as delays in the observation of system performance and in the execution of control actions. We assume management has the dual objectives of maximizing profits (revenues on completed work less the costs of labor employed) and keeping cycle times below a predetermined ceiling. In order to achieve these objectives it observes the state of the process and adjusts its labor force accordingly. Management must choose which of several aspects of process performance to measure (cycle time, backlog, or demand) and the parameters governing the control process. Our analysis highlights the interactions among the demand environment faced by the process (e.g., random or seasonal), the control signal chosen (e.g., cycle time or backlog), and the type of control used (e.g., proportional or differential). Our results suggest that, regardless of the demand environment, a control process based on system backlog is generally more robust than the alternatives in the sense that adequate performance is achieved over a broader range of control parameters. We also find that, in most cases, proportional control by itself is inadequate to provide effective performance and that differential control is a necessary adjunct. We conclude the article with a discussion of the managerial implications of this research. Copyright © 2001 John Wiley & Sons, Ltd.

*Syst. Dyn. Rev.* 17, 63–91, (2001)

In the early part of the 1990s a great deal of attention was devoted by consultants and popular business writers to “business processes” (Davenport and Short 1990; Davenport 1993; Hammer 1990; Hammer and Champy 1993). The result was a somewhat faddish movement, which is only now maturing, called variously Business Process Reengineering, Business Process Redesign, or simply BPR. While the early proponents of these ideas were given to making exaggerated claims, at its core this movement was based on a few simple and powerful ideas. One of these core ideas is that business processes, or “how the work gets done around here”, are worthy of management

\* Corresponding author

<sup>a</sup> Amos Tuck School of Business Administration, Dartmouth College, Hanover, NH 03755, USA

<sup>b</sup> Institute of Management, University of St Gallen, CH-9000 St Gallen, Switzerland

Markus Schwaninger is Professor of Management at the University of St Gallen, Switzerland. His research is focused on Managerial Cybernetics and System Dynamics, applied to issues of organizational intelligence, management systems, and to the design, transformation, and learning of organizations. He is the author of 130 publications in six languages on systemic aspects of management, including *Organizational Transformation and Learning, A Cybernetic Approach to Management* (Wiley, Chichester, 1996, with co-authors) and six books in German, published with Campus, Frankfurt; Duncker & Humblot, Berlin; and Haupt, Berne.

attention. Clearly, business processes generate most of the costs of any business, so improving efficiency generally requires improving processes. Business processes also strongly influence the *quality* of the product and the satisfaction of the customer, both of which are of fundamental importance in the marketplace. A second idea at the foundation of BPR is the recognition that recent developments in information technology have created dramatically more effective alternatives for organizing work. Finally, the proponents of BPR recognized that many of the proven ideas and tools of industrial engineering were applicable to service-sector processes. In fact, Davenport referred to the application of BPR to a service business as "The New Industrial Engineering" (Davenport and Short 1990).

Many different approaches have been proposed for creating and managing business processes (Bach *et al.*, 1996), but it is taken for granted that any well-engineered business process is one in which management establishes *measurements* of process performance, and influences process performance in a desired direction by using these measurements to *control* the process. Control involves, among other things, changing the levels of resources available to the process. Thus the fundamental notion of *feedback control* is also a key part of BPR. Feedback control involves controlling the behavior of a system by measuring its state, comparing its current state to a desired state, and selecting control actions based on differences between actual and desired states (Forrester 1961; Sterman 2000).

While the science of feedback control is highly evolved in the engineering disciplines, within BPR it has not evolved beyond the rudimentary level of analogy. Consider, for example, the following prescription for managing a business process, (Oesterle 1995):

Process management plans, structures, and observes the process. ... It determines the performance indicators for evaluating the process, sets objectives, compares targets with achievements, derives actions for improving the process and monitors process implementation.

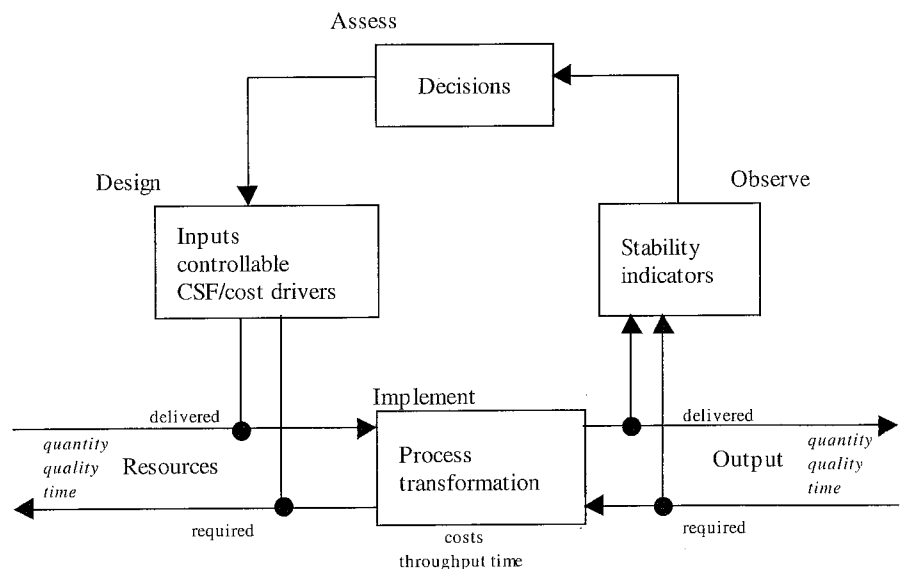
While the idea of feedback control is clearly implicit in this description, no further details are offered as to how one might implement the idea at the operational level. In fact, Oesterle suggests that the actual determination of managerial actions in response to gaps between targets and achievements be left to a committee, without suggesting how a committee would develop an understanding of the dynamic features of the system it must control. This omission is particularly disturbing in light of the fact that the control of engineering systems using feedback is subject to numerous pitfalls, such as cycling and instability (Franklin *et al.*, 1994; Sterman 1989). And most engineering systems are far less complex and better understood than the socio-technical systems we know as business processes.

Chirs Trimble is an Adjunct Associate Professor of Business Administration at the Tuck School, where he teaches Dynamic Modelling, and is the Executive Director of Tuck's William F. Achtmeyer Center for Global Leadership. Prior to coming to Tuck, he was a Senior Consultant at High Performance Systems, where he assisted a variety of clients in applying system dynamics modeling techniques to a variety of management issues.

Feedback control is also the defining concept of cybernetics, originally the engineering science of communication and control. *Management cybernetics*, which has its origins in engineering cybernetics, has evolved its own terms and tools as it attempts to apply cybernetic ideas to the social world. In a recent text on organizational cybernetics (Espejo *et al.* 1996), which addresses itself in part to the management of business processes, we find similar recommendations for managing processes using feedback. Figure 1, which is taken from Espejo *et al.*, shows that a business process uses resources to transform inputs into outputs for a customer. The management of such a process consists of three activities: observation, assessment, and design. Management must first specify "stability indicators" for its processes: process measurements that reflect the efficiency and effectiveness of the process. Next, management must assess the health of the process, that is, make decisions as to the appropriate value of controllable inputs based on the values of stability indicators and their goals for the process. Finally, management must decide which controllable inputs it will use to influence process performance. For example, if raw material quality is critical to successful process performance, management may need to negotiate contracts with suppliers that provide it the ability to influence raw material quality effectively in the future.

While management cybernetics offers many useful insights into the effective design and management of business processes, like other approaches to BPR it offers little specific guidance to managers at the operational level. A manager or management team in charge of operating a business process would benefit from knowing, for example, which measurements of a process most accurately

Fig. 1. A process control model (after Espejo *et al.* 1996, © John Wiley & Sons Limited, reproduced with permission.)



---

reveal its performance. They would also like to know over how long a period of time one should observe a process in order to get a dependable sense of its performance, and in what ways one should change resources in order to close gaps between observed and desired performance. This is just a sample of the kinds of concrete questions that need to be addressed if the idea of feedback control is to be made operational in an organizational context.

This article represents an attempt to identify general principles for the measurement and control of business processes. We take a normative viewpoint here. That is, we examine what management *should* do to control processes, not what it actually does. Our strategy is to build a mathematical model of a generic business process that has all the essential ingredients of such processes, including time lags and randomness. Such a model involves a level of complexity that is realistic with respect to actual business processes, and thus provides a useful laboratory within which to study different approaches to process control. By studying how simple control mechanisms work in this laboratory setting we hope to derive useful insights for the successful control of real-world business processes.

In what follows we first describe our model of the business process itself and the control system that governs it. Then we analyze the performance of alternative control regimes within five distinct demand environments. For each of these demand environments we compare the performance of control policies based on cycle time, backlog, and demand. Next we discuss the sensitivity of our results to underlying system parameters. We conclude with a discussion of the managerial implications of our inquiry and suggest some directions for further research.

### **A model for the control of business processes**

In this section, we describe a mathematical model that has the main ingredients of a generic business process. The following characteristics are essential to our model:

- Work arrives at an unpredictable rate.
- Work builds up in a queue whenever the work on hand exceeds the available labor force.
- The size of the labor force must be adjusted by management to process the work on hand and the work likely to arrive.
- Labor force productivity varies randomly according to a fixed distribution.
- Management can measure three aspects of the performance of the process: the cycle time for completed work, the backlog of work on hand, and the arrival rate of new work.
- Management's primary goal is to maximize profit, but it must also keep the cycle time below a certain level or suffer customer dissatisfaction.

- 
- Profit is measured as the revenue gained on completed work less wage and hiring costs.
  - The labor force changes only at discrete time intervals.

We emphasize that this model represents only the operational aspects of the business; we deliberately ignore in this analysis the aspects of strategic and normative management.

### *Overview of the model*

The model reflects a business process of the type typically encountered in service sector businesses, e.g., insurance claims processing, credit card approval, or college admissions applications processing. These processes have a small number of defining characteristics. First, they are *labor intensive*—although information technology can and has been used to speed the flow of work, it has not proven capable of replacing human knowledge workers who are required to make expert judgements about individual cases. Second, *demand* for service, or the arrival of customers, is generally unpredictable and non-postponable—for example, insurance applications arrive by mail whenever a customer sends one in and, although the process does not have to begin work on each application immediately, the customer generally determines his or her waiting time from the time of submission, not the time work begins. In contrast to the typical manufacturing process, in these processes the work required cannot be scheduled to match the available process resources. Instead, the resources must be provided to match the work. Third, there are *no binding limits* to the amount of work that can build up within the system. For example, even when insurance claims are filed on paper forms (rather than electronically), they do not consume enough space to encounter any practical storage limits. So again in contrast to the situation in a typical factory, controlling work-in-progress and cycle time by the arrival of work is not a feasible option for management. Finally, the *time required* to process a given customer may vary substantially as a result of differences among customers, workers, or other factors in the process.

Our model seeks to reproduce each of these characteristics of business processes. Work arrives unpredictably, perhaps randomly (according to a specified probability distribution). That is to say, management can observe the pattern of work arriving but cannot entirely anticipate future arrivals (in the sense of making a precise prediction). Each unit of work (or job) is pushed through the process, that is, work begins as soon as a worker is available. Newly arriving work waits in a queue whenever the available labor force is fully occupied. There is unlimited storage capacity for work in process. Finally, the processing times for each job are again governed by a probability distribution. (Although we assume these distributions are given exogenously, we do not assume they are known to management.)

---

Management's central goal for this process is simply to make as much profit as possible. That requires processing as many customers as possible in a given time period (each of which brings in a fixed amount of revenue upon completion) but with the minimum labor force possible, since the major cost of the process is the cost of labor. In order to penalize control policies that depend on excessive changes to the labor force, we also assume there is a cost to hiring labor. We will measure profit with a *Profit Index*, given by the ratio between the actual profit achieved and the profit that could have been achieved with no hiring costs.

In addition to maximizing profits, management will want to ensure that the cycle time of the process does not exceed a level that is acceptable to customers. We will operationalize this second goal by measuring the percentage of customers for whom the actual cycle time exceeds a cycle time limit. We refer to this measure as the *Lateness Index*. The cycle time limit is established by the competitive environment and is, therefore, not a decision variable.

The task faced by management is to adjust the labor force so as to produce as much profit as possible while simultaneously ensuring that cycle times do not exceed the cycle time limit. In order to achieve these goals, management measures three aspects of the state of the process: cycle time, backlog, and demand. Each of these measures is taken over an averaging interval. If the averaging interval is short, management gets a quick signal that the system is deviating from its previous state; however, in some environments this signal may be largely noise. A longer averaging interval would give management a more accurate picture of the true state of the system, but may delay the recognition of significant problems. We will also assume that management action to change the labor force occurs only at discrete points in time, say monthly. We will refer to the time between changes in the labor force as the staffing interval.

Why might management chose one of the three process measures (cycle time, backlog and demand) over the others? It would seem appropriate to base the control policy on cycle time, since this is a direct reflection of customer satisfaction. However, customers are satisfied by the cycle time they receive as individuals, whereas (in our model) management measures cycle time as an average over many customers. Thus, individual customers may be dissatisfied even though management's measure of cycle time appears to be acceptable. Using backlog as the control signal may be appealing since problems in the system should show up in rising backlogs before they show up in longer cycle times, allowing management more time to take corrective action. By the same logic, using demand itself as the control signal may be appropriate because demand provides the earliest possible indication of a change in operating conditions.

According to Little's Law (Allen 1978), in the steady state the average backlog is equal to the average demand times the cycle time. Thus, our three measures of system performance can be expected to more or less follow each

other. This might suggest that the system can be controlled equally well using any one of the three measures, but a dynamic system with lags, feedback, and randomness may not reach a steady state, especially over realistic time intervals. Moreover, management must control these systems in real time, on the basis of recent information. Thus, it is an open question which of the available system measures is most effective in controlling this system.

The actual rules management uses to link observations of the process state to changes in resources involve two types of control: *proportional* and *differential* (Franklin *et al.* 1994; Sterman 2000). For illustrative purposes, assume for the moment that management has chosen to base its control policy on its measurement of cycle time (similar control rules are used for the backlog and demand controllers). Proportional control is identical to the first-order feedback rule commonly used in system dynamics models (Forrester 1961; Sterman 2000). If, for example, the measured cycle time is 18 days and the target cycle time is 15 days, the labor force will be increased in proportion to the gap of three days between actual and target values. The constant of proportionality in this relationship is called the *proportional gain*. A proportional gain of 0.5, for example, implies that one worker will be added to the labor force each period when the gap between measured and target cycle time is two days.

Differential control also uses first-order feedback, but this control adjustment is proportional to the *first difference* in the gap between measured and target cycle times. Thus if the gap is positive and *widening*, the use of differential control in addition to proportional control will lead to a stronger control response than otherwise. We refer to the proportionality constant in this case as the *differential gain*. A differential gain of 0.5, for example, implies that one worker will be added to the labor force each time period in which the gap between measured and target cycle times *changes* by two days. Differential control is widely used in engineering systems in conjunction with proportional control since it can help stabilize systems that would otherwise be difficult to control.

We can express these control signals in the form of equations as follows:

$$\left. \begin{array}{l} \text{Proportional} \\ \text{adjustment}_t \end{array} \right\} = a * (\text{Measured Cycle Time}_t - \text{Target Cycle Time}_t)$$

$$\left. \begin{array}{l} \text{Differential} \\ \text{adjustment}_t \end{array} \right\} = b * [(\text{Measured Cycle Time}_t - \text{Target Cycle Time}_t) \\ - (\text{Measured Cycle Time}_{t-1} - \text{Target Cycle Time}_{t-1})]$$

At the operational level, then, the task facing management is to choose which of the three measurements of the process state (cycle time, backlog, or demand) to use in its control policy, and then to choose the control parameters (or gains) for the proportional and differential control inputs.

### Detailed description of the model

We have implemented this model in *ithink* (High Performance Systems, Inc., Hanover, New Hampshire, USA), a software package designed to support simulation of dynamic systems. The complete model diagrams, equations and user interface are available from the authors.

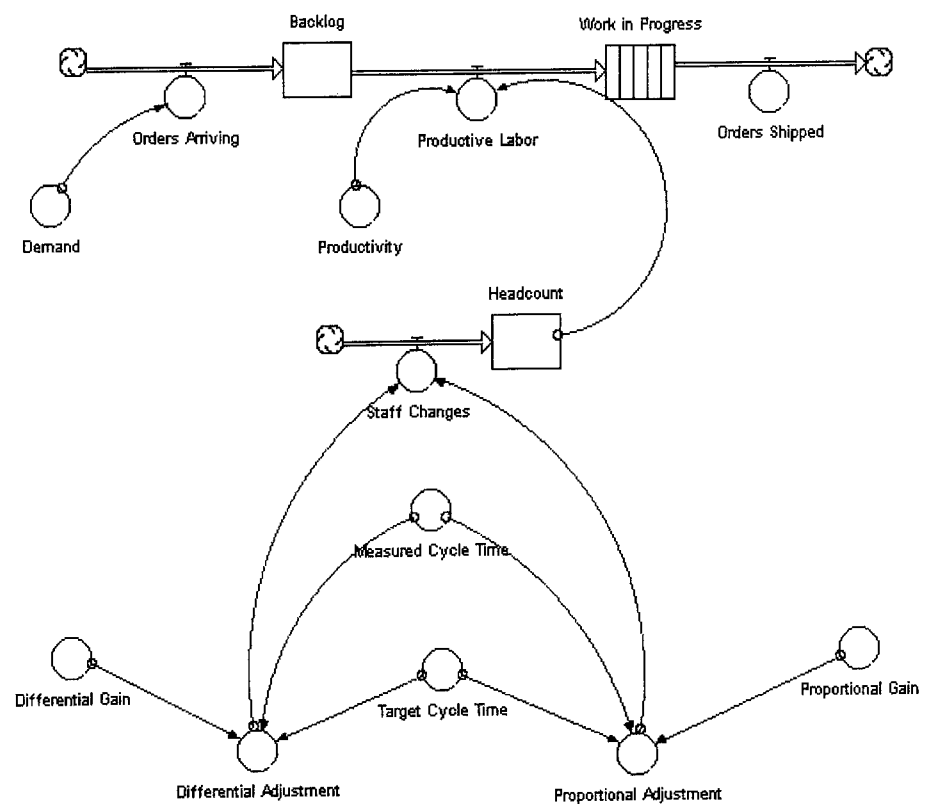
In Figure 2 we display the essence of the model in stock-flow format. The process involves a backlog stock that is fed by arriving orders and drained by the stock of work-in-process. Work is completed at a rate determined by the current headcount and productivity. The remaining stock is headcount, which is adjusted up or down using proportional and differential feedback based on the gap between measured and target cycle times.

The actual model consists of the five sectors described below.

### Demand sector

The demand generator includes a set of parameters by which different environmental scenarios can be simulated. We consider five distinct types of

Fig. 2. A simplified stock-flow representation of the business process model





---

demand environment: Step, Ramp, Spike, Random and Seasonal. In the Step environment, demand increases once by a set amount. In the Ramp environment demand increases at a constant rate. The Spike environment includes a single discrete spike in demand. In the Random demand environment, average demand is constant but actual demand varies each period according to a normal distribution. Finally, in the Seasonal environment demand cycles with a fixed amplitude and period.

#### *Production sector*

Orders arrive according to the specified demand environment. If all workers are busy, arriving orders accumulate in a backlog. The available labor force multiplied by labor force productivity (itself a random variable) determines the rate at which orders can be withdrawn from the backlog. Once withdrawn, there is a minimum time delay associated with fulfilling an order, representing the actual work time needed to process a job or customer. This time delay plus any time spent waiting in the backlog is equal to the cycle time for that order.

#### *Measurement sector*

Three aspects of the process state are measured: cycle time, backlog, and demand. Each is averaged over an Averaging Interval specified by management. Measured Cycle Time is the average cycle time over all work completed during the Averaging Interval (e.g., the last 30 days). Similarly, Measured Demand is the average over the Averaging Interval of arriving work. Backlog is measured in days of work on hand by dividing the number of units of work in the Backlog by the amount of labor on hand and averaging over the Averaging Interval.

Each of these measures of system performance can be interpreted as management's *forecast* of future performance. Essentially these are simple moving averages of the most recently available data. We have not implemented more complex forecasting procedures for two reasons: one is that moving averages are themselves a commonly used forecasting technique; the other is that we have attempted to keep the model as simple as possible and yet capture the essence of business processes. We plan to study the interaction between forecasting and control in a future paper.

#### *Adjustment sector*

Labor force is the resource management uses to achieve its goals of high profit with low cycle times. It accomplishes this by choosing first on which of the measurements (cycle time, backlog, or demand) to base its control policy. Then it chooses parameter values for the proportional and differential gains for these measurements. For example, management could choose to control using only Measured Cycle Time, and to set the proportional gain to 0.5 while setting the differential gain to 0.0. For each control signal there is a target value toward which the system aims: in the case of cycle time, we set a target of

15 days. The minimum production delay is 10 days and the cycle time limit (set by the market) is 20 days. When the observed cycle time exceeds 15 days the labor force is increased; when it falls below 15 days it is decreased. When the backlog is used as the control signal, the labor force is increased when the Measured Backlog exceeds five days (with a minimum production delay of 10 days this results in a cycle time of 15 days). Finally, when demand is the control signal, labor force increases whenever the Measured Demand exceeds the current headcount.

### *Results sector*

The results sector accounts for profits and actual cycle times. Profit includes revenues for each unit of work completed less wage and hiring costs.

### *Base case parameters*

The essential parameters used in our model are listed in Table 1. Incoming demand averages 10 units of work per day. Ten workers are initially in place, so with a productivity of one unit of work per worker per day this staff is adequate to handle the load. The actual work time on any one unit is 10 days, but any time spent waiting in queue for a worker to become available adds

Table 1. Base case parameters

Demand	
Average demand	10 units per day
Step demand	Demand increases permanently to 20 units on day 51
Ramp demand	Demand increases by 0.02 units per day starting on day 51
Spike demand	One-time demand of 100 arrives on day 51; otherwise demand is 10
Random demand	Demand is normally distributed with a mean of 10 and a standard deviation of 5 units per day
Seasonal demand	Demand varies periodically around a mean of 10 units per day with a period of 150 days and a range from 7.5 to 12.5 units
Production	
Minimum production time	10 days
Staffing delay	30 days
Averaging interval	30 days
Distribution productivity	Uniform with a range of 0.5 to 1.5 units per person per day
Performance	
Unit price	\$ 1 per unit
Labor cost per day	\$ 0.5 per person per day
Staffing cost	\$ 10 per person
Minimum cycle time	10 days
Target cycle time	15 days
Cycle time limit	20 days
Modeling	
DT	1 day

to the overall cycle time for that unit of work. Customers accept cycle times up to 20 days, and management uses 15 days as the cycle time above which it initiates efforts to increase the labor force. It also decreases the labor force when the cycle time is below 15 days. The control signal is either demand, backlog, or cycle time. Each is measured as an arithmetic average over the preceding 30 days. Decisions are made every 30 days. Each worker added to the labor force costs \$ 10; workers are paid \$ 0.5 per day and the product is sold for \$ 1.

### *Study design*

Within the boundaries of this research we view management as making two sets of decisions: which measurement of system performance to use: demand, backlog, or cycle time; and what values to assign to the proportional and differential control gains. We assume that management takes as given all other parameters of the system, including the averaging interval over which the process measurements are taken, the frequency with which staffing decisions are made, fluctuations in worker productivity, and the target for control actions.

Our strategy for analyzing this problem is to specify the demand environment and the control signal, and then to search for effective values of the two control gains. There are 15 combinations of the five demand environments and three control signals. For ease of comparison we use the same sets of values for the control gains for all 15 combinations of demand and control signal. We chose 10 values each for the two control gains, spanning four orders of magnitude, and giving rise to 100 cases. After simulating each of these 100 cases for each of the 15 combinations we impose a coarse screen to identify acceptable outcomes. We define as *acceptable* any case in which profit is at least 75 percent of the maximum achievable for the given demand environment and no more than 20 percent of the completed work has a cycle time that exceeds the cycle time limit (20 days). That is to say, acceptable results require that the Profit Index exceeds 75 percent and the Lateness Index does not exceed 20 percent.

In general we seek control policies that are *robust*. In this case robust has two meanings. One is that for a given demand environment and control signal there are many combinations of control gains that give acceptable results. This requirement simply ensures that management is not expected to fine tune its control policy to the given environment, but in fact can be imprecise in its choice of control gains and still achieve acceptable results. The second, stronger, sense in which a control policy can be robust is if a wide range of control gains give acceptable performance over most or all demand environments. The ideal control signal is one that controls the system adequately, regardless of the demand environment and the numerical values of the control gains.

## Effective control in different demand environments

The system begins in equilibrium at time zero with 10 workers and a steady stream of 10 units of work arriving each day. The demand changes begin on day 51. The simulation then runs for another 1200 days. We will first discuss the effects of using each of the three control signals (cycle time, backlog, or demand) in each of the five demand environments. Then we will investigate how robust each control signal is over all five demand environments. Finally, we will consider the problem of selecting the optimal values of the control gains for backlog control in each demand environment.

### Control within a given demand environment

*Step demand*

In this environment, demand increases on day 51 from 10 to 20 units and remains there for the remainder of the simulation. The equilibrium or desired labor force also increases from 10 to 20 workers. The goal of an effective control regime in this case is to recognize the demand increase quickly and to increase the labor force to the new equilibrium level without lengthy delays or expensive cycling. In Figure 3 we show which combinations of proportional and differential gains result in acceptable performance for each of the three control signals.

Figure 3 shows that nine combinations (out of the 100 tested) of proportional and differential gains result in acceptable performance when cycle time is the control signal. These are primarily cases with intermediate values for the proportional gain and high values of the differential gain. The best Profit Index achieved is 97.03 percent (with a Lateness Index of 19.85 percent); the minimum Lateness Index achieved is 12.84 percent (with a Profit Index of 94.79 percent). Thus, if the appropriate control gains can be found, there is little need in this case to trade off profit for cycle time performance.

Fig. 3. Acceptable combinations of control gains for Step demand

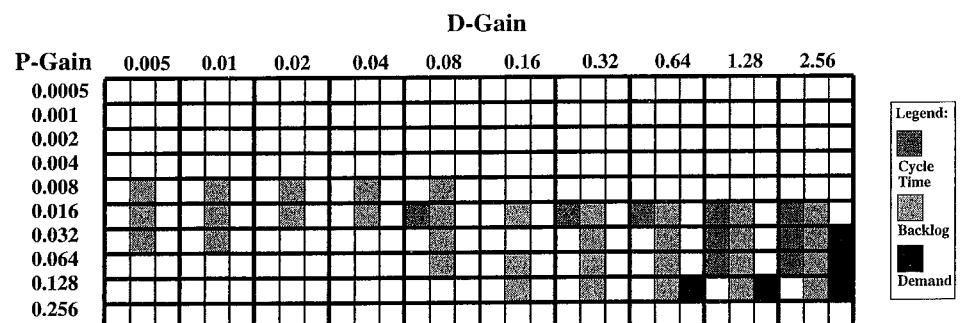


Figure 3 also shows that 33 combinations of the control gains result in acceptable performance when backlog is the control signal. These tend to be cases with intermediate to high values for the proportional gain. When the differential gain is high, a wide variety of values for the proportional gain are acceptable. The best Profit Index achieved is 97.66 percent (with a Lateness Index of 16.43 percent); the minimum Lateness Index achieved is 0.42 percent (with a Profit Index of 91.73 percent). In this case there is some trade off between profit and cycle time performance: achieving minimal cycle time costs six percentage points in profit. These results suggest that in a step demand environment, backlog control is both more robust and more effective than cycle time control: many more values of the control gains result in acceptable performance, and almost perfect control of cycle time is possible with a modest penalty in profit.

Finally, we see that five out of 100 cases provide acceptable performance when demand is the control signal. These are cases with high values of both proportional and differential gains. The best Profit Index achieved is 99.04 percent (with a Lateness Index of 18.01 percent); the minimum Lateness Index is 0.00 percent (with a Profit Index of 98.68 percent). Demand control would appear to be the least robust of the three options in the Step environment; however, it does achieve perfect control of cycle time in one case.

Figure 4 shows a time plot of headcount and excess cycle time for the Step demand case under cycle time control. The control gains are chosen from the set that produces acceptable performance. We see that the step in demand at time period 50 induces cycling in both the headcount and excess cycle time. Cycle time occasionally rises as high as 25 (an excess of 10 over the target of 15 days), but generally is held under 20 days.

### *Ramp demand*

In this environment, demand increases at a constant rate of 0.02 units per day starting on day 51. At this rate, demand reaches 33 units per day at the end of 1250 days. In an environment in which demand is growing steadily, there is no equilibrium toward which the system can move. An effective control policy must constantly increase the labor force, without significantly overshooting (which would reduce profits as a result of excess hiring costs) or undershooting (which would cause cycle times to exceed the cycle time limit). Figure 5 shows which combinations of proportional and differential gains result in acceptable performance for each of the three control signals.

Figure 5 shows that only one combination of proportional and differential gains results in acceptable performance when cycle time is the control signal. This case involves high proportional and differential gains. The Profit Index achieved is 88.93 percent (with a Lateness Index of 19.68 percent). In general, the control gains that result in high profit are nearly identical to those that give low cycle time, and in either case little is lost on the other outcome measure.

Fig. 4. Time plot of headcount and excess cycle time for Step demand

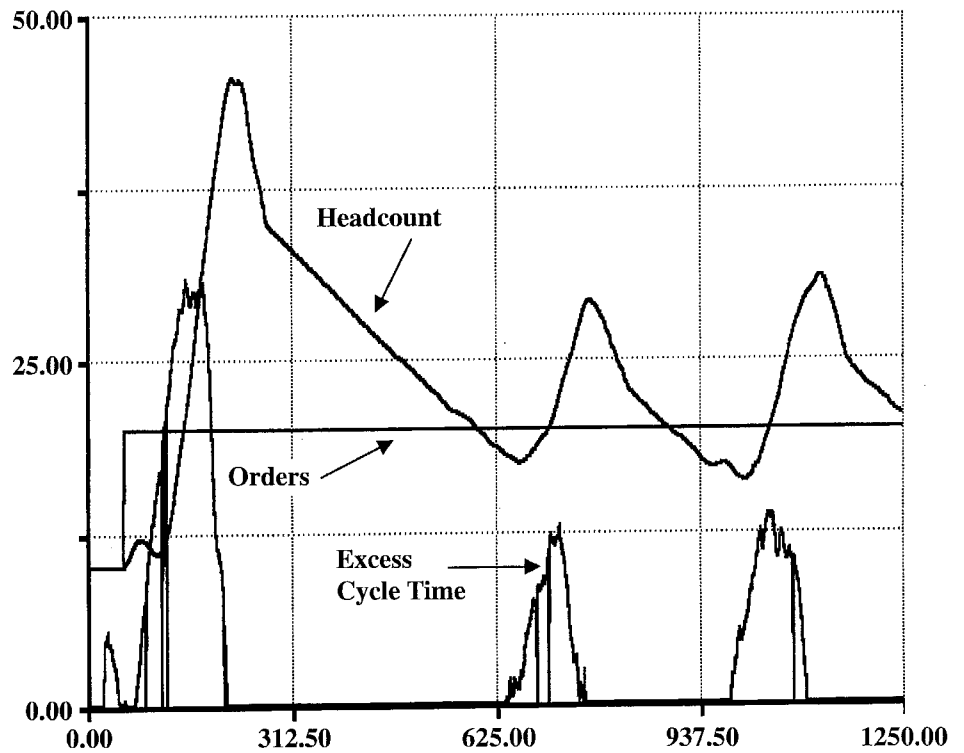


Fig. 5. Acceptable combinations of control gains for Ramp demand

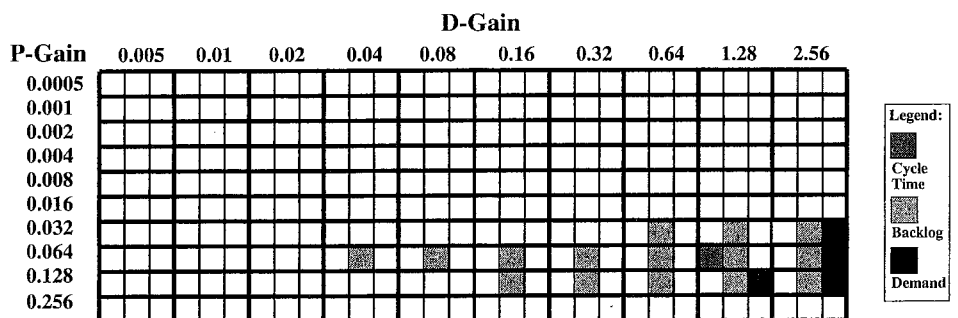


Figure 5 also shows comparable results for backlog control. Here 15 cases are acceptable, with high proportional control gains and medium to high differential gains predominating. The highest Profit Index is 95.95 percent (with a Lateness Index of 6.09 percent); the minimum Lateness Index is 2.92 percent (with a Profit Index of 84.81 percent). In this case, minimizing cycle time reduces profit by a quite substantial amount. As in the case of Step

demand, it appears that backlog control is more robust than cycle time control and can achieve lower cycle times (if the values of the control gains are chosen precisely).

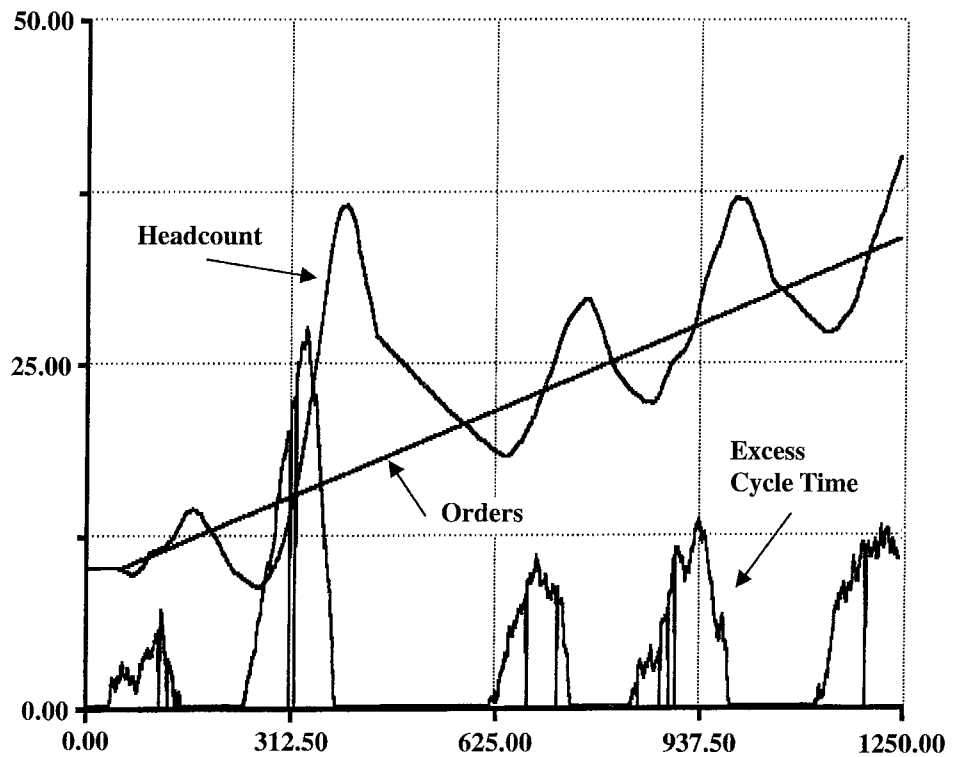
Finally, Figure 5 shows that in four cases out of 100 control based on demand achieves acceptable results. These typically involve high proportional and differential control gains. The highest Profit Index is 97.31 percent and the Lateness Index is 0.0 percent. Perfect control of cycle time is possible in this scenario, but only with quite precise choice of the control gains.

Figure 6 shows a time plot of headcount and excess cycle time for the Ramp demand case under cycle time control. The control gains are chosen from the set that produces acceptable performance. This control policy does not eliminate cycling, but it does manage to keep excess cycle time under control despite the steady growth in demand. Headcount increases to match growing demand, as it must, but it also cycles widely.

#### *Spike demand*

A spike in demand is a one-time increase that is never repeated. In this environment, demand is 10 units per day except on day 51, when it jumps

Fig. 6. Time plot of headcount and excess cycle time for Ramp demand



to 100 units. Since this demand increase is a one-time event, it would be easy to overreact to it and add too many workers for too long. On the other hand, simply ignoring the demand spike would lead to unacceptable cycle times forever, as the backlog created by the spike would never be removed. We would expect, then, that an effective control policy would add enough workers to remove the backlog created by this spike, but would then quickly reduce the labor force to save costs. Poor control policies will result in cycling and overshoot. Our results are summarized in Figure 7.

A glance at this figure shows that for all three control signals there are many cases of acceptable performance: 51 for cycle time, 68 for backlog, and 60 for demand. Low values of the proportional gain predominate among acceptable cases for cycle time and backlog control, while high values predominate for demand control (differential control is largely irrelevant in this scenario). For cycle time control the highest Profit Index reached is 99.72 percent (with a Lateness Index of 9.17 percent); the minimum Lateness Index is 2.67 percent (with a Profit Index of 89.01 percent). For backlog control the highest Profit Index is 99.86 percent (with a Lateness Index of 3.92 percent); the minimum Lateness Index is 0.0, which is achieved in 14 cases with high differential gains (the lowest Profit Index among these cases is 95 percent). Finally, in the case of demand control, the highest Profit Index achieved is 99.94 percent (with a Lateness Index of 18.35 percent); the minimum Lateness Index is 0.92 percent (with a Profit Index of 98.82 percent).

It would be difficult to choose a preferred control signal for this environment. Even though a spike of 100 units represents ten days of normal demand arriving in one day, all three control regimes handle the disruption fairly well. Perhaps the most robust policy is to use backlog control, since it achieves an almost perfect outcome on the cycle time dimension (Lateness Index below 1.0 percent) in seven cases with differential gain of 12.8.

Figure 8 shows a time plot of headcount and excess cycle time for the Spike demand case under cycle time control. The control gains are chosen from the set that produces acceptable performance. In this case very little cycling is evident in headcount, although there are several severe spikes in cycle time.

Fig. 7. Acceptable combinations of control gains for Spike demand

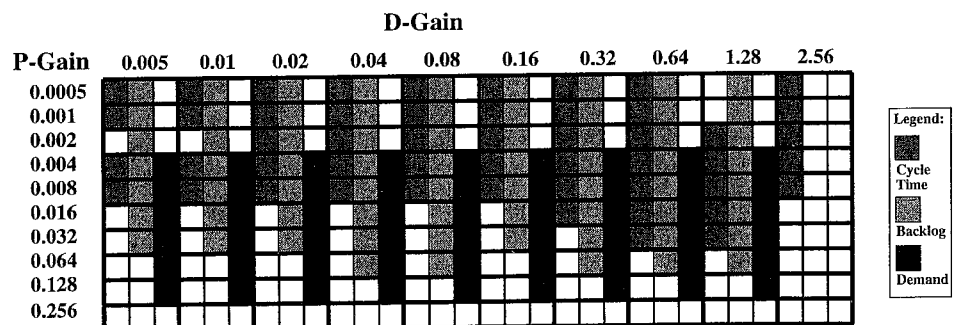
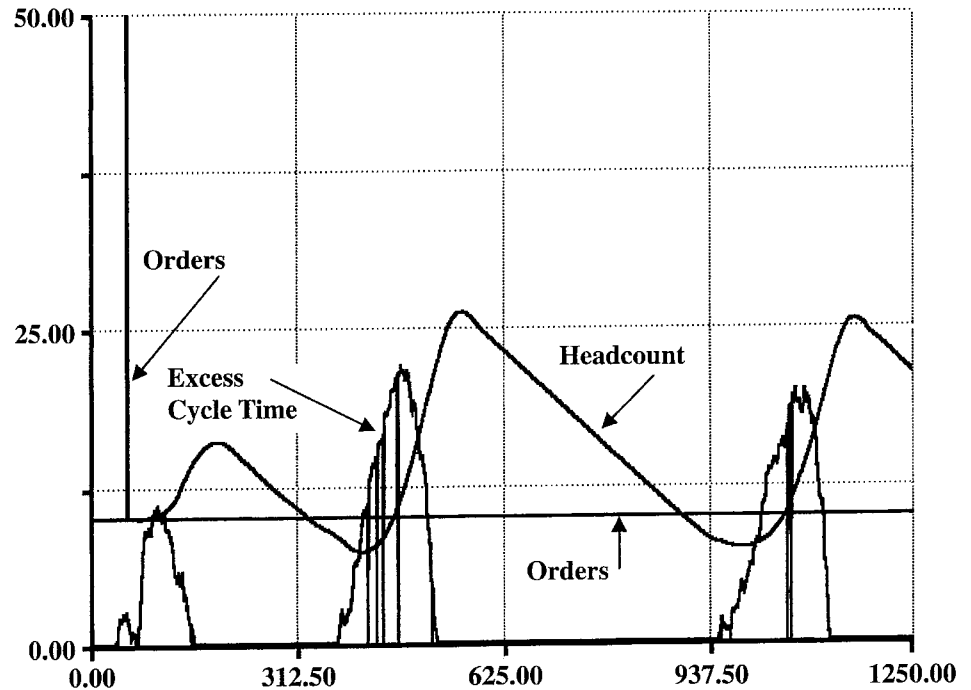




Fig. 8. Time plot of headcount and excess cycle time for Spike demand



The spike in demand creates a large backlog of work in the system, but this backlog can be eliminated rather quickly with modest increases in headcount. Given our definition of acceptable performance, it is preferable to react slowly to this demand spike, and allow the cycle time to exceed its target for a brief period, than to increase headcount drastically and eliminate all excess cycle time.

#### *Random demand*

In the random demand environment, daily demand is governed by a normal probability distribution with a mean of 10 units and a standard deviation of 5 units. A random demand environment is somewhat analogous to a spike environment in that *average* demand never deviates from the original value of 10 units per day. However, instead of a single positive spike in demand, in the random demand case there are daily fluctuations above and below the mean value. If actual demand is below the mean for a period of time the backlog will fall and the labor force will become idle at some point; on the other hand, if demand is above the mean for a period of time the backlog will build and additional labor will be required. In this environment an effective control policy must react appropriately to these random events and not ignore the build up of backlog or growing cycle times. However, it also must not overreact

and add too much labor for too long in response to an increase in demand (relative to the available labor) that is in fact merely random. Our results are summarized in Figure 9.

Starting with the case of cycle time control, we see that acceptable performance is achieved in 42 of 100 cases. These are generally cases with low to medium values of the proportional gain. We also note that the value of the differential gain is irrelevant in this case, presumably because random demand creates no real trends in system performance that differential control is needed to react to. The best Profit Index obtained is 98.18 percent (with a Lateness Index of 17.04 percent); the minimum Lateness Index is 9.29 percent (with a Profit Index of 86.25 percent). Low cycle times can be achieved in this case but with significant loss in profit. Cycle time control appears to perform well in the face of this level of demand uncertainty, over a wide range of control parameters.

Backlog control also appears highly effective in this demand environment. Acceptable results are achieved in 59 percent of the cases tested. Generally, the acceptable cases involve low to medium values of the proportional control gain (most values of the differential gain appears to work equally well). The best Profit Index obtained is 97.72 percent (with a Lateness Index of 7.52 percent); the minimum Lateness Index is 0.92 percent (with a Profit Index of 93.05 percent). Again, perfect cycle time control is achievable here with only a negligible loss in profit. In fact, perfect cycle time control is achieved with four different combinations of control gains.

Finally, Figure 9 shows that demand control works effectively in 31% of the cases tested, typically for high values of the proportional gain (the differential gain again makes little difference). The best value for the Profit Index is 98.66 percent (with a Lateness Index of 14.36 percent); the minimum Lateness Index is 1.34 (with a Profit Index of 97.71 percent).

Although all three control signals provide adequate performance in many cases in a random demand environment, backlog control appears to work well over the widest range of parameters, while demand control also achieves good

Fig. 9. Acceptable combinations of control gains for Random demand

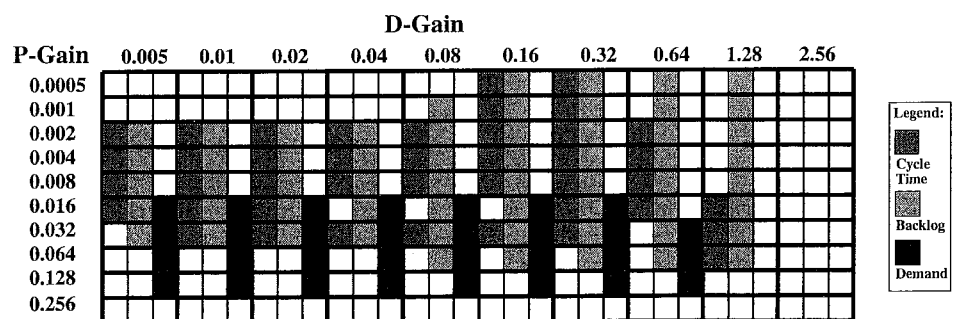
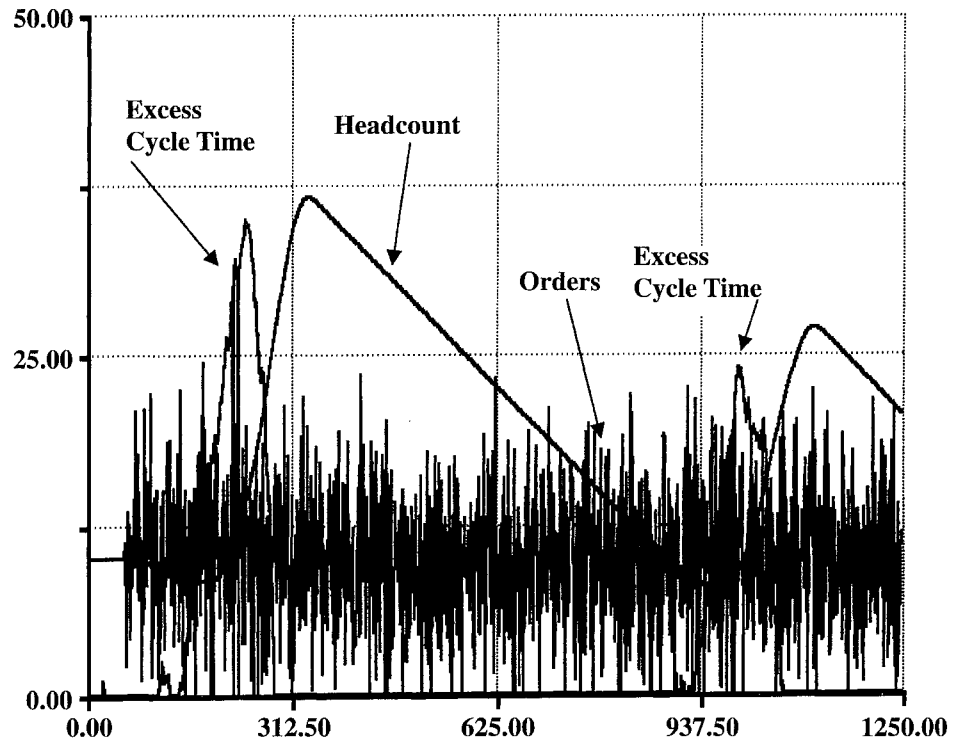


Fig. 10. Time plot of headcount and excess cycle time for Random demand



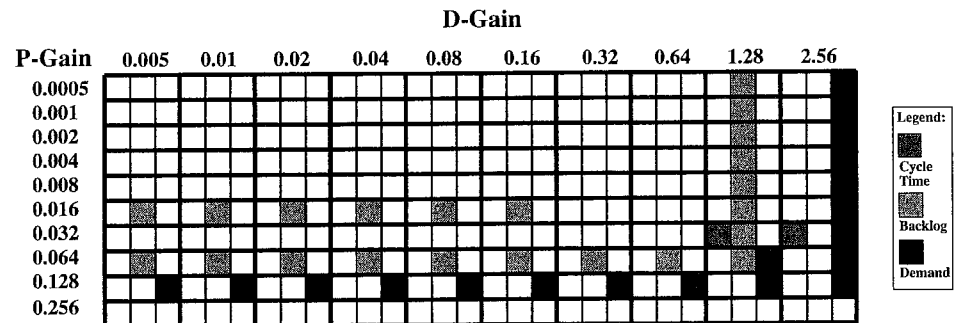
cycle time control in many cases. It is also noteworthy that in all three cases the value of the differential control gain is not critical.

Figure 10 shows a time plot of headcount and excess cycle time for the Random demand case under cycle time control. The control gains are chosen from the set that produces acceptable performance. As in the case of Spike demand, very little cycling is evident in headcount, although there are several modest spikes in cycle time. Under random demand the goal of control should be not to overreact to noisy demand. In the case illustrated here, random demands eventually create a backlog against which headcount must be increased. As the backlog is worked off, cycle time drops to an acceptable level and headcount is reduced. Eventually, another bulge is generated randomly in backlog, cycle time increases, headcount increases, and the cycle repeats.

#### *Seasonal demand*

Our final demand scenario involves seasonal fluctuations around a constant mean value. In this case, demand cycles from a high value of 12.5 units per day to a low of 7.5, within a cycle of 250 days. A glance at Figure 11 shows that this demand environment is among the more challenging for effective control.

Fig. 11. Acceptable combinations of control gains for Seasonal



In the case of cycle time control we observe only two cases in which both profit and cycle time are acceptable. All three involve a high differential gain and moderate proportional gains. The best Profit Index achieved is 85.69 percent (with a Lateness Index of 18.27 percent); this is also the minimum Lateness Index among these cases. It appears that basing control on observed cycle time in a rapidly changing environment puts extreme burdens on management to pick control gains precisely.

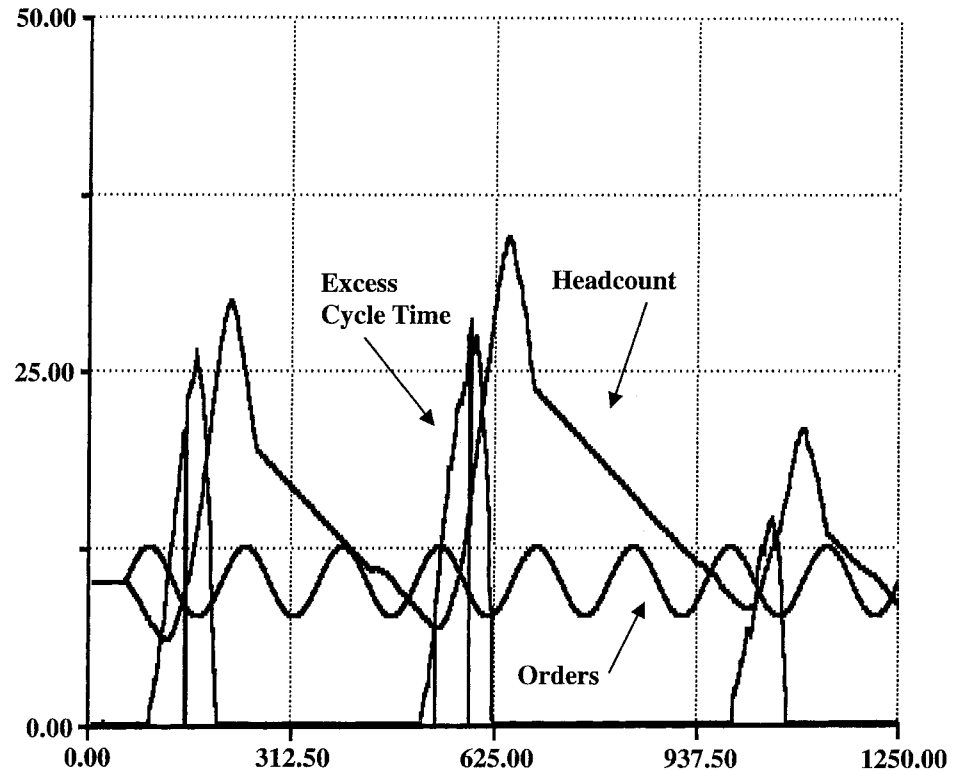
Figure 11 shows 22 cases of acceptable performance when backlog is the control signal. High values of the differential control gain or moderate values of the proportional gain work in this case. The best value of the Profit Index is 95.19 percent (with a Lateness Index of 14.18 percent); the minimum Lateness Index achieved is 5.42 percent (with a Profit Index of 85.69 percent).

Finally, we observe 19 cases in Figure 6 in which demand control leads to acceptable performance. These also involve high values of either the proportional or differential control gains. The best value of the Profit Index is 97.11 percent (with a Lateness Index of 9.59 percent); the minimum value of the Lateness Index is 0.00 percent (with a Profit Index of 95.81 percent).

Overall, it appears that demand control is most effective in a environment in which demand is cyclical. Although backlog control is acceptable in a few more cases, the best results for cycle time require a significant loss in profit. Demand control achieves perfect cycle time control with very little loss in profit.

Figure 12 shows a time plot of headcount and excess cycle time for the Seasonal demand case under cycle time control. The control gains are chosen from the set that produces acceptable performance. Not surprisingly, when demand is seasonal and management cannot anticipate demand, feedback control leads to cycling. However, these control gains manage to keep cycle time from rising much above 25 days, at the expense of a constantly changing labor force.

Fig. 12. Time plot of headcount and excess cycle time for Seasonal demand

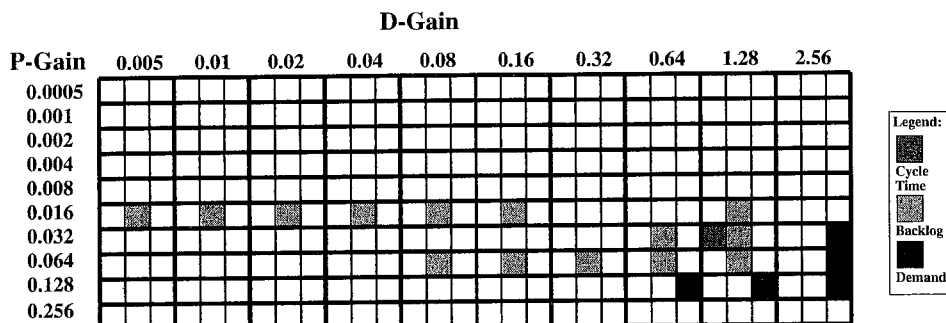


#### *Control across all demand environments*

It is true, of course, that in most cases management will not know in advance whether future demand is likely to grow steadily, fluctuate randomly, or exhibit some other pattern. We have used distinct, ideal-type demand patterns in our analysis not because they are likely to be known in advance but because they help us to understand and compare the performance of different control policies. But what control policy should management adopt if it does not know the demand environment it will face in the future?

One way to address this question is to examine the performance of the three control regimes over all five of the demand environments simultaneously. In Figure 13 we show the combinations of control gains for which each control signal gives acceptable results in *at least four* demand environments. On this measure, cycle time control is effective in only one case, demand in five cases, and backlog in 14 cases. Moreover, backlog control is effective in all five demand environments in five of these 14 cases, and it is the only control regime effective in all five. The important conclusion is that *backlog control is the most robust of the three alternatives when the*

Fig. 13. Acceptable combinations of control gains over all demand



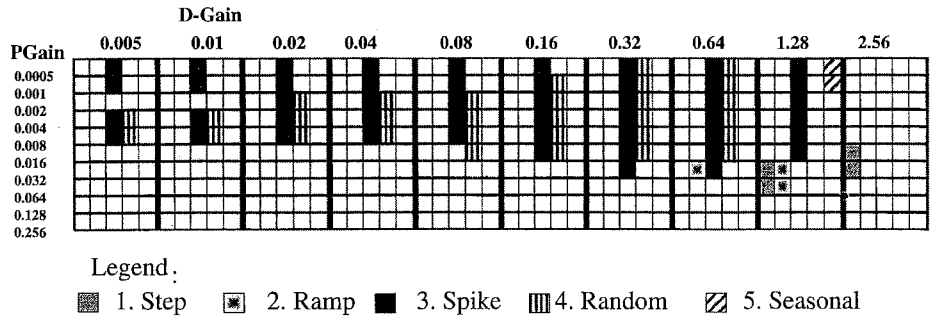
*demand environment is unknown.* Management can be assured that if it chooses backlog control (with proportional and differential gains around 0.064 and 0.32, respectively) it will achieve acceptable profit and cycle time performance in most of the demand environments it is likely to encounter. Of course, there is no guarantee that backlog control will continue to be effective in other demand environments. However, these results do suggest that backlog control may have properties that make it particularly attractive.

#### *Optimal control using backlog control*

Another question of interest to management is how to achieve *optimal* performance, not just acceptable performance. Up to this point we have concentrated on how to achieve a Profit Index above 75 percent and a Lateness Index below 20 percent. As we tighten our standards on each of these dimensions we can expect the number of acceptable cases to shrink. We would like to know whether the number of acceptable cases shrinks quickly or slowly as we tighten our standards; clearly, a control policy that achieves very high performance over many combinations of control gains is to be preferred.

We will illustrate this idea using backlog control since it appears to be robust over most environments. In Figure 13 we have shown the combinations of control gains for which backlog control achieves at least 75 percent of the optimal profit and does not exceed 20 percent unacceptable cycle times. In Figure 14 we have tightened the standards to 95 percent of optimal profit and 15 percent unacceptable cycle times. Backlog control generates acceptable results by these standards in four cases with Step demand, three cases with Ramp demand, 49 cases with Spike demand, 31 cases with Random demand, and two cases with Seasonal demand. A glance at the figure shows that low to moderate values of the proportional gain alone are sufficient to achieve excellent results in the Spike and Random environments. In the other three environments (Step, Ramp, and Seasonal), high values of the differential gain

Fig. 14. Optimal combinations of control gains for backlog



are essential to achieving best results. There is a zone around a proportional gain of 0.032 and a differential gain of 1.28 that gives excellent performance in all but the Seasonal environment. (Figure 11 shows that these parameters give acceptable performance even under that scenario.) So a fairly well-defined zone appears to exist for backlog control in the demand environments tested here in which good to excellent performance is assured. No doubt this zone is somewhat dependent on the structure of the production process and the parameters used here, but it is reassuring to find that wildly different parameters are not required for good control in many demand environments.

#### Concluding observations

First, we have observed that *the demand pattern faced by the business process has a strong influence on the qualitative and quantitative nature of an effective control regime*. For example, in an environment where demand increases once for all time, backlog control generally outperforms cycle time control. Furthermore, under random demand and backlog control it may be unnecessary to utilize the *change* in the gap between actual and desired states (differential control), while under seasonal demand it may be vital.

Second, we have found that *a control policy based on cycle time itself is workable in some demand environments but not in others*. In particular, in Spike and Random environments it performs acceptably over a wide range of control gains, but it does so for very few values of these parameters in Step, Ramp, or Seasonal environments. This somewhat surprising result suggests that, while monitoring cycle time is critical to having feedback on customer satisfaction, basing a control policy upon it may not be advisable.

Third, we find that *basing control on the incoming demand stream also has weaknesses*. Generally speaking, a control scheme based on the incoming demand performs poorly because it cannot recognize the existence of a large backlog of work in the system. Such a backlog can arise not only from high demands, but also from internal behavior such as low worker productivity.

Demand control performs acceptably only when the demand environment is stable enough not to cause high backlogs on its own and the control scheme and internal system behavior are such that backlogs are not created.

Fourth, *backlog control appears to be the best overall choice for a control signal upon which to base a control policy*. It not only performs adequately for many combinations of control gains in most of the demand environments we tested, but it performs very well in *all* environments for the same set of parameters. Backlog control appears to strike a balance between measuring changes too early (as demand control does) and measuring it too late (as cycle time control does). Backlog itself is a leading indicator of the cycle time customers will be seeing, and it reflects not only imbalances between demand and the labor force but variations in labor force productivity that can by themselves cause cycle times to grow.

### **Sensitivity to system parameters**

We cannot know without testing whether the conclusions drawn in the previous section continue to apply in other demand environments or with other process parameters. In this section we report the results of such tests. We first vary the underlying parameters that govern the five demand environments and test which values of proportional and differential control gains continue to give acceptable performance under each of the three control regimes. Then we repeat the tests of all five demand environments and three control regimes for alternative values of three critical process parameters: averaging interval, staffing interval, and productivity variability. The results are summarized in Table 2. Each entry in this table represents the number of combinations of control gains (out of 100) in which the Profit Index exceeds 75 percent and the Lateness Index does not exceed 20 percent.

#### *Demand environments*

In four of the cases reported here we have made the demand environment more variable than in the base case. The exception is Random Demand, in which we reduced the already high variability. In the case of Step Demand, demand increases to 30 (versus 20 in the base case); for Ramp Demand the daily increase is 0.4 units per day (versus 0.2 in the base case); for Spike Demand the one-time demand is 150 units (versus 100 in the base case); for Random Demand the standard deviation of demand is 2.5 units (versus 5 in the base case); for Seasonal Demand the range is from 5 to 15 units (versus 7.5 to 12.5 in the base case).

None of the conclusions drawn in the previous section are contradicted by these results. Cycle time and demand control policies continue to perform poorly in the Step, Ramp, and Seasonal environments. Backlog control



Table 2. Sensitivity Analysis (percentage of acceptable cases)

	Base case	Demand environment	Staffing interval	Averaging interval	Productivity variability
Step demand					
Cycle time	10	5	11	10	8
Backlog	33	20	29	43	15
Demand	5	4	32	10	2
Ramp demand					
Cycle time	10	0	3	4	5
Backlog	15	9	24	32	23
Demand	4	2	28	14	2
Spike demand					
Cycle time	51	39	23	45	38
Backlog	68	62	38	65	52
Demand	60	40	80	0	40
Random demand					
Cycle time	42	52	29	45	45
Backlog	59	68	43	46	68
Demand	31	81	44	0	26
Seasonal demand					
Cycle time	2	0	6	16	3
Backlog	22	3	21	41	12
Demand	19	0	34	28	19

continues to be the best overall control regime, although as expected it performs less well under more stringent conditions. If we look over all five environments as we did in Figure 13, we see that backlog control performs adequately in four out of five environments for one combination of control gains, and this combination is included in those that gave adequate control in all demand environments in the base case. This result strengthens our conclusion that backlog control is the most robust of the control variants tested here.

#### *Process parameters*

Three aspects of our model business process can be expected to make the control problem more challenging: changing the labor force at discrete points in time, averaging the cycle time measure, and variability in the productivity of the labor force. We chose what we felt were realistic values for these three variables for the base case. We now test whether our conclusions are sensitive to these parameters, which in some cases could be changed by management. The results are summarized in Table 2.

In the base case we assumed that changes to the labor force could only be implemented every 30 days. In Table 2 we report comparable results for changes every five days. A shorter staffing interval has a uniformly positive effect on the performance of demand control, but mixed effects on cycle time

and backlog control. In the Step, Ramp, and Seasonal environments both cycle time and backlog control improve (or remain about the same); in the Spike and Random environments both do considerably worse. In contrast, the number of acceptable cases when demand is the control signal increases in all five demand environments. In fact, demand control has the highest percentage of acceptable cases of all three control regimes in all five demand environments when the staffing interval is short. However, as we pointed out earlier, demand control is inherently limited because it cannot recognize the existence of a large backlog in the system. In the results we report here it is fortuitous that the demand and process uncertainties are such that demand control appears to work fairly well. More substantial variations in worker productivity, for example, would lead to more frequent backlogs and poor system performance under demand control.

It is almost universal practice to measure process performance using averages, if only because the variation in cycle times or other process measures at the individual level is so large that it can obscure rather than illuminate trends. Accordingly, we set an averaging interval of 30 days in our base case. Table 2 shows the results for an averaging interval of five days. The effects of this change are to improve the performance of backlog and cycle time control regimes in several demand environments. In no case do these regimes perform noticeably worse. However, the demand regime worsens dramatically in the two demand environments (Spike and Random) in which it performed best in the base case. In fact, with a short averaging interval in these environments there is no combination of control gains that results in acceptable performance. As with the staffing interval, we see that demand control is most sensitive to the averaging interval. These results suggest that backlog control performs well whether one is measuring it over short or long periods of time.

The final parameter we examine is the variability in labor force productivity. In the base case we assumed a fairly high degree of variability within the process itself, since service sector processes (with human labor) are notoriously variable (Hopp and Spearman 1995). To determine how significant this assumption is to our conclusions we reduced the variability to zero. Again the results are given in Table 2. Somewhat surprisingly, this change to the model has very little overall effect on the performance of either cycle time or demand control, except in the case of Spike demand where reduced variability reduces the number of acceptable cases. The pattern with backlog control is more complex: in the Step, Spike, and Seasonal environments it performs worse with reduced variability, while in the Ramp and Random environments it performs better. It is difficult to find a simple explanation for this complex pattern of results. What does stand out is the somewhat unexpected finding that *reducing internal process variability does not result in uniformly better performance*. This suggestive observation, which merits further research, would appear to run counter to much of the recent emphasis within operations planning on reducing process variability (Hopp and Spearman 1995).

## Managerial implications

We noted at the outset that the Business Process Redesign movement has recommended the use of feedback control as part of good process management but, to our knowledge, has not spelled out in any detail how to implement an effective control regime. The literature suggests that management must decide three things in order to control processes:

- which aspects of the process to measure;
- what targets to set for those measurements;
- which resources to use to control the gap between target and actual performance.

Little information is available, however, on how these choices interact with each other, or on how the external and internal environments of the organization impact the choice of an effective control regime. Our research is a first step toward deepening our understanding of these issues and toward providing practical advice to managers.

Were we asked to advise a manager on how to design an effective control regime for a particular business process, we would now, on the basis of our research, raise a number of new and more detailed questions. For example:

- *What demand patterns are we likely to face?* We have shown that effective control depends on the demand pattern. Some industries face stable but random demands, others are cyclical. Our results suggest that some control policies are robust across many different types of demands, while others work well only in certain types of demand patterns. Also, since firms have many opportunities to change the demands they face (with, for example, pricing policies and other incentive programs), demand itself can be thought of as a resource (partly) under the control of management.
- *What aspects of internal process performance should we measure?* Any control regime must have a target to aim for, so management must decide which aspect or aspects of its business process to measure and use as the target. We have examined three generic process measures: demand, backlog, and cycle time. All processes have such measures, but rarely is thought given to the implications of focusing on one or the other. We have shown that the choice of a process measure itself can be a critical aspect of an effective control policy.
- *How should we measure these aspects?* Cycle time, demand and backlog themselves are simply concepts, not operational measures. To make such a measure operational, management must decide whether to monitor these aspects of performance every day, every week, or every month (not to mention whether to aggregate product types, customer types, and so on). We

have shown that this aspect of control also must be coordinated with the other choices.

- *How should we change resources when internal measures are off target?* We have found that a control policy that simply reacts to the *current* gap between actual performance and the target can be totally ineffectual in certain cases. Proportional control often has to be supplemented with differential control to be effective. This means tracking not just the current gap but the *change* in the gap over time. We have also found that appropriate values for the control gains for these two measures are often different from each other and vary depending on the control environment, which means that management must be prepared to weight these two gaps differently and to tie the choice of control gains to assumptions about demand. This is a level of sophistication we suspect few control regimes have achieved.

Substantive and rather general insights related to some of these questions have been developed in this article. We should like to stress, however, that it will be crucial for managers themselves to apply modeling and simulation to come up with solid answers to these questions in order to cope with the concrete situations they face. Besides general principles, an organization also needs to take into account its specific circumstances in order to deal effectively with the complexities that confront it.

The forgoing summarizes our current thinking on the problem of business process control. It also provides many suggestions for future research (see also Schwaninger, *et al.* 1999). To cite just one example, what control regime is best for maximizing profit while ensuring that the worst case cycle time (for example, the 95th percentile) does not exceed a certain level? Future research should examine the interrelated problems of how to process the available information on system performance and how to use these measures to best control the system. Optimization techniques, time series analysis and perhaps Kalman filtering, which are the subjects of our current research, have the potential to significantly improve the performance of business processes when incorporated in control regimes. Finally, another important research challenge is to extend the system dynamics-based design of control systems to the realm of strategic and normative management (Schwaninger 2000).

## References

- Allen A. 1978. *Probability, Statistics, and Queueing Theory*. Academic Press: New York.
- Bach V, Brecht L, Hess T, Oesterle H. 1996. *Enabling Systematic Business Change*. View eg: Wiesbaden.
- Davenport TH. 1993. *Process Innovation*. Harvard Business School Press: Boston.
- Davenport TH, Short JE. 1990. The new industrial Engineering: information technology and business process redesign. *Sloan Management Review* (Summer) 11–26.

- 
- Espejo R, Schuhmann W, Schwaninger M, Bilello U. 1996. *Organizational Transformation and Learning, A Cybernetic Approach to Management*. John Wiley & Sons, Ltd.: Chichester.
- Forrester J. 1961. *Industrial Dynamics*. Wiley: New York.
- Franklin G, Powell J, Emami-Naeini A. 1994. *Feedback Control of Dynamic Systems*. Addison-Wesley: Reading, MA.
- Hammer M. 1990. Reengineering work: don't automate, obliterate. *Harvard Business Review* July-August: 104–112.
- Hammer M, Champy J. 1993. *Reengineering the Corporation*. HarperCollins: New York.
- Hopp WJ, Spearman M. 1995. *Factory Physics*. Irwin: Chicago.
- Oesterle H. 1995. *Business in the Information Age*. Springer: Berlin.
- Schwaninger M, 2000. Managing complexity—the path towards intelligent organizations. *Systemic Practice and Action Research* 13(2): 207–241.
- Schwaninger M, Powell S, Trimble C. 1999. Modeling a Control System for Organizational Processes. *Proceedings of the 43rd Annual Conference of the International Society for the Systems Sciences*, Asilomar: California (forthcoming).
- Sterman J. 1989. Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. *Management Science* 35(3): 321–339.
- Sterman J. 2000. *Business Dynamics*. McGraw-Hill: Boston.