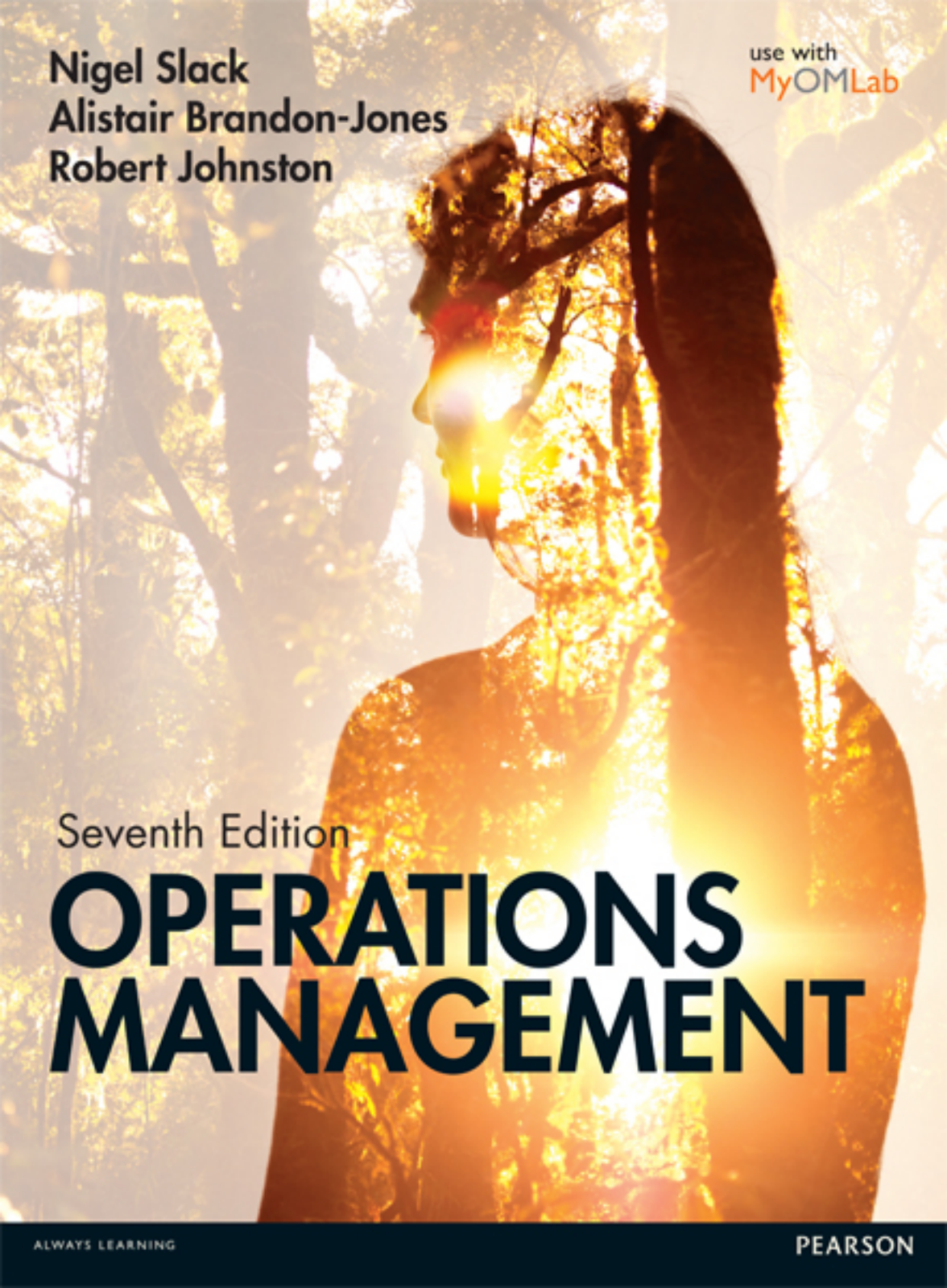




Nigel Slack
Alistair Brandon-Jones
Robert Johnston

use with
MyOMLab

Seventh Edition

OPERATIONS MANAGEMENT

Key questions

- › What is capacity management?
- › How is capacity measured?
- › What are the ways of coping with demand fluctuation?
- › How can operations plan their capacity level?
- › How is capacity planning a queuing problem?

INTRODUCTION

Providing sufficient capability to satisfy current and future demand is a fundamental responsibility of operations management. Get the balance right between the capacity of an operation and the demand it is subjected to and it can satisfy its customers cost-effectively. Get it wrong and it could both fail to satisfy demand and have excessive costs. Capacity planning and control is also sometimes referred to as *aggregate* planning and control. This is because, at this level of planning and control, demand and capacity calculations are usually performed on an aggregated basis which does not discriminate between the different products and services that an operation might produce. The essence of the task is to reconcile, at a general and

aggregated level, the supply of capacity with the level of demand which it must satisfy (see Fig. 11.1). There is also a supplement that deals with analytical queuing models, which is one way of considering capacity management, especially in some service operations.

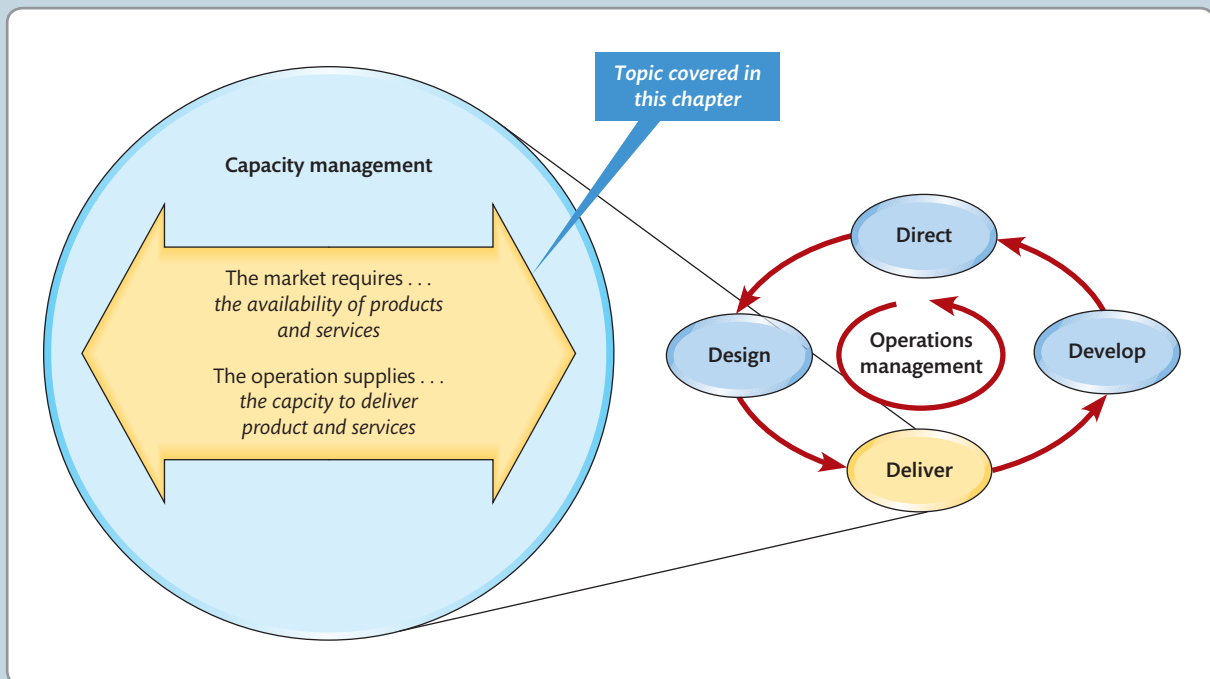


Figure 11.1 A definition of capacity planning and control

Amazon's 'Cyber Monday'¹

Founded as an online bookseller by Jeff Bezos, Amazon, now the world's biggest online retailer, started business in Europe in 1998. Since then the Seattle-based firm has experienced remarkable growth, employing around 40,000 people around the world and dominating a fiercely competitive consumer market, where its success is unquestionably based partly on its keen pricing. But low prices are not the only thing supporting Amazon's success. Without fast, accurate and efficient delivery it could not have secured its 'top e-retailer' position.

This is why Amazon devotes so much investment and effort into its fulfilment centres, customer service centres and software development centres across Europe, North America, Latin America and Asia.

These organize the shipment of millions of individual items, from bird tables to baby clothes. (Amazon says that the exact number of different items it sells is difficult to define, especially if you take into account the articles that are offered via the 'marketplace', and changes every day.) Typical of Amazon's shipment operations is its 46,000-square-metre Milton Keynes warehouse in the UK, one of eight in Britain. In the warehouse products are stocked within its extensive shelving arrangement, with the position of every item referenced using a portable satellite navigation system. Amazon says that it saves time when their staff retrieve items. *'The computer shows the shortest pick-path'*, said Arthur Valdez, the vice-president of Amazon's British operations. The item is then scanned and picked, after which it moves along a conveyor belt to be packed or gift-wrapped and then labelled. At this point an email is automatically sent to the customer informing them that their product is on its way.

Mr Valdez manages a network of fast-moving operations that must always maintain a tight control of its activities, but no time is more testing than the run-up to Christmas. The gift-buying habits of Western consumers mean that up to 40 per cent of annual sales value can come in the final three months of the year, with half of the multi-billion online Christmas sales taking place over the end of November and the first two weeks of December. The average number of articles being sold each day soars from 300,000 to, at its peak, 3.6 million being sold in one day. In the UK, this day – which it calls



Source: Alamy Images/M405 Photos

'Cyber Monday' – is at the beginning of December; or, to be more precise, 9 pm on that day, when shoppers, having normally been paid for the month and having spent the weekend browsing the high street, return from work to begin their Christmas shopping in earnest. It makes for a hectic time. *'A full truck is dispatched every 3 minutes and 24 seconds on our busiest trading day'*, says Mr Valdez. But careful forecasting can at least stop the Christmas peak being a surprise. And careful monitoring of customer behaviour has revealed a further trend – after 'Cyber Monday', now comes 'Boomerang Thursday', when customers start to return their unwanted items. *'As the online retail sector continues to grow, so too has consumer demand and confidence to return items, often before Christmas'*, says Mark Lewis, chief executive of CollectPlus, which allows customers to return items to a local convenience store. *'This suits retailers. They want to get [items] back as soon as possible, so they can sell them on.'* Mark Lewis says that half of his customers return items at off-peak times. *'It peaks at 7 pm. It reflects how we live our lives these days.'*

However, some retail analysts believe that the advance of technology in the form of mobile phone transactions and broadband has also meant the significance of 'Cyber Monday' and 'Boomerang Thursday' will diminish because such technology makes it easier to stagger transactions. But for Mr Valdez, it is continual vigilance that allows Amazon to keep up with demand trends. *'Every year it feels like [Christmas starts on] 1 January. We are all year long focused on understanding the lessons learnt from the previous Christmas'*, he says.

WHAT IS CAPACITY MANAGEMENT?

The most common use of the word capacity is in the static, physical sense of the fixed *volume* of a container, or the space in a building. This meaning of the word is also sometimes used by operations managers. For example: a pharmaceutical manufacturer may invest in new 1,000-litre capacity reactor vessels; a property company purchases a 500-vehicle capacity city-centre car park; and a 'multiplex' cinema is built with 10 screens and a total capacity of 2,500 seats. Although these capacity measures describe the *scale* of these operations, they do not reflect the processing capacities of these investments. To do this we must incorporate a *time* dimension appropriate to the use of assets. So the pharmaceutical company will be concerned with the level of output that can be achieved using the 1,000-litre reactor vessel. If a batch of standard products can be produced every hour, the planned processing capacity could be as high as 24,000 litres per day. If the reaction takes four hours, and two hours are used for cleaning between batches, the vessel may only produce 4,000 litres per day. Similarly, the car park may be fully occupied by office workers during the working day, 'processing' only 500 cars per day. Alternatively, it may be used for shoppers staying on average only one hour, and theatre-goers occupying spaces for three hours, in the evening. The processing capacity would then be up to 5,000 cars per day. Thus the definition of the capacity of an operation is the *maximum level of value-added activity over a period of time* that the process can achieve under normal operating conditions.

Capacity constraints

Many organizations operate at below their maximum processing capacity, either because there is insufficient demand completely to 'fill' their capacity, or as a deliberate policy, so that

* Operations principle

Any measure of capacity should reflect the ability of an operation or process to supply demand.

the operation can respond quickly to every new order. Often, though, organizations find themselves with some parts of their operation operating below their capacity while other parts are at their capacity 'ceiling'. It is the parts of the operation that are operating at their capacity 'ceiling' which are the capacity constraint for the whole operation. For example, a retail superstore might offer a gift-wrapping

service which at normal times can cope with all requests for its services without delaying customers unduly. At Christmas, however, the demand for gift-wrapping might increase proportionally far more than the overall increase in custom for the store as a whole. Unless extra resources are provided to increase the capacity of this micro-operation, it could constrain the capacity of the whole store.

Planning and controlling capacity

Capacity planning and control is the task of setting the effective capacity of the operation so that it can respond to the demands placed upon it. This usually means deciding how the operation should react to fluctuations in demand. We have faced this issue before (in Chapter 6) where we examined long-term changes in demand and the alternative capacity strategies for dealing with the changes. These strategies were concerned with introducing (or deleting) major increments of physical capacity. We called this task long-term capacity strategy. Here we are treating the shorter timescale where capacity decisions are being made largely within the constraints of the physical capacity limits set by the operation's long-term capacity strategy.

Medium- and short-term capacity

Having established long-term capacity, operations managers must decide how to adjust the capacity of the operation in the medium term. This usually involves an assessment of the demand forecasts over a period of 2–18 months ahead, during which time planned output can be varied; for example, by changing the number of hours the equipment is used. In practice, however, few forecasts are accurate, and most operations also need to respond to changes

in demand which occur over a shorter timescale. Hotels and restaurants have unexpected and apparently random changes in demand from night to night, but also know from experience that certain days are on average busier than others. So operations managers also have to make short-term capacity adjustments, which enable them to flex output for a short period, either on a predicted basis (for example, bank checkouts are always busy at lunchtimes) or at short notice (for example, a sunny warm day at a theme park).

* Operations principle

Capacity management decisions should reflect both predictable and unpredictable variations in capacity and demand.

Aggregate demand and capacity

The important characteristic of capacity management, as we are treating it here, is that it is concerned with setting capacity levels over the medium and short terms in aggregated terms. That is, it is making overall, broad capacity decisions, but is not concerned with all of the detail of the individual products and services offered. This is what 'aggregated' means – different products and services are bundled together in order to get a broad view of demand and capacity. This may mean some degree of approximation, especially if the mix of products or services being produced varies significantly (as we shall see later). Nevertheless, as a first step in management, aggregation is necessary. For example, a hotel might think of demand and capacity in terms of 'room nights per month'; this ignores the number of guests in each room and their individual requirements, but it is a good first approximation. A woollen knitwear factory might measure demand and capacity in the number of units (garments) it is capable of making per month, ignoring size, colour or style variations. Aluminium producers could use tonnes per month, ignoring types of alloy, gauge and batch size variation. The ultimate aggregation measure is money. For example, retail stores, who sell an exceptionally wide variety of products, use revenue per month, ignoring variation in spend, number of items bought, the gross margin of each item and the number of items per customer transaction. If all this seems very approximate, remember that most operations have sufficient experience of dealing with aggregated data to find it useful.

The objectives of capacity management

The decisions taken by operations managers in devising their capacity plans will affect several different aspects of performance:

- **Costs** will be affected by the balance between capacity and demand (or output level if that is different). Capacity levels in excess of demand could mean under-utilization of capacity and therefore high unit costs.
- **Revenues** will also be affected by the balance between capacity and demand, but in the opposite way. Capacity levels equal to or higher than demand at any point in time will ensure that all demand is satisfied and no revenue lost.
- **Working capital** will be affected if an operation decides to build up finished goods inventory prior to demand. This might allow demand to be satisfied, but the organization will have to fund the inventory until it can be sold.
- **Quality** of goods or services might be affected by a capacity plan which involved large fluctuations in capacity levels, by hiring temporary staff for example. The new staff and the disruption to the routine working of the operation could increase the probability of errors being made.
- **Speed** of response to customer demand could be enhanced, either by the build-up of inventories (allowing customers to be satisfied directly from the inventory rather than having to wait for items to be manufactured) or by the deliberate provision of surplus capacity to avoid queuing.
- **Dependability** of supply will also be affected by how close demand levels are to capacity. The closer demand gets to the operation's capacity ceiling, the less able it is to cope with any unexpected disruptions and the less dependable its deliveries of goods and services could be.
- **Flexibility**, especially volume flexibility, will be enhanced by surplus capacity. If demand and capacity are in balance, the operation will not be able to respond to any unexpected increase in demand.

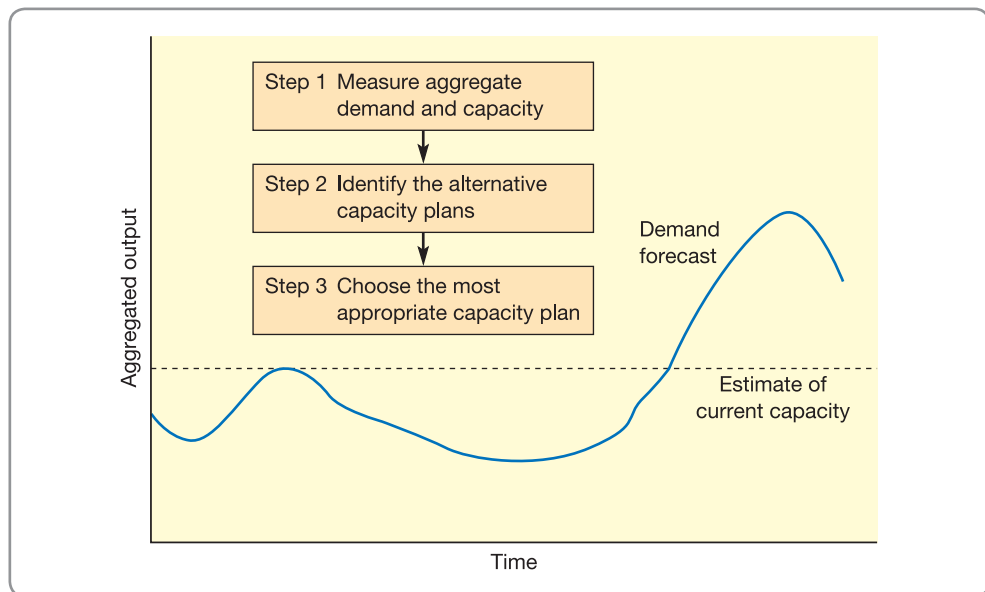


Figure 11.2 The steps in capacity management

The steps of capacity management

The sequence of capacity management decisions which need to be taken by operations managers is illustrated in Figure 11.2. Typically, operations managers are faced with a forecast of demand which is unlikely to be either certain or constant. They will also have some idea of their own ability to meet this demand. Nevertheless, before any further decisions are taken, they must have quantitative data on both capacity and demand. So the first step will be to *measure the aggregate demand and capacity* levels for the planning period. The second step will be to *identify the alternative capacity plans* which could be adopted in response to the demand fluctuations. The third step will be to *choose the most appropriate capacity plan* for their circumstances.

HOW IS CAPACITY MEASURED?

Forecasting demand fluctuations

Although demand forecasting is usually the responsibility of the sales and/or marketing functions, it is a very important input into the capacity management decision, and so is of interest to operations managers. After all, without an estimate of future demand it is not possible to plan effectively for future events, only to react to them. It is therefore important to understand the basis and rationale for these demand forecasts. (See the supplement on forecasting after Chapter 6.) As far as capacity planning and control is concerned, there are three requirements from a demand forecast:

- *It is expressed in terms which are useful for capacity management.* If forecasts are expressed only in money terms and give no indication of the demands that will be placed on an operation's capacity, they will need to be translated into realistic expectations of demand, expressed in the same units as the capacity (for example, machine hours per year, operatives required, space, etc.).
- *It is as accurate as possible.* In capacity management, the accuracy of a forecast is important because, whereas demand can change instantaneously, there is a lag between deciding to change capacity and the change taking effect. Thus many operations managers are faced with a dilemma. In order to attempt to meet demand, they must often decide output in advance, based on a forecast which might change before the demand occurs, or worse, prove not to reflect actual demand at all.

- *It gives an indication of relative uncertainty.* Decisions to operate extra hours and recruit extra staff are usually based on forecast levels of demand, which could in practice differ considerably from actual demand, leading to unnecessary costs or unsatisfactory customer service. For example, a forecast of demand levels in a supermarket may show initially slow business that builds up to a lunchtime rush. After this, demand slows, only to build up again for the early evening rush, and it finally falls again at the end of trading. The supermarket manager can use this forecast to adjust (say) checkout capacity throughout the day. But although this may be an accurate average demand forecast, no single day will exactly conform to this pattern. Of equal importance is an estimate of how much actual demand could differ from the average. This can be found by examining demand statistics to build up a distribution of demand at each point in the day. The importance of this is that the manager now has an understanding of when it will be important to have reserve staff, perhaps filling shelves, but still on call to staff the checkouts should demand warrant it. Generally, the advantage of probabilistic forecasts such as this is that it allows operations managers to make a judgement between possible plans that would virtually guarantee the operation's ability to meet actual demand, and plans that minimize costs. Ideally, this judgement should be influenced by the nature of the way the business wins orders: price-sensitive markets may require a risk-avoiding cost-minimization plan that does not always satisfy peak demand, whereas markets that value responsiveness and service quality may justify a more generous provision of operational capacity.

Seasonality of demand

Most markets are influenced by some kind of seasonality – that means that they vary depending on the time of year. Sometimes the causes of seasonality are climatic (holidays), sometimes festive (gift purchases), sometimes financial (tax processing), or social, or political; in fact there are many factors that affect the volume of activity in everything from construction materials to clothing, from health care to hotels. It may be demand seasonality or supply seasonality, but in many organizations, capacity management is largely about coping with these seasonal fluctuations. These fluctuations in demand or supply may be reasonably forecastable, but some are usually also affected by unexpected variations in the weather and by changing economic conditions.

Consider the four different types of operation described previously: a woollen knitwear factory, a city hotel, a supermarket and an aluminium producer. Their demand patterns are shown in Figure 11.3. The woollen knitwear business and the city hotel both have seasonal sales demand patterns, but for different reasons: the woollen knitwear business because of climatic patterns (cold winters, warm summers) and the hotel because of demand from business people, who take vacations from work at Christmas and in the summer. The retail supermarket is a little less seasonal, but is affected by pre-vacation peaks and reduced sales during vacation periods. The aluminium producer shows virtually no seasonality, but is showing a steady growth in sales over the forecast period.

Weekly and daily demand fluctuations

Seasonality of demand occurs over a year, but similar predictable variations in demand can also occur for some products and services on a shorter cycle. The daily and weekly demand patterns of a supermarket will fluctuate, with some degree of predictability. Demand might be low in the morning, higher in the afternoon, with peaks at lunchtime and after work in the evening. Demand might be low on Monday and Tuesday, build up during the latter part of the week and reach a peak on Friday and Saturday. Banks, public offices, telephone sales organizations and electricity utilities all have weekly and daily, or even hourly, demand patterns which require capacity adjustment. The extent to which an operation will have to cope with very short-term demand fluctuations is partly determined by how long its customers are prepared to wait for their products or services. An operation whose customers are incapable of waiting, or unwilling to wait, will have to plan for very short-term demand fluctuations. Emergency services, for example, will need to understand the hourly variation in the demand for their services and plan capacity accordingly.

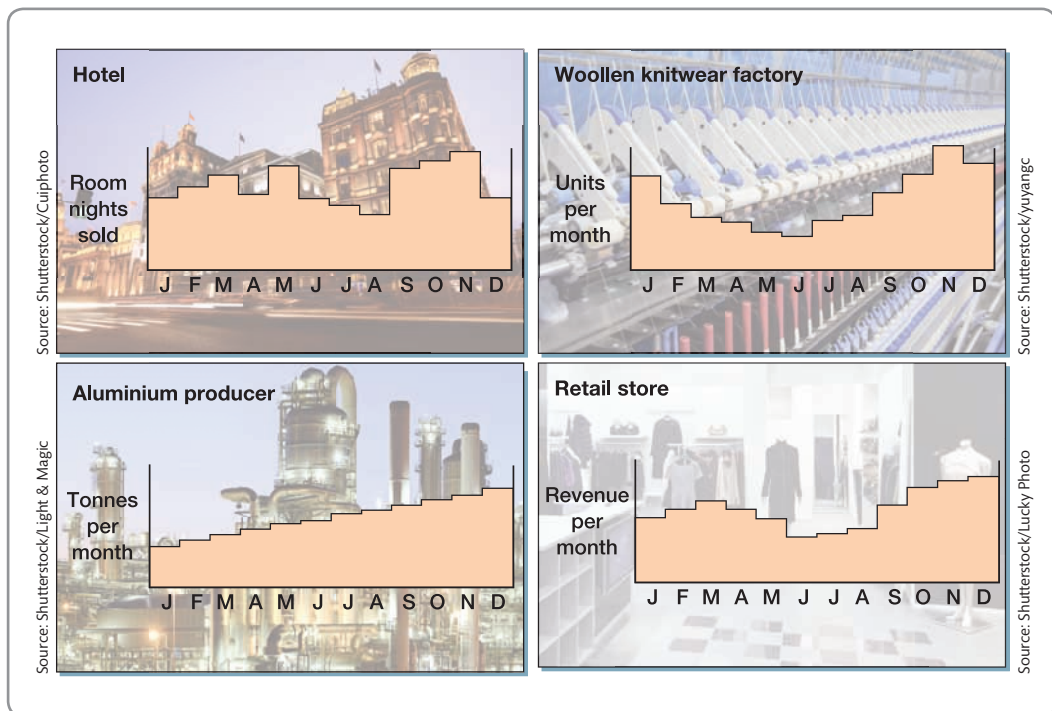


Figure 11.3 Aggregate demand fluctuations for four organizations

SHORT CASE

Raining on their parade²

Some fluctuations in demand are predictable, others far less so. Amongst the most difficult to forecast, especially in the medium to long term that is of interest to business, is the weather. And many services and products are profoundly affected by the weather. Theme parks, garden centres, sunglasses, sunscreen, waterproof clothing and ice-cream are all obvious examples. Yet the range of operations interested in weather forecasting has expanded significantly. Energy utilities, soft drink producers and fresh food producers and retailers are all keen to purchase the latest weather forecasts. But so are operations such as banking call centres and mobile phone operators. It would appear that the demand for telephone banking falls dramatically when the sun shines, as does the use of mobile phones. A motorway catering group were surprised to find that their sales of hot meals fell predictably by €110,000 per day for each degree temperature rise above 20°C. Similarly, insurance companies have found it wise to sell their products when the weather is poor and likely customers are trapped indoors rather than relaxing outside in the sun, refusing to worry about the future. In the not-for-profit sector new understanding is being developed about the link between various illnesses and temperature. Here



Source: AL RF (PhotoDisc/Don Farrell)

temperature is often used as a predictor of demand. So, for example, coronary thrombosis cases peak two days after a drop in temperature, for strokes the delay is around five days, while deaths from respiratory infections peak twelve days after a temperature drop. Knowing this, hospital managers can plan for changes in their demand.

SHORT CASE

The London Eye

The London Eye is the world's largest observation wheel and one of the UK's most spectacular tourist attractions. The 32 passenger capsules, fixed on the perimeter of the 135-metre diameter rim, each hold 25 people. The wheel rotates continuously, so entry requires customers to step into the capsules which are moving at 0.26 metres per second, which is a quarter of normal walking speed. One complete 360-degree rotation takes 30 minutes, at the end of which the doors open and passengers disembark. Boarding and disembarkation are separated on the specially designed platform which is built out over the river. The attraction has a 'timed admissions booking system' (TABS) for both individual and group bookings. This allocates requests for 'flights' on the basis of half-hour time slots. At the time of writing, the London Eye is open every day except Christmas Day. Admission is from 10.00 am to 9.30 pm (for the 9.30 to 10.00 pm slot) in the summer, from the beginning of April to mid-September. For the rest of the year, the winter season, admission begins at 10.00 am, and last admissions are for the 5.30 to 6.00 pm slot. The London Eye forecasts anticipated that 2.2 million passengers would fly the London Eye in 2000, excluding January, which was reserved for final testing and admission of



Source: AL RF (Imagestate/Michael Duerinckx)

invited guests only. An early press release told journalists that the London Eye would rotate an average of 6,000 revolutions per year.

COPING WITH DEMAND FLUCTUATION

With an understanding of both demand and capacity, the next step is to consider the alternative methods of responding to demand fluctuations. There are three 'pure' options available for coping with such variation:

* Operations principle

Capacity planning will use some mix of three 'pure' approaches, level capacity, chase demand, and demand management.

- Ignore the fluctuations and keep activity levels constant (level capacity plan).
- Adjust capacity to reflect the fluctuations in demand (chase demand plan).
- Attempt to change demand to fit capacity availability (demand management).

In practice, most organizations will use a mixture of all of these 'pure' plans, although often one plan might dominate. The short case 'Panettone: how Italy's bakers cope with seasonal demand' describes how one operation pursues some of these options.

Level capacity plan

In a level capacity plan, the processing capacity is set at a uniform level throughout the planning period, regardless of the fluctuations in forecast demand. This means that the same number of staff operate the same processes and should therefore be capable of producing the same aggregate output in each period. Where non-perishable materials are processed, but not immediately sold, they can be transferred to finished goods inventory in anticipation of sales at a later time. Thus this plan is feasible (but not necessarily desirable) for our examples of the woollen knitwear company and the aluminium producer (see Fig. 11.5).

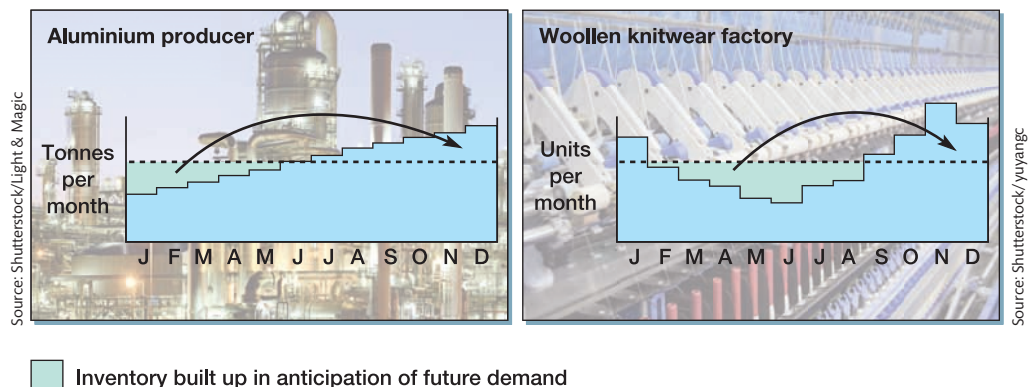


Figure 11.5 Level capacity plans which use anticipation inventory to supply future demand

SHORT CASE

Panettone: how Italy's bakers cope with seasonal demand³

Panettone has become a national symbol of the Italian Christmas. The light and fluffy dome-shaped confection is dotted with sultanas and candied citrus peel, and is *the* Italian Christmas cake. Traditionally made in Milan, Italy, about 40 million of them are consumed throughout Italy over the holiday period. Now, they are becoming popular around the world. Over a million are exported to the US, while an endorsement from Delia Smith, a celebrity chef, caused a surge in demand in Britain with a well-publicized recipe for trifle made with panettone. This boost to production is good news for the big Italian manufacturers, but although volumes are higher, the product is still seasonal, which poses a problem for even the experienced Milanese confectioners. Smaller 'artisan' producers simply squeeze a few batches of panettoni into their normal baking schedules as Christmas approaches. But for the large industrial producers who need to make millions for the Christmas season, it is not possible. And no panettone manufacturer is larger than the Bauli group. It is one of the foremost manufacturers of confectionery in Europe. Founded over 70 years ago, and in spite of its mass production approach, it has a reputation for quality and technological improvement. The company's output of panettone accounts for 38 per cent of Italian sales. The key to its success, according to the company, is in having '*combined the skill of home-made recipes with high technology [and] quality guaranteed by high standards that are unattainable in craftsman production, but that can only be reached by selecting top quality raw materials, by thousands of tests and checks on the entire production line and the production process*'. In fact, the company says that its size is an advantage. '*High*



investment in research and technology allows us to manage natural fermentation and guarantee a uniform quality that artisanal bakeries find hard to achieve.'

In fact, although Bauli has diversified into year-round products like croissants and biscuits, it has acquired a leadership role in the production of products for festive occasions. Seasonal cakes account for over 50 per cent of its turnover of around €420 million. And so successful has it been in its chosen markets that in 2009 it bought Motta and Alemagna, the two big Milanese brands that pioneered the manufacture of panettone. So how does Bauli cope with such seasonality? Partly it is by hiring

large numbers of temporary seasonal workers to staff its dedicated production lines. At peak times there can be 1,200 seasonal workers in the factory, in comparison to its permanent staff of around 800. It also starts to build up inventories of panettone before demand begins to increase for the Christmas peak. Production of panettone lasts about four months, starting in September. 'Attention

to ingredients and the use of new technologies in production give a shelf-life of five months without preservatives', says Michele Bauli, deputy chairman, who comes from the firm's founding family. Temporary workers are also hired to bake other seasonal cakes such as the *colomba*, a dove-shaped Easter treat, which keeps them occupied for a month and a half in the spring.

Level capacity plans of this type can achieve the objectives of stable employment patterns, high process utilization, and usually also high productivity with low unit costs. Unfortunately, they can also create considerable inventory which has to be financed and stored. Perhaps the biggest problem, however, is that decisions have to be taken as to what to produce for inventory rather than for immediate sale. Will green woollen sweaters knitted in July still be fashionable in October? Could a particular aluminium alloy in a specific sectional shape still be sold months after it has been produced? Most firms operating this plan, therefore, give priority to only creating inventory where future sales are relatively certain and unlikely to be affected by changes in fashion or design. Clearly, such plans are not suitable for 'perishable' products, such as foods and some pharmaceuticals, for products where fashion changes rapidly and unpredictably (for example, fashion garments), or for customized products.

A level capacity plan could also be used by the hotel and supermarket, although this would not be the usual approach of such organizations, because it usually results in a waste of staff resources, reflected in low productivity. Because service cannot be stored as inventory, a level capacity plan would involve running the operation at a uniformly high level of capacity availability. The hotel would employ sufficient staff to service all the rooms, to run a full restaurant, and to staff the reception even in months when demand was expected to be well below capacity. Similarly, the supermarket would plan to staff all the checkouts, warehousing operations, and so on, even in quiet periods (see Fig. 11.6).

Low utilization can make level capacity plans prohibitively expensive in many service operations, but may be considered appropriate where the opportunity costs of individual lost sales are very high; or example, in the high-margin retailing of jewellery and in (real) estate agents. It is also possible to set the capacity somewhat below the forecast peak demand level in order to reduce the degree of under-utilization. However, in the periods where demand is expected to exceed planned capacity, customer service may deteriorate. Customers may have to queue for long periods or may be 'processed' faster and less

* Operations principle

The higher the base level of capacity, the less capacity fluctuation is needed to satisfy demand.

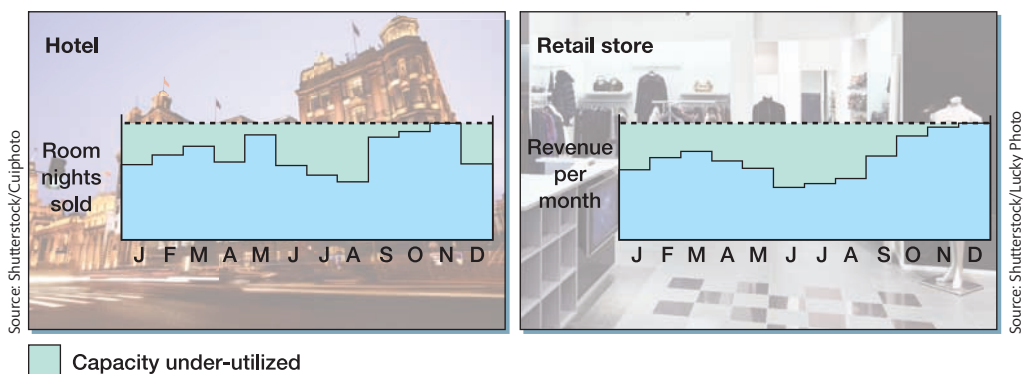


Figure 11.6 Level capacity plans with under-utilization of capacity

sensitively. While this is obviously far from ideal, the benefits to the organization of stability and productivity may outweigh the disadvantages of upsetting some customers.

Chase demand plan

The opposite of a level capacity plan is one which attempts to match capacity closely to the varying levels of forecast demand. This is much more difficult to achieve than a level capacity plan, as different numbers of staff, different working hours, and even different amounts of equipment may be necessary in each period. For this reason, pure chase demand plans are unlikely to appeal to operations which manufacture standard, non-perishable products. Also, where manufacturing operations are particularly capital-intensive, the chase demand policy would require a level of physical capacity, all of which would only be used occasionally. It is for this reason that such a plan is less likely to be appropriate for the aluminium producer than for the woollen garment manufacturer (see Fig. 11.7). A pure chase demand plan is more usually adopted by operations which cannot store their output, such as customer-processing operations or manufacturers of perishable products. It avoids the wasteful provision of excess staff that occurs with a level capacity plan, and yet should satisfy customer demand throughout the planned period. Where output can be stored, the chase demand policy might be adopted in order to minimize or eliminate finished goods inventory.

Sometimes it is difficult to achieve very large variations in capacity from period to period. If the changes in forecast demand are as large as those in the hotel example (see Fig. 11.8), significantly different levels of staffing will be required throughout the year. This would mean

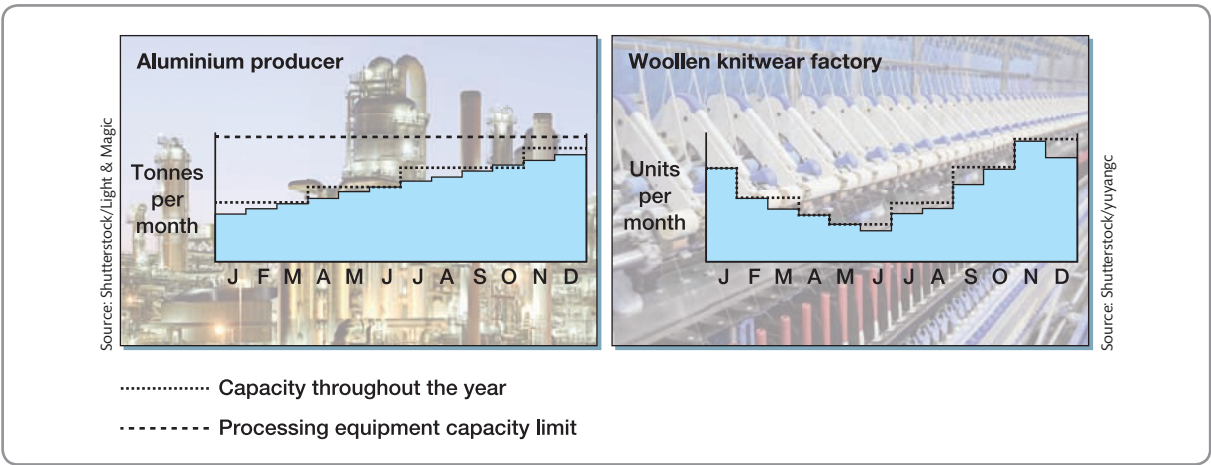


Figure 11.7 Chase demand capacity plans with changes in capacity which reflect changes in demand



Figure 11.8 Chase demand capacity plans with changes in capacity which reflect changes in demand

employing part-time and temporary staff, requiring permanent employees to work longer hours, or even bringing in contract labour. The operations managers will then have the difficult task of ensuring that quality standards and safety procedures are still adhered to, and that the customer service levels are maintained.

Methods of adjusting capacity

The chase demand approach requires that capacity is adjusted by some means. There are a number of different methods for achieving this, although they may not all be feasible for all types of operation. Some of these methods are listed below.

Overtime and idle time Often the quickest and most convenient method of adjusting capacity is by varying the number of productive hours worked by the staff in the operation. When demand is higher than nominal capacity, overtime is worked, and when demand is lower than nominal capacity the amount of time spent by staff on productive work can be reduced. In the latter case, it may be possible for staff to engage in some other activity such as cleaning or maintenance. This method is only useful if the timing of the extra productive capacity matches that of the demand. For example, there is little to be gained in asking a retail operation's staff to work extra hours in the evening if all the extra demand is occurring during their normal working period. The costs associated with this method are either the extra payment which is normally necessary to secure the agreement of staff to work overtime, or in the case of idle time, the costs of paying staff who are not engaged in direct productive work. Further, there might be costs associated with the fixed costs of keeping the operation heated, lit and secure over the extra period staff are working. There is also a limit to the amount of extra working time which any workforce can deliver before productivity levels decrease. Annualized hours approaches, as described in the short case 'Annualized hours help Lowwaters to retain its core team', are one way of flexing working hours without excessive extra costs.

Varying the size of the workforce If capacity is largely governed by workforce size, one way to adjust it is to adjust the size of the workforce. This is done by hiring extra staff during periods of high demand and laying them off as demand falls, or hire and fire. However, there are cost and ethical implications to be taken into account before adopting such a method. The costs of hiring extra staff include those associated with recruitment, as well as the costs of low productivity while new staff go through the learning curve. The costs of lay-off may include possible severance payments, but might also include the loss of morale in the operation and loss of goodwill in the local labour market. At a micro-operation level, one method of coping with peaks in demand in one area of an operation is to build sufficient flexibility into job design and job demarcation so that staff can transfer across from less busy parts of the operation. For example, the French hotel chain Novotel has trained some of its kitchen staff to escort customers from the reception area up to their rooms. The peak times for registering new customers coincide with the least busy times in the kitchen and restaurant areas.

Using part-time staff A variation on the previous strategy is to recruit part-time staff; that is, for less than the normal working day. This method is extensively used in service operations such as supermarkets and fast-food restaurants but is also used by some manufacturers to staff an evening shift after the normal working day. However, if the fixed costs of employment for each employee, irrespective of how long he or she works, are high, then using this method may not be worthwhile.

Subcontracting In periods of high demand, an operation might buy capacity from other organizations, called subcontracting. This might enable the operation to meet its own demand without the extra expense of investing in capacity which will not be needed after the peak

in demand has passed. Again, there are costs associated with this method. The most obvious one is that subcontracting can be very expensive. The subcontractor will also want to make sufficient margin out of the business. A subcontractor may not be as motivated to deliver on time or to the desired levels of quality. Finally, there is the risk that the subcontractors might themselves decide to enter the same market.

* Operations principle

There are always costs, as well as benefits, associated with changing capacity levels.

Critical commentary

To many, the idea of fluctuating the workforce to match demand, either by using part-time staff or by hiring and firing, is more than just controversial. It is regarded as unethical. It is any business's responsibility, they argue, to engage in a set of activities which are capable of sustaining employment at a steady level. Hiring and firing merely for seasonal fluctuations, which can be predicted in advance, is treating human beings in a totally unacceptable manner. Even hiring people on a short-term contract, in practice, leads to them being offered poorer conditions of service and leads to a state of permanent anxiety as to whether they will keep their jobs. On a more practical note, it is pointed out that, in an increasingly global business world where companies may have sites in different countries, those countries that allow hiring and firing are more likely to have their plants 'downsized' than those where legislation makes this difficult.

Manage demand plan

The most obvious mechanism of demand management is to change demand through price. Although this is probably the most widely applied approach in demand management, it is less common for products than for services. For example, some city hotels offer low-cost 'city break' vacation packages in the months when fewer business visitors are expected. Skiing and camping holidays are cheapest at the beginning and end of the season and are particularly expensive during school vacations. Ice-cream is 'on offer' in many supermarkets during the winter. The objective is invariably to stimulate off-peak demand and to constrain peak demand, in order to smooth demand as much as possible. Organizations can also attempt to increase demand in low periods by appropriate advertising. For example, turkey farmers in the UK and the USA make vigorous attempts to promote their products at times other than Christmas and Thanksgiving.

SHORT CASE

Annualized hours help Lowaters to retain its core team⁴

Lowaters Nursery is a garden plant and horticulture specialist in the south of England. Like any business that depends on seasonal weather conditions, it faces fluctuating demand for its services and products. It also prides itself on offering *'the best service in partnership with our customers, by communicating in a friendly professional manner and listening to our customers to provide the result required'* (Lowaters mission statement). But to maintain its quality of service throughout the seasonal ups and down in workload means keeping your core team happy and employed throughout the year. This is why Lowaters introduced its annualized hours scheme.



Source: Alistair Brandon-Jones

An annual hours work plan is a method of fluctuating capacity as demand varies throughout the year, without many of the costs associated with overtime or hiring temporary staff. It involves staff contracting to work a set number of hours per year rather than a set number of hours per week. The advantage of this is that the amount of staff time available to an organization can be varied throughout the year to reflect the real state of demand. Annual hours plans can also be useful when supply varies throughout the year.

Lowaters specialize in ornamental plants and employ around 25 people – most of them have been with the company for a long time and the company has a relatively low staff turnover of about 9 per cent per annum. Maria Fox, one of the management team at Lowaters, says that annualized hours give the company several advantages. *'It simplifies administration and gives us the flexibility we need to run the business while delivering some real advantages to the employees. They are all effectively on salary with fixed monthly payments. We can flex the hours worked over the year – when we are busy we work longer and when things are quiet, in the winter, they can take time off. Everyone other than directors is contracted to work 39 hours on average over 52 weeks of the year. This gives us a total of 2,028 hours that are available. In fact, supervisors are contracted to do a hundred more hours to cover planning and paperwork, making a total of 2,128.'*

The company created a simple spreadsheet that sets out the actual hours worked and compares them with a target distribution of the annualized hours expected to be worked over the year. This allows employees to see at a glance whether someone is over or under target. *'We email them a copy of their sheet at the beginning of the year so they can keep track of their own progress as they go',* says Maria. *'It also allows us to keep track of how many hours they do. If at the end of the year they come in plus or minus 50 hours we simply adjust it up or down for the next year. If there is a bigger discrepancy than that we'll look at the job structure – or some retraining. We end up with more control over our wage budget and we can flex staff up and down according to weather and workload.'* It also gives flexibility to staff, claims Lowaters. Their salary doesn't vary from month to month and they can deal with time off for sickness because they have the whole year to make up their hours.

However, not all experiments with annualized hours have been as successful as that at Lowaters. In cases where demand is very unpredictable, staff can be asked to come in to work at very short notice. This can cause considerable disruption to social and family life. For example, at one news broadcasting company, the scheme caused problems. Journalists and camera crew who went to cover a foreign crisis found that they had worked so many hours, they were asked to take the whole of one month off to compensate. Since they had no holiday plans, many would have preferred to work.

Alternative products and services

Sometimes, a more radical approach is required to fill periods of low demand, such as developing alternative products or services which can be produced on existing processes, but have different demand patterns throughout the year (see the short case 'Getting the message' for an example of this approach). Most universities fill their accommodation and lecture theatres with conferences and company meetings during vacations. Ski resorts provide organized mountain activity holidays in the summer. Some garden tractor companies in the US now make snow movers in the autumn and winter. The apparent benefits of filling capacity in this way must be weighted against the risks of damaging the core product or service, and the operation must be fully capable of serving both markets. Some universities have been criticized for providing sub-standard, badly decorated accommodation which met the needs of impecunious undergraduates, but which failed to impress executives at a trade conference.

Mixed plans

Each of the three 'pure' plans is applied only where its advantages strongly outweigh its disadvantages. For many organizations, however, these 'pure' approaches do not match their required combination of competitive and operational objectives. Most operations managers are required simultaneously to reduce costs and inventory, to minimize capital investment, and yet to provide a responsive and customer-orientated approach at all times. For this reason, most organizations choose to follow a mixture of the three approaches. This can be best illustrated by the woollen knitwear company example (see Fig. 11.9). Here some of the peak

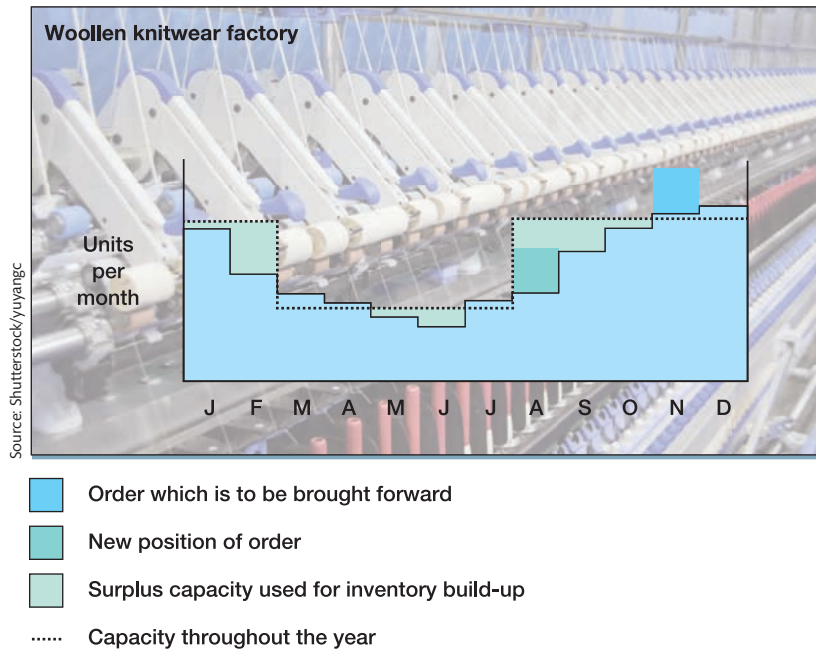


Figure 11.9 A mixed capacity plan for the woollen knitwear factory

demand has been brought forward by the company offering discounts to selected retail customers (manage demand plan). Capacity has also been adjusted at two points in the year to reflect the broad changes in demand (chase demand plan). Yet the adjustment in capacity is not sufficient to avoid totally the build-up of inventories (level capacity plan).

Yield management⁵

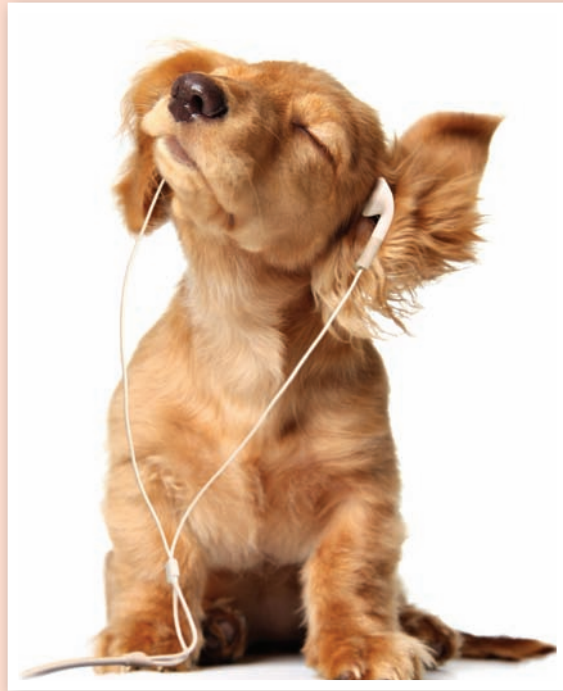
In operations which have relatively fixed capacities, such as airlines and hotels, it is important to use the capacity of the operation to maximize its potential to generate profit. One approach used by such operations is called yield management. It is really a variety of methods and analytical tools. The term is used in many service operations to mean techniques that can be used to allocate limited resources, among different categories of customers, such as business or leisure travellers. Because these techniques are used by operations with services that cannot be stored, yield management is sometimes called 'perishable asset revenue management' or simply 'revenue management'. But whatever name it goes by, the basic concept of yield management is based on the economic principle of supply and demand. When supplies are short, prices go up; when supply is high, prices go down. Yield management simply provides a systematic method for positioning customers within the supply-demand spectrum in such a way that they can obtain the highest yield for their services or products. So, a customer who has relatively little flexibility in his or her travel plans is the customer who is most likely to pay a higher price for airline tickets and hotel rooms. The customer with a great deal of flexibility is not as inclined to pay a higher price. Yield management is especially useful where:

- capacity is relatively fixed;
- the market can be fairly clearly segmented;
- the service cannot be stored in any way;
- the services are sold in advance;
- the marginal cost of making a sale is relatively low.

Companies which traditionally operate in seasonal markets can demonstrate some considerable ingenuity in their attempts to develop counter-seasonal products. One of the most successful industries in this respect has been the greetings card industry. Mother's Day, Father's Day, Halloween, Valentine's Day and other occasions have all been promoted as times to send (and buy) appropriately designed cards. Now, having run out of occasions to promote, greetings card manufacturers have moved on to 'non-occasion' cards, which can be sent at any time. These have the considerable advantage of being less seasonal, thus making the companies' seasonality less marked.

Hallmark Cards, one of the best-known card producers, has been the pioneer in developing non-occasion cards. Their cards include those intended to be sent from a parent to a child with messages such as 'Would a hug help?', 'Sorry I made you feel bad', and 'You're perfectly wonderful – it's your room that's a mess'. Other cards deal with more serious adult themes such as friendship ('You're more than a friend, you're just like family'), alcoholism ('This is hard to say, but I think you're a much neater person when you're not drinking'), or losing your job ('Don't think of it as losing your job. Think of it as a time out between stupid bosses'). Now Hallmark Cards have founded a 'loyalty marketing group' that 'helps companies communicate with their customers at an emotional level'. It promotes the use of greetings cards for corporate use, to show that customers and employees are valued.

The greetings card industry remains healthy, especially in some countries. Britons, for example, spend more than any other country on cards – sending 31 cards per person each year, according to the Greeting Card Association (GCA), an industry body. Now, prompted by new online 'greetings cards by email' or 'via online



Source: Shutterstock.com/Hammamariah

social networks' operations who offer a speedy service without buying a stamp, traditional card companies have also moved into the online business. And although card companies insist that e-cards do not pose a grave threat, because people use them as a supplement to physical cards or for less important dates, Hallmark and others are dealing with the threat by allowing customers to send free e-cards from their websites, but encouraging them to purchase annual subscriptions for access to 'special' or 'exclusive' card designs. Whatever else these products may be, they are not seasonal!

Airlines, for example, fit all these criteria. They adopt a collection of methods to try to maximize the yield (i.e. profit) from their capacity. These include the following:

- **Over-booking capacity.** Not every passenger who has booked a place on a flight will actually show up for the flight. If the airline did not fill this seat it would lose the revenue from it. Because of this, airlines regularly book more passengers onto flights than the capacity of the aircraft can cope with. If they over-book by the exact number of passengers who fail to show up, they have maximized their revenue under the circumstances. Of course, if more passengers show up than they expect, the airline will have a number of upset passengers to deal with (although they may be able to offer financial inducements for the passengers to take another flight). If they fail to over-book sufficiently, they will have empty seats. By studying past data on flight demand, airlines try to balance the risks of over-booking and under-booking.

- *Price discounting.* At quiet times, when demand is unlikely to fill capacity, airlines will also sell heavily discounted tickets to agents who then themselves take the risk of finding customers for them. In effect, this is using the price mechanism to affect demand.
- *Varying service types.* Discounting and other methods of affecting demand are also adjusted depending on the demand for particular types of service. For example, the relative demand for first-, business-, and economy-class seats varies throughout the year. There is no point discounting tickets in a class for which demand will be high. Yield management also tries to adjust the availability of the different classes of seat to reflect their demand. It will also vary the number of seats available in each class by upgrading or even changing the configuration of airline seats.

HOW CAN OPERATIONS PLAN THEIR CAPACITY LEVEL?

Before an operation can decide which of the capacity plans to adopt, it must be aware of the consequences of adopting each plan in its own set of circumstances. Two methods are particularly useful in helping to assess the consequences of adopting particular capacity plans:

- cumulative representations of demand and capacity;
- queuing theory.

Cumulative representations

Figure 11.10 shows the forecast aggregated demand for a chocolate factory which makes confectionery products. Demand for its products in the shops is greatest at Christmas. To meet this demand and allow time for the products to work their way through the distribution system, the factory must supply a demand which peaks in September, as shown. One method of assessing whether a particular level of capacity can satisfy the demand would be to calculate the degree of over-capacity below the graph which represents the capacity levels (areas A and C) and the degree of under-capacity above the graph (area B). If the total over-capacity is greater than the total under-capacity for a particular level of capacity, then that capacity could be regarded as adequate to satisfy demand fully, the assumption being that inventory has been accumulated in the periods of over-capacity. However, there are two problems with this approach. The first is that each month shown in Figure 11.10 may not

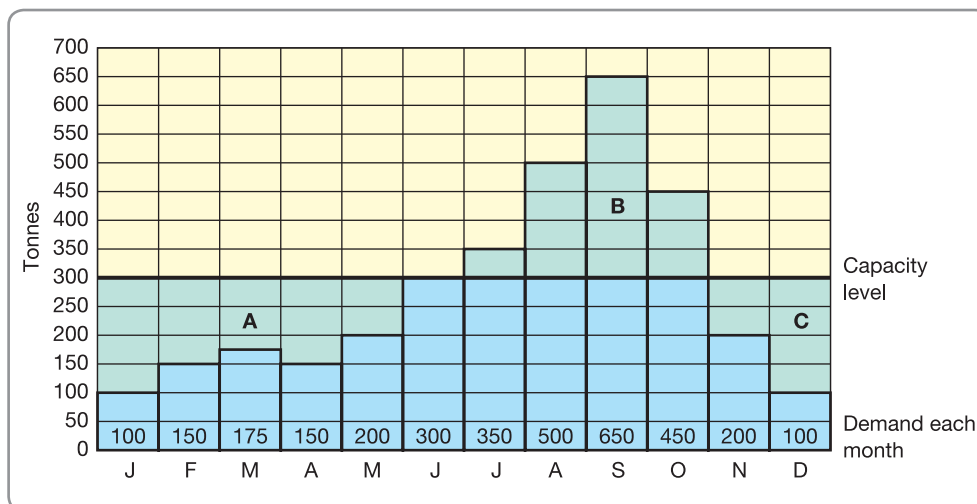
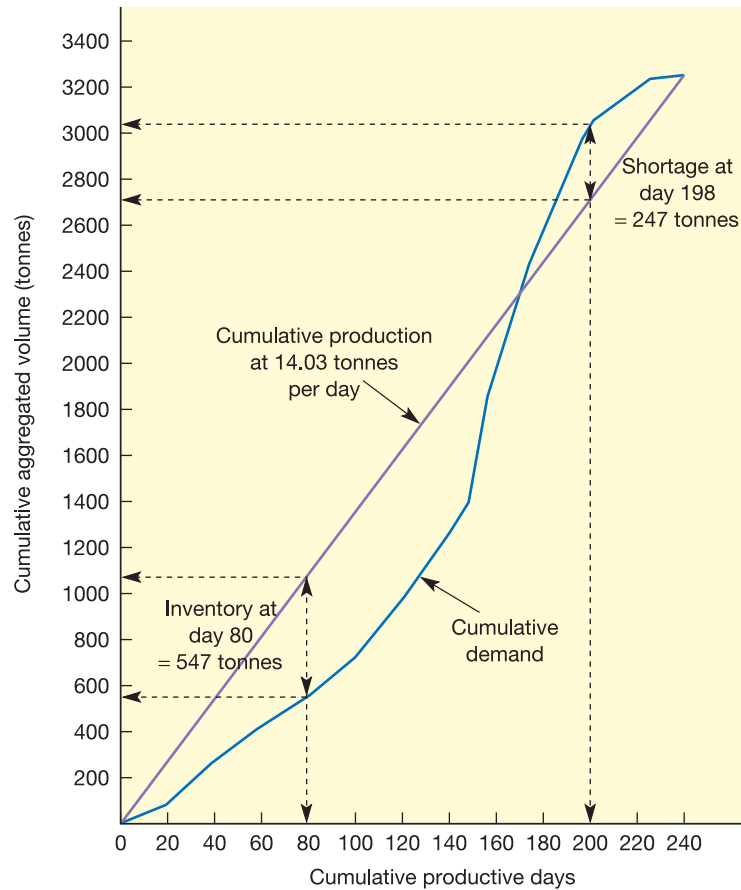


Figure 11.10 If the over-capacity areas (A+C) are greater than the under-capacity area (B), the capacity level seems adequate to meet demand. This may not necessarily be the case, however



	J	F	M	A	M	J	J	A	S	O	N	D
Demand (tonnes/month)	100	150	175	150	200	300	350	500	650	450	200	100
Productive days	20	18	21	21	22	22	21	10	21	22	21	18
Demand (tonnes/day)	5	8.33	8.33	7.14	9.52	13.64	16.67	50	30.95	20.46	9.52	5.56
Cumulative days	20	38	59	80	102	124	145	155	176	198	219	237
Cumulative demand	100	250	425	575	775	1075	1425	1925	2575	3025	3225	3325
Cumulative production (tonnes)	281	533	828	1122	1431	1740	2023	2175	2469	2778	3073	3325
Ending inventory (tonnes)	181	283	403	547	656	715	609	250	(106)	(247)	(150)	0

Figure 11.11 A level capacity plan which produces shortages in spite of meeting demand at the end of the year

have the same amount of productive time. Some months (August, for example) may contain vacation periods which reduce the availability of capacity. The second problem is that a capacity level which seems adequate may only be able to supply products *after* the demand for them has occurred. For example, if the period of under-capacity occurred at the beginning of the year, no inventory could have accumulated to meet demand. A far superior way of assessing capacity plans is first to plot demand on a *cumulative* basis. This is shown as the blue line in Figure 11.11.

The cumulative representation of demand immediately reveals more information. First, it shows that although total demand peaks in September, because of the restricted number of available productive days, the peak demand per productive day occurs a month earlier in August. Second, it shows that the fluctuation in demand over the year is even greater than

it seemed. The ratio of monthly peak demand to monthly lowest demand is 6.5:1, but the ratio of peak to lowest demand per productive day is 10:1. Demand per productive day is more relevant to operations managers, because productive days represent the time element of capacity.

The most useful consequence of plotting demand on a cumulative basis is that, by plotting capacity on the same graph, the feasibility and consequences of a capacity plan can be assessed. Figure 11.11 also shows a level capacity plan which produces at a rate of 14.03 tonnes per productive day. This meets cumulative demand by the end of the year. It would also pass our earlier test of total over-capacity being the same as or greater than under-capacity.

However, if one of the aims of the plan is to supply demand when it occurs, the plan is inadequate. Up to around day 168, the line representing cumulative production is above that representing cumulative demand. This means that at any time during this period, more product has been produced by the factory than has been demanded from it. In fact the vertical distance between the two lines is the level of inventory at that point in time. So by day 80, 1,122 tonnes have been produced but only 575 tonnes have been demanded. The surplus of production above demand, or inventory, is therefore 547 tonnes. When the cumulative demand line lies above the cumulative production line, the reverse is true. The vertical distance between the two lines now indicates the shortage, or lack of supply. So by day 198, 3,025 tonnes have been demanded but only 2,778 tonnes produced. The shortage is therefore 247 tonnes.

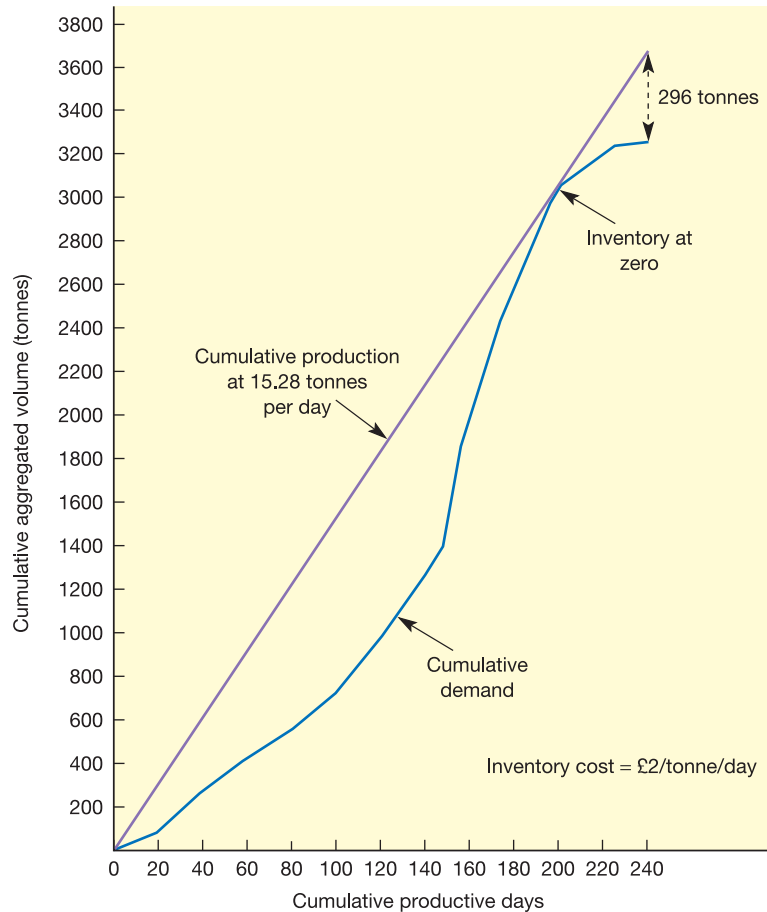
For any capacity plan to meet demand as it occurs, its cumulative production line must always lie above the cumulative demand line. This makes it a straightforward task to judge the adequacy of a plan, simply by looking at its cumulative representation. An impression of the inventory implications can also be gained from a cumulative representation by judging the area between the cumulative production and demand curves. This represents the amount of inventory carried over the period. Figure 11.12 illustrates an adequate level capacity plan for the chocolate manufacturer, together with the costs of carrying inventory. It is assumed that inventory costs £2 per tonne per day to keep in storage. The average inventory each month is taken to be the average of the beginning- and end-of-month inventory levels, and the inventory-carrying cost each month is the product of the average inventory, the inventory cost per day per tonne and the number of days in the month.

* Operations principle

For any capacity plan to meet demand as it occurs, its cumulative production line must always lie above its cumulative demand line.

Comparing plans on a cumulative basis

Chase demand plans can also be illustrated on a cumulative representation. Rather than the cumulative production line having a constant gradient, it would have a varying gradient representing the production rate at any point in time. If a pure demand chase plan was adopted, the cumulative production line would match the cumulative demand line. The gap between the two lines would be zero and hence inventory would be zero. Although this would eliminate inventory-carrying costs, as we discussed earlier, there would be costs associated with changing capacity levels. Usually, the marginal cost of making a capacity change increases with the size of the change. For example, if the chocolate manufacturer wishes to increase capacity by 5 per cent, this can be achieved by requesting its staff to work overtime – a simple, fast and relatively inexpensive option. If the change is 15 per cent, overtime cannot provide sufficient extra capacity and temporary staff will need to be employed – a more expensive solution which also would take more time. Increases in capacity of above 15 per cent might only be achieved by subcontracting some work out. This would be even more expensive. The cost of the change will also be affected by the point from which the change is being made, as well as the direction of the change. Usually, it is less expensive to change capacity towards what is regarded as the ‘normal’ capacity level than away from it.



	J	F	M	A	M	J	J	A	S	O	N	D
Demand (tonnes/month)	100	150	175	150	200	300	350	500	650	450	200	100
Productive days	20	18	21	21	22	22	21	10	21	22	21	18
Demand (tonnes/day)	5	8.33	8.33	7.14	9.52	13.64	16.67	50	30.95	20.46	9.52	5.56
Cumulative days	20	38	59	80	102	124	145	155	176	198	219	237
Cumulative demand	100	250	425	575	775	1075	1425	1925	2575	3025	3225	3325
Cumulative production (tonnes)	306	581	902	1222	1559	1895	2216	2368	2689	3025	3346	3621
Ending inventory (tonnes)	206	331	477	647	784	820	791	443	114	0	121	296
Average inventory (tonnes)	103	270	404	562	716	802	806	617	279	57	61	209
Inventory cost for month (£)	4120	9720	16,968	23,604	31,504	35,288	33,852	12,340	11,718	2508	2562	7524

Total inventory cost for year = £191,608

Figure 11.12 A level capacity plan which meets demand at all times during the year

Worked example

Suppose the chocolate manufacturer, which has been operating the level capacity plan as shown in Figure 11.13, is unhappy with the inventory costs of this approach and decides to explore two alternative plans, both involving some degree of demand chasing.

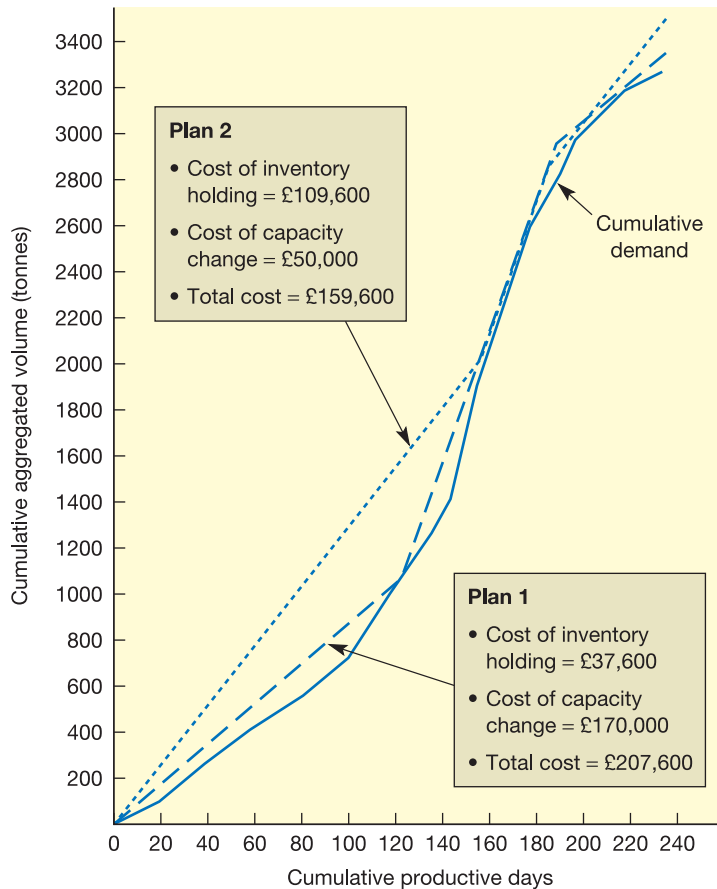


Figure 11.13 Comparing two alternative capacity plans

Plan 1

- Organize and staff the factory for a 'normal' capacity level of 8.7 tonnes per day.
- Produce at 8.7 tonnes per day for the first 124 days of the year, then increase capacity to 29 tonnes per day by heavy use of overtime, hiring temporary staff and some subcontracting.
- Produce at 29 tonnes per day until day 194, then reduce capacity back to 8.7 tonnes per day for the rest of the year.

The costs of changing capacity by such a large amount (the ratio of peak to normal capacity is 3.33:1) are calculated by the company as being:

Cost of changing from 8.7 tonnes/day to 29 tonnes/day = £110,000

Cost of changing from 29 tonnes/day to 8.7 tonnes/day = £60,000

Plan 2

- Organize and staff the factory for a 'normal' capacity level of 12.4 tonnes per day.
- Produce at 12.4 tonnes per day for the first 150 days of the year, then increase capacity to 29 tonnes per day by overtime and hiring some temporary staff.
- Produce at 29 tonnes/day until day 190, then reduce capacity back to 12.4 tonnes per day for the rest of the year.

The costs of changing capacity in this plan are smaller because the degree of change is smaller (a peak to normal capacity ratio of 2.34:1), and they are calculated by the company as being:

Cost of changing from 12.4 tonnes/day to 29 tonnes/day = £35,000

Cost of changing from 29 tonnes/day to 12.4 tonnes/day = £15,000

Figure 11.13 illustrates both plans on a cumulative basis. Plan 1, which envisaged two drastic changes in capacity, has high capacity change costs but, because its production levels are close to demand levels, it has low inventory carrying costs. Plan 2 sacrifices some of the inventory cost advantage of Plan 1 but saves more in terms of capacity change costs.

HOW IS CAPACITY PLANNING A QUEUING PROBLEM?

Cumulative representations of capacity plans are useful where the operation has the ability to store its finished goods as inventory. However, for operations where it is not possible to produce products and services *before* demand for them has occurred, a cumulative representation would tell us relatively little. The cumulative 'production' could never be above the cumulative demand line. At best, it could show when an operation failed to meet its demand. So the vertical gap between the cumulative demand and production lines would indicate the amount of demand unsatisfied. Some of this demand would look elsewhere to be satisfied, but some would wait. This is why, for operations which by their nature cannot store their output, such as most service operations, capacity planning and control is best considered using waiting or queuing theory.

Queuing or 'waiting line' management

When we were illustrating the use of cumulative representations for capacity planning and control, our assumption was that, generally, any production plan should aim to meet demand at any point in time (the cumulative production line must be above the cumulative demand line). Looking at the issue as a queuing problem (in many parts of the world queuing concepts are referred to as 'waiting line' concepts) accepts that, while sometimes demand may be satisfied instantly, at other times customers may have to wait. This is particularly true when the arrival of individual demands on an operation are difficult to predict, or the time to produce a product or service is uncertain, or both. These circumstances make providing adequate capacity at all points in time particularly difficult. Figure 11.14 shows the general form of this capacity issue. Customers arrive according to some probability distribution and wait to be processed (unless part of the operation is idle); when they have reached the front of the queue, they are processed by one of the n parallel 'servers' (their processing time also being described by a probability distribution), after which they leave the operation. There are many examples of this kind of system. Table 11.2 illustrates some of these. All of these examples can be described by a common set of elements that define their queuing behaviour.

The source of customers – sometimes called the calling population – is the source of supply of customers. In queue management 'customers' are not always human. 'Customers' could, for example, be trucks arriving at a weighbridge, orders arriving to be processed or machines waiting to be serviced, etc. The source of customers for a queuing system can be either *finite* or *infinite*. A finite source has a known number of possible customers. For example, if one maintenance person serves four assembly lines, the number of customers for the maintenance person is known, i.e. four. There will be a certain probability that one of the assembly lines will break down and need repairing. However, if one line really does break down the probability of another line needing repair is reduced because there are now only three lines to break down. So, with a finite source of customers, the probability of a customer arriving depends on the number of customers already being serviced. By contrast, infinite customer sources