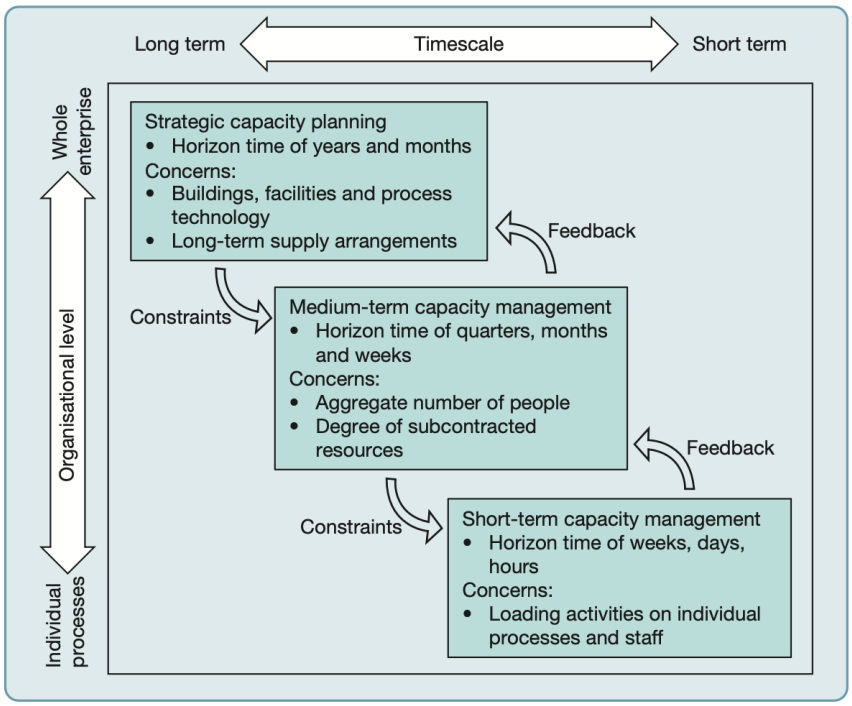
Capacity management

11.1 What is capacity management?

Capacity management = concerned with understanding the nature of demand and supply (capacity) and attempting to reduce mismatches between them in a way that reconciles the competing demands of customer satisfaction and resource efficiency.

The decisions are made within the constraints of the operation, the ability of its suppliers to supply, the availability of staff and so on. Each **level of capacity decision** is made with the constraints of a higher level

Short-term decisions = provide important feedback for planning over longer-term time horizons



Medium-term capacity management = involves assessing demand forecasts with a time horizon of between 2 and 18 months, during which time planned output can be varied

Short-term capacity management = few forecasts are accurate and most operations need to respond to changes in demand that occur over a shorter timescale

Capacity management performance objectives

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

- **Costs** will be affected by the balance between demand and capacity. Capacity levels in excess of demand could mean underutilization of capacity and therefore high unit-cost
- **Revenues** are affected by the balance between demand and capacity. Capacity levels ≥ than demand → ensures that all demand is satisfied and no revenue lost
- **Working capital** will be affected if an operation decides to build up finished product inventory prior to demand. Might allow demand to be satisfied but the organization will have to fund the inventory until it can be sold
- **Quality** of services might be affected by a capacity plan that involves large fluctuations in capacity levels, i.e. hiring temporary staff which disrupts the routine working of the operation and increases probability of errors
- **Speed** of response to customer demand could be enhanced either by the deliberate provision of surplus capacity to avoid queuing or through the build-up of product inventories
- **Dependability** of supply is 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
- **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.

A framework for capacity management

The first step on the demand side of the equation is to measure/forecast demand for services and products over different time periods. It involves selecting from a range of **qualitative** (panel, Delphi and scenario planning) and **quantitative** (time series and causal models) tools to support more accurate prediction of demand

Second step is on the supply side of the framework and involves measuring the capacity to deliver services and products. The impacts of mix, time frame and output specification should be considered

Third step is considering if and how to manage demand using demand management and yield management techniques

Fourth step is to manage the supply side by determining the appropriate level of average capacity and then deciding whether to either keep it constant (level capacity plan) or to adjust capacity in line with changing demand patterns (chase capacity plan)

Last step is to understand the consequences of different capacity management decisions on both the demand side and supply side of the framework

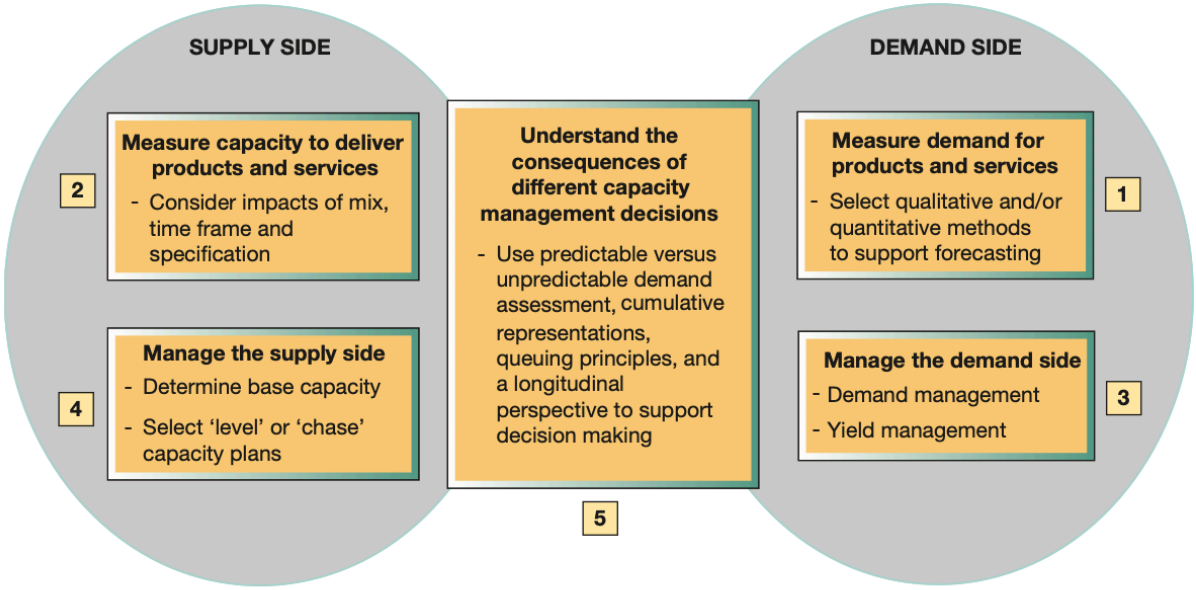


Figure 11.3 Capacity management framework

11.2 How is demand measured?

First task of capacity management is to understand the patterns of demand for products and services over different time frames. Knowing the rate of change is vital for business planning

Qualitative approaches to forecasting

Panel = acts like a focus group allowing everyone to talk openly. Advantage that several brains are included, disadvantage that it is difficult to reach a consensus and that everyone could get it wrong

Delphi method = formal method that attempts to reduce the influences from procedures of face-to-face meetings. Employs a survey of experts where replies are analyzed and anonymous summaries are sent back to all experts. Experts are then asked to reconsider their original forecasts in the light of the replies and arguments put forward by other experts. Process is **repeated** several times to conclude either with a consensus or narrower range of decisions. Problems associated with the approach is constructing an appropriate questionnaire and selecting an appropriate panel of experts

Scenario planning = applied to long-range forecasting using a panel. The panel members are asked to devise a range of future scenarios, each scenario can then be discussed with inherent risks considered. It looks at a range of options and puts plans in place to try to avoid the ones that are least desired and taking action to follow the most desired

Quantitative approaches to forecasting

Time series analysis = examines the pattern of time series data and extrapolates future behavior by removing underlying variations with assignable causes

Simple moving-average forecasting = a time series analysis used to estimate demand for a future time period by averaging the demand for the n most recent time periods. The value of n can be set at any level but is usually in the range of 3 to 7. Next period's demand is forecasted by taking the moving average of the previous n periods' actual demand $\frac{F_t = A_{t-1} + A_{t-2} + A_{t-3} + A_{t-4}}{4}$ Here n = 4 and t = week and F_t = forecast demand for week t, A_t = actual demand for week t

11.3 How is capacity measured?

Nature of capacity = the ability to supply

The capacity of an operation = the maximum level of value-added activity over a period of time that the process can achieve under normal operating conditions

Measuring capacity can be hard to define unambiguously unless the operation is standardized and repetitive

Output capacity measure = appropriate to use when the output from an operation does not vary in its nature

Input capacity measure = appropriate when a wider range of outputs places varying demands on the process

Table 11.4 Input and output capacity measures for different operations

Operation	Input measure of capacity	Output measure of capacity
Hospital	**Beds available**	Number of patients treated per week
Air-conditioner plant	Machine hours available	**Number of units per week**
Theatre	**Number of seats**	Number of customers entertained per week
University	**Number of students**	Students graduated per year
Electricity company	Generator size	**Megawatts of electricity generated**
Retail store	**Sales floor area**	Number of items sold per day
Airline	**Number of seats available**	Number of passengers per week
Brewery	Volume of fermentation tanks	**Litres per week**

(*Note*: The most commonly used measure is shown in bold.)

The effect of activity mix on capacity management

An operation's ability to supply is partly dependent on what it is being required to do. Some of the problems caused by variation mix can be partially overcome by using aggregated capacity measures (different products/services are bundled together)

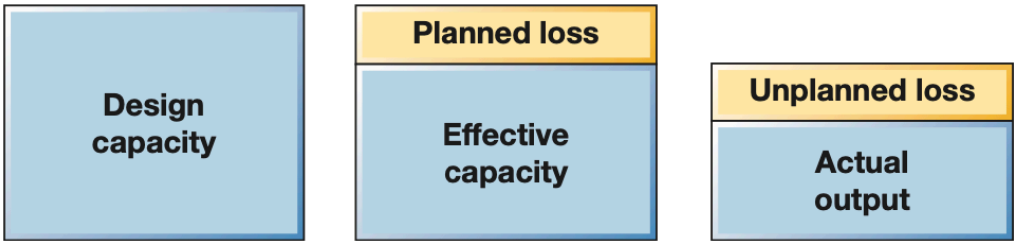
The effect of time frame on capacity measurement

When measuring capacity, operations managers should consider 3 different measures of capacity

Design capacity = the theoretical capacity of an operation that its technical designers had in mind when they commissioned it

Effective capacity = the capacity of an operation after planned losses are accounted for

Actual output = the capacity of an operation after both planned and unplanned losses are accounted for



$$\text{Utilisation} = \frac{\text{Actual output}}{\text{Design capacity}}$$

$$\text{Efficiency} = \frac{\text{Actual output}}{\text{Effective capacity}}$$

Capacity leakage

Overall equipment effectiveness OEE (capacity leakage) = the reduction in capacity caused by both predictable and unpredictable losses. It is calculated as

OEE = a x p x q, where a = availability of a process, p = performance/speed of a process, q = quality of product/services that the process creates

OEE works on the assumption that some capacity leakage occurs, causing reduced availability

For processes to operate effectively, they need to achieve high levels of performance against all 3 dimensions, availability, performance (speed) and quality

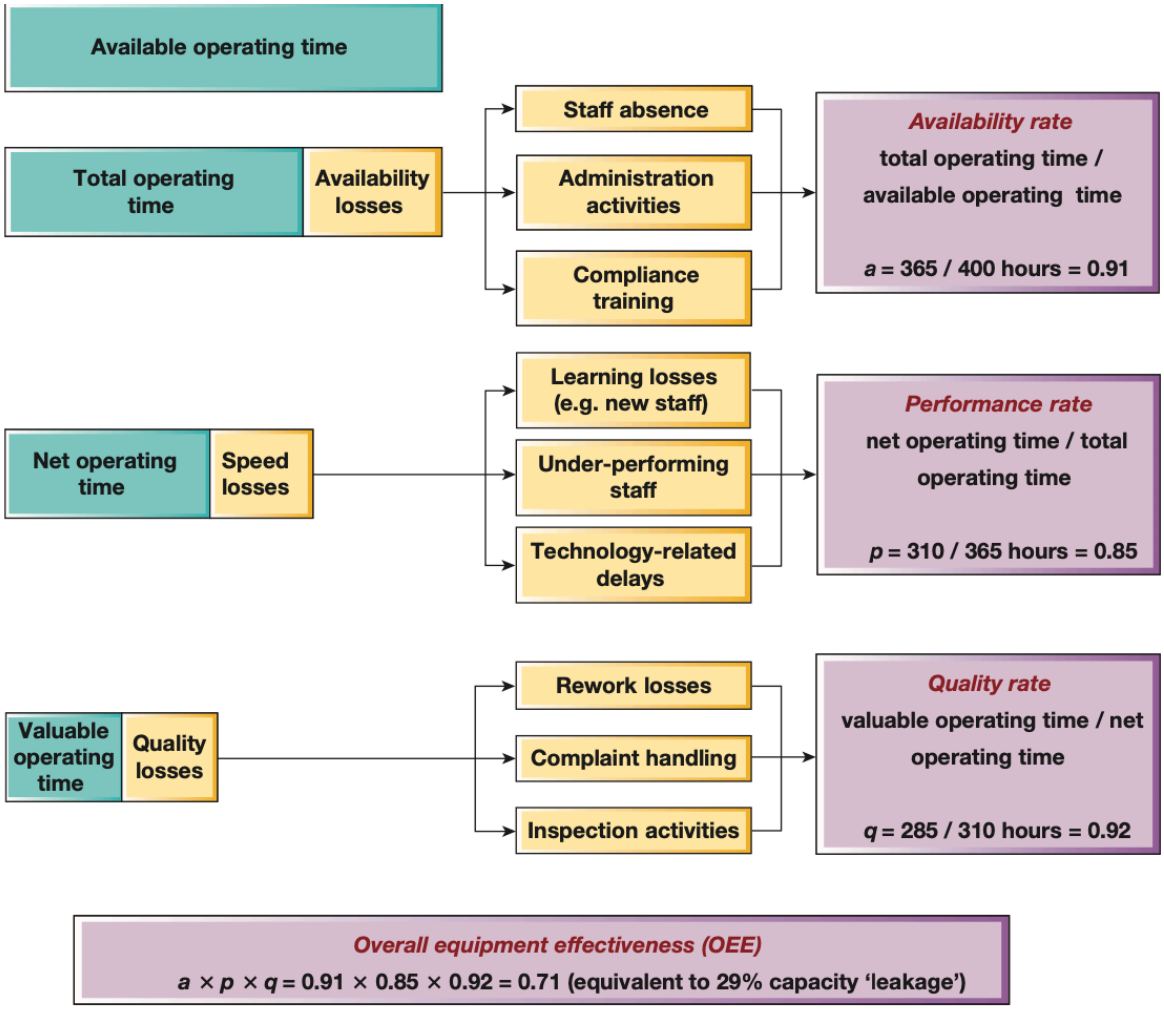


Figure 11.7 Overall equipment effectiveness (OEE) for a client support service team in a small software company

Understanding changes in capacity

Variation in capacity = the ability to supply

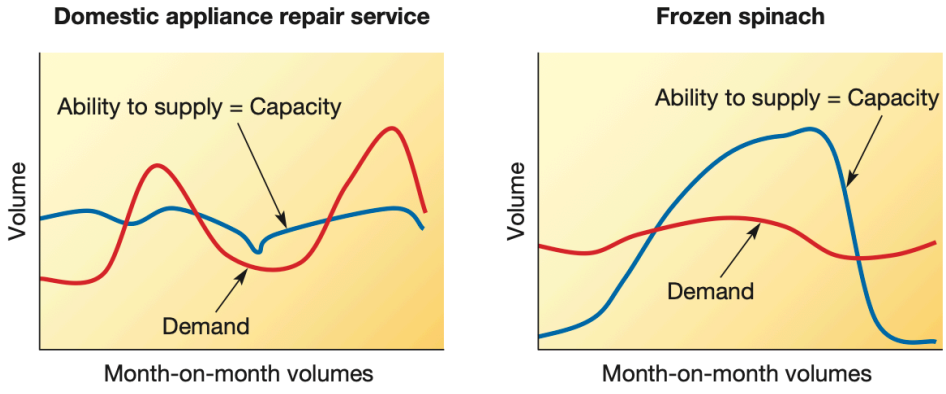


Figure 11.8 Volatility in demand versus volatility in capacity

11.4 How is the demand side managed?

Demand management = changing the pattern of demand to bring it closer to available capacity. Can be achieved by stimulating **off-peak demand** or by constraining **peak demand**

Methods used in managing demand are:

- **Price differentials** = adjusting price to reflect demand
- **Scheduling promotion** = varying the degree of market stimulation through promotion and advertising in order to encourage demand during normally low periods
- **Constraining customer access** = customers may only be allowed access to the operation's products or services at particular times
- **Service differentials** = allowing service levels to reflect demand (implicitly or explicitly) by letting service deteriorate in periods of high demand and increase in periods of low demand
- **Creating alternative products or services** = developing services or products aimed at filling capacity in quiet periods

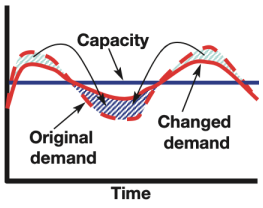


Figure 11.9 Demand management plan

11.6 How can operations understand the consequences of their capacity management decisions?

When making capacity management decisions, managers are attempting to balance the need to provide a responsive and customer-oriented service with the need to minimize costs. There are 3 main capacity management plans:

Demand management plan = Peak demand can be brought forward by offering discounts to selected clients.

Chase capacity plan = Capacity may be increased through the use of outsourced suppliers during the busiest months of the year

Level capacity plan = Capacity may be constrained and clients may still experience delays during high demand periods

4 methods used in assessing the consequences of the capacity management plans are:

- Factoring in predictable versus unpredictable demand variation
- Using cumulative representations of demand and capacity

- Using queuing principles to make capacity management decisions
- Taking a longitudinal perspective that considers short- and long-term outlooks

Factoring in predictable versus unpredictable demand variation

If demand is changeable but the change is predictable, then capacity adjustments may be needed but at least be planned in advance

Unpredictable variation in demand makes operations have to react quickly, otherwise the change in capacity will have little effect on the operation's ability to deliver products and services as needed by their customers

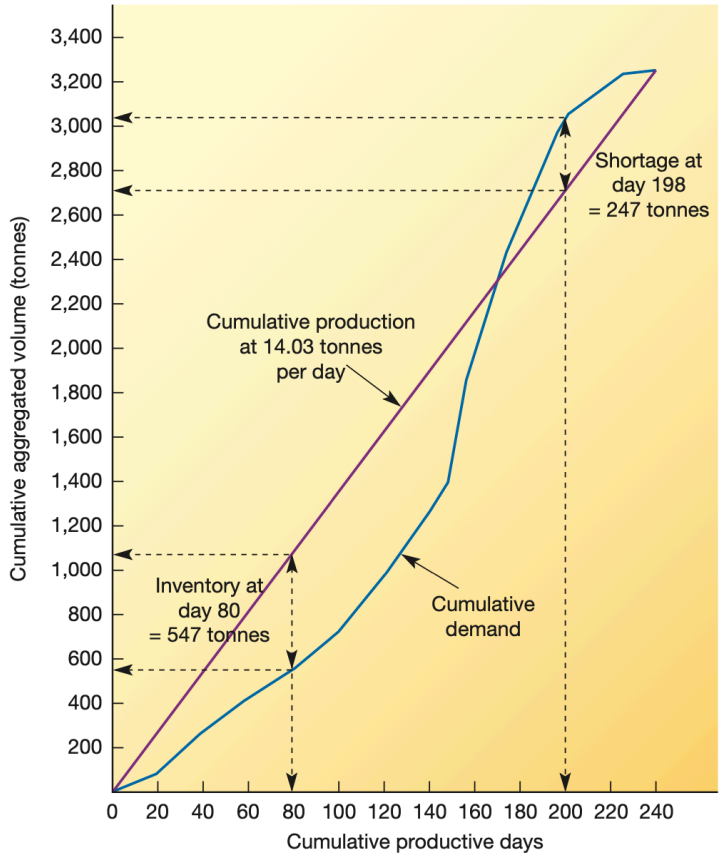
		Unpredictable variation	
		Low	**High**
Predictable variation	**High**	*Objective* – Adjust planned capacity as efficiently as possible *Capacity management tasks* • Evaluate optimum mix of methods for capacity fluctuation • Work on how to reduce cost of putting plan into effect	*Objective* – Adjust planned capacity as efficiently as possible and enhance capability for further fast adjustments *Capacity management tasks* • Combination of those for predictable and unpredictable variation
	Low	*Objective* – Make sure the base capacity is appropriate *Capacity management tasks* • Seek ways of providing steady capacity effectively	*Objective* – Adjust capacity as fast as possible *Capacity management tasks* • Identify sources of extra capacity and/or uses for surplus capacity • Work on how to adjust capacity and/or uses of capacity quickly

Figure 11.13 The nature of capacity management depends on the mixture of predictable and unpredictable demand and capacity variation

Using cumulative representations of demand and capacity

Demand per productive day is more relevant to operations managers because productive days represent the time element of capacity

By plotting demand and capacity on a cumulative basis, the feasibility and consequences of a plan can be assessed

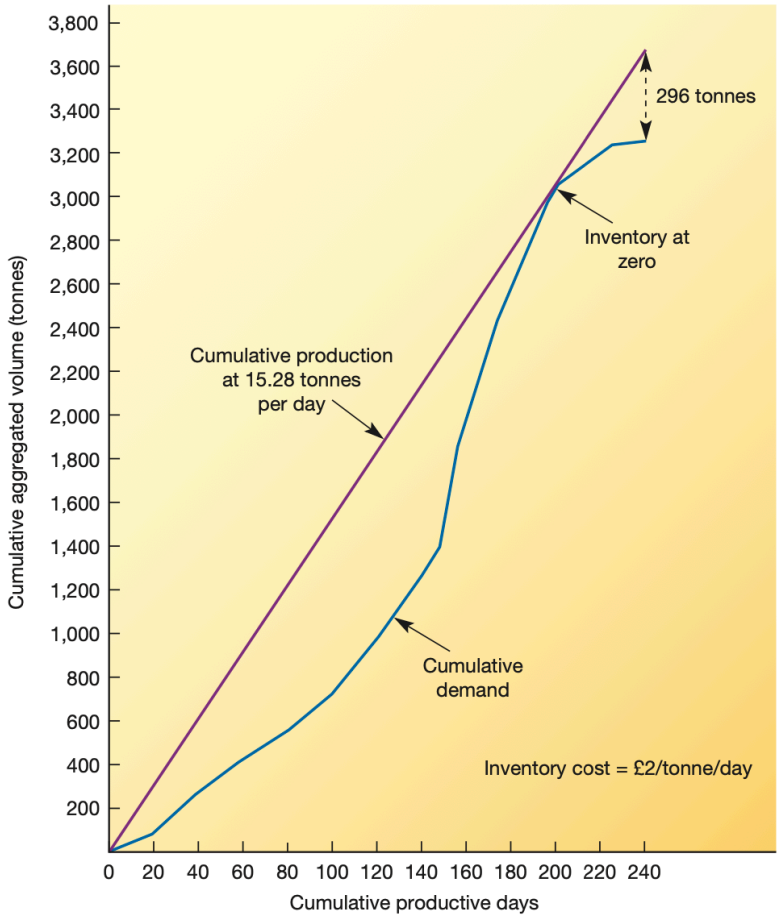


	J	F	M	A	M	J	J	A	S	O	N	D
Demand (tonnes/month)	100	150	175	150	200	300	350	500	650	450	200	100
Productive days	20	18	21	21	22	22	21	10	21	22	21	18
Demand (tonnes/day)	5	8.33	8.33	7.14	9.52	13.64	16.67	50	30.95	20.46	9.52	5.56
Cumulative days	20	38	59	80	102	124	145	155	176	198	219	237
Cumulative demand	100	250	425	575	775	1,075	1,425	1,925	2,575	3,025	3,225	3,325
Cumulative production (tonnes)	281	533	828	1,122	1,431	1,740	2,023	2,175	2,469	2,778	3,073	3,325
Ending inventory (tonnes)	181	283	403	547	656	715	609	250	(106)	(247)	(150)	0

Figure 11.14 A level capacity plan that produces shortages in spite of meeting demand at the end of the year

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

By judging the area between the cumulative production and demand curves, an impression of the inventory implications can be gained



	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	1,075	1,425	1,925	2,575	3,025	3,225	3,325
Cumulative production (tonnes)	306	581	902	1,222	1,559	1,895	2,216	2,368	2,689	3,025	3,346	3,621
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 (£)	4,120	9,720	16,968	23,604	31,504	35,288	33,852	12,340	11,718	2,508	2,562	7,524

Total inventory cost for year = £191,608

Figure 11.15 A level capacity plan that meets demand at all times during the year

The marginal cost of making a capacity change increases with the size of the change

Using queuing principles to make capacity management decisions

Cumulative representations of capacity plans are useful where the operation has the ability to store its finished goods as inventory. For operations which by their nature cannot store their output such as service operations, capacity management decisions are best considered using **waiting line or queuing theory**

Queuing theory = operations managers accept that while some demand may be satisfied instantly, at other times customers may have to wait. True when the arrival of individual demands on an operation are difficult to predict, the time to create a service or product is uncertain, or both

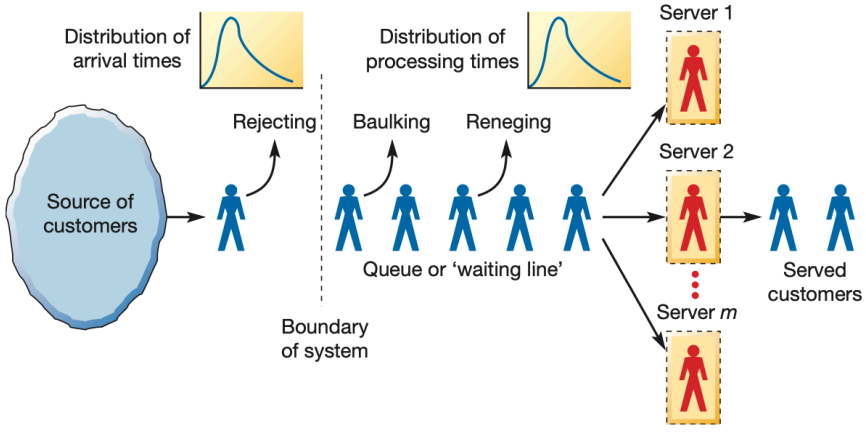
The general form of capacity issue is that customers arrive according to some probability distribution and wait to be processed (unless part of the operation is idle); when they have reached the front of the queue they are processed by one of the n parallel servers (their processing time also being described by a probability distribution) after which they leave the operation

The source of customers

- In queue management, customers are not always human, they could be trucks, orders etc.
- Source of customers for queuing systems can either be finite (known number of possible customers) or infinite
    - With a finite source of customers the probability of a customer arriving depends on the number of customers already being serviced
    - With an infinite customer source we assume that there is a large number of potential customers so that it is always possible for another customer to arrive no matter how many are being serviced

Servers

- **Server** = the facility that processes the customers in the queue
- Servers can be configured in parallel or in a series arrangement
- It is likely to be variation in how long it takes to process each customer and even if they have the same needs, human servers will vary in the time they take to perform repetitive serving tasks.
    - Processing time is described by a probability distribution



- **Figure 11.16** Capacity management as a queuing problem

The arrival rate

- The rate at which customers needing to be served arrive at the server or servers. Customers rarely arrive at a steady and predictable rate and there is usually variability
- Arrival rates are therefore described in terms of probability distributions

The queue

- Customers waiting to be served form the queue or waiting line itself.
- Queues are not always physical but can be a waiting list for example

Queue discipline

- The set of rules that determine the order in which customers waiting in the queue are served for example first in first out

Rejecting

- If the number of customers in a queue is already at the maximum number allowed, then the customer could be rejected by the system

Balking

- **Balking** = When a customer is a human being with free will and with the ability to get annoyed, they may refuse to join the queue and wait for service if it is judged to be too long

Reneging

- Customer has queued for a certain length of time and then leaves the queue and therefore the chance of being served

Variability effects on queues

Dilemma in managing the capacity of a queuing system is how many servers to have available at any point in time in order to avoid unacceptably long queuing times or unacceptably low utilization of the servers

When the **average capacity (processing capability)** of the operation matches the **average demand (arrival rate)** on the system, both queues and idle time will occur

Too few servers (low capacity) will build up queues to a level where customers become dissatisfied with the waiting time although the server utilization is high. If too many servers are used (high capacity) the time that customers expect to wait is low but so is the utilization of the servers.

The capacity planning and control problem is a trade-off between customer waiting time and system utilization

Taking a longitudinal perspective that considers short- and long-term outlooks

The capacity control process can be seen as a sequence of partially reactive capacity decisions. At the beginning of each period the operations management considers its forecasts of demand, understanding of current capacity and how much inventory has been carried forward from the previous period

Based on that information plans for the following period’s capacity are made

Success of capacity management is measured by a combination of **costs, revenue, working capital and customer satisfaction** which is influenced by the actual product or service and the capacity available to the operation in any period

If long-term outlook for demand is good (higher than current capacity can cope with) then it is unlikely that poor (demand less than capacity) short-term demand would cause an operation to make large cuts in capacity.

If long-term outlook for demand is poor (lower than current capacity) then it is unlikely that good (demand more than capacity) short-term demand would cause an operation to take on large or difficult to reverse extra capacity.

Long-term outlook for volume	Short-term outlook for volume: Decreasing below current capacity	Short-term outlook for volume: Level with current capacity	Short-term outlook for volume: Increasing above current capacity
Decreasing below current capacity	Reduce capacity (semi) permanently. For example, reduce staffing levels; reduce supply agreements.	Plan to reduce capacity (semi) permanently. For example, freeze recruitment; modify supply agreements.	Increase capacity temporarily. For example, increase working hours, and/or hire temporary staff; modify supply agreements.
Level with current capacity	Reduce capacity temporarily. For example, reduce staff working hours; modify supply agreements.	Maintain capacity at current level.	Increase capacity temporarily. For example, increase working hours, and/or hire temporary staff; modify supply agreements.
Increasing above current capacity	Reduce capacity temporarily. For example, reduce staff working hours, but plan to recruit; modify supply agreements.	Plan to increase capacity above current level; plan to increase supply agreements.	Increase capacity (semi) permanently. For example, hire staff; increase supply agreements.