Part IV - Ch 2 Performance quantification

van Koningsveld, M.; de Vriend, H.J.

Publication date
2021

Document Version
Final published version

Published in
Ports and Waterways

Citation (APA)
van Koningsveld, M., & de Vriend, H. J. (2021). Part IV - Ch 2 Performance quantification. In M. V. Koningsveld, H. J. Verheij, P. Taneja, & H. J. de Vriend (Eds.), *Ports and Waterways: Navigating a changing world* (pp. 365-396). TU Delft OPEN Publishing.

2 Performance quantification

2.1 Data acquisition and validation

[1]Data is needed to establish empirical relationships between parameters, but also to calibrate and validate models. The quantity and quality of those data has to be sufficient to achieve the accuracy required.

In order to gather the necessary data, one may go out and make one's own observations during some time. Rijkswaterstaat engineer Kooman, for instance, observed in the 1970's vessel passages and waiting times at various locks in the Netherlands. His observations are the basis of a number of empirical rules used in the so-called Kooman method to estimate waiting and passage times at locks (Kooman and De Bruijn, 1975).

Another approach is monitoring. This can be done at a specific location with equipment taking continuous measurements and transmitting or logging the data. But it can also be done by asking actors to collect and transfer data during their operations. The International Maritime Organization (IMO) requirement for ocean-going vessels to share their data on fuel oil consumption is one example, the request to Inland Water Transport (IWT) vessels to share their soundings for Covadem is another (also see Part I – Section 2.1.1).

This multi-source data acquisition may conflict with prevailing legislation about privacy or data ownership. Privacy problems can be avoided by anonymisation, but in many countries ownership is a more problematic issue, especially if data is traded against money, or if they are considered of strategic importance. Yet, it is important that data can be shared, if only because they are of more value together than in isolation.

When applied to similar situations, a well-validated model may produce results that correlate so well that they can be captured in a rule of thumb. In that case the model can be considered the data source.

Before being made available for further use, the data quality needs to be checked. Not all observations are equally accurate; sometimes values may lie far out side the normal distribution (outliers) or be obviously wrong. Such values need to be explained or removed from the set, otherwise they may lead to a false correlation. Figure 2.1 shows an extreme example, in this case for the value of imported agricultural goods in the EU.

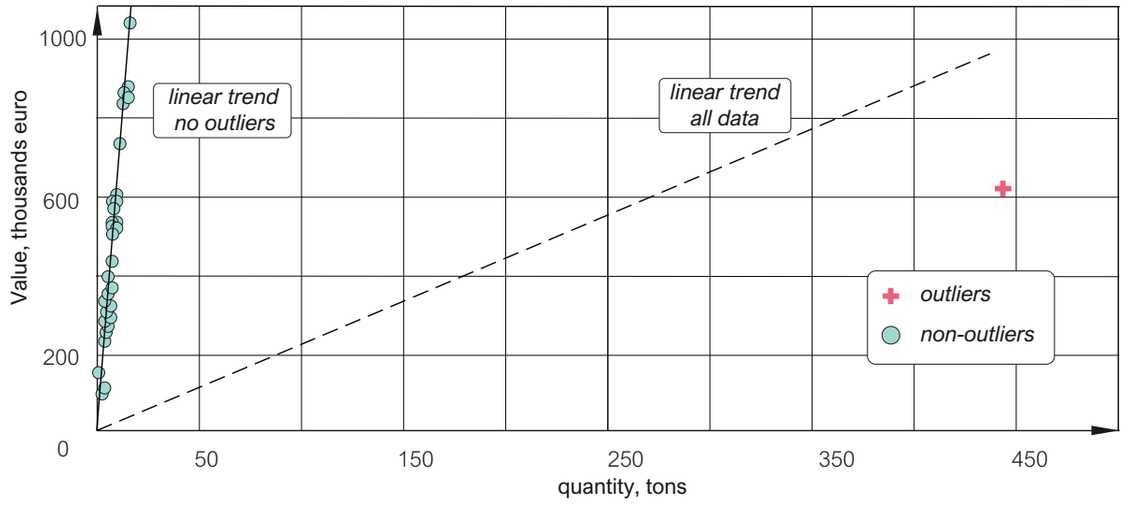

Figure 2.1: Outlier effect on the relationship between two parameters (reworked from Kopustinskas and Arsenis, 2012, by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

[1]This chapter made use of '*Service systems in ports and inland waterways*' (Groenveld, 2001), lecture notes for the Ports and Waterways courses CIE4330 and CIE5306 at TU Delft.

2.2 Empirical rules

Empirical rules, or rules of thumb, are based on experience, observations or model results concerning the relationship between two or more parameters, like in Figure 2.1. In principle, their validity is restricted to the locations and time intervals in which the observations have been made, or to which the model has been applied. Extrapolation to other situations can become very inaccurate and therefore requires explicit validation.

The observations used are mostly measured or recorded data on the parameters to be correlated. As such data are usually scattered, statistical methods such as regression analysis are used to establish the relationship. Artificial Intelligence techniques can be used, also to derive more complex relationships from data. In any case, an accuracy measure, such as the correlation coefficient R or the coefficient of determination R^2, is a useful addition (see, for instance, Figure 2.2).

Rules of thumb are widely used in waterborne transport systems. Many dimension requirements regarding waterways or water bodies in ports (see Part II and Part III) are experience-based and therefore actually rules of thumb. They are also often used in (pre-)feasibility studies and in the early design phases of infrastructure. In the following sections we will give some performance-related examples.

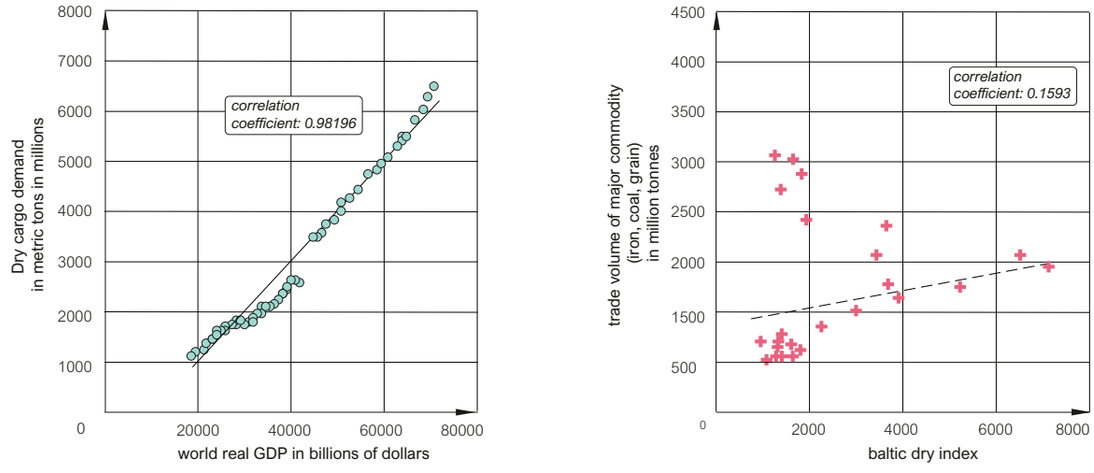

Figure 2.2: Relationships with different correlation coefficients (reworked from Akyar, 2018, by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

2.2.1 Port performance

In Part II we have shown examples of terminal throughput and storage capacity calculations. If such calculations are too laborious, e.g. in a (pre-)feasibility study or an early design phase, one may use rules of thumb to gain a rough indication of what to expect.

One example is the capacity of a berth, which is of course proportional to its availability, its occupancy and the capacity of the equipment handling a ship. In a formula (also see Ligteringen, 2017):

$$C_b = PNn_d m_b \tag{2.1}$$

in which:

C_b	=	estimated berth capacity in ton/yr or TEU/yr,
P	=	(un)loading capacity per unit handling equipment in ton/hr or TEU/hr,
N	=	number of (un)loading units operating on a ship of average size,
n_d	=	berth availability = number of operational hours per year,
m_b	=	berth occupancy factor = fraction of the time that the berth is occupied.

The values to be substituted into Equation 2.1 are experience-based, which makes this a rule of thumb. Table 2.1 gives typical ranges of results, expressed in throughput capacity per unit quay length or per berth (in the case of a crude oil jetty).

Cargo type	Capacity range	Units
Conventional general cargo	500 – 1,000	ton/yr per m^1 quay length
Containers	800 – 1,700	TEU/yr per m^1 quay length
Coal import (Van Vianen et al., 2011)	10,000 – 30,000	ton/yr per m^1 quay length
Coal export (Van Vianen et al., 2011)	50,000 – 150,000	ton/yr per m^1 quay length
Iron ore import (Van Vianen et al., 2011)	25,000 – 75,000	ton/yr per m^1 quay length
Iron ore export (Van Vianen et al., 2011)	20,000 – 60,000	ton/yr per m^1 quay length
Crude oil (VLCC and ULCC)	70 million	ton/yr per berth

Table 2.1: Empirical capacity ranges for different types of cargo (Ligteringen, 2017).

One may also use rules of thumb to estimate the throughput capacity per unit storage area. Table 2.2 gives some ranges per cargo type. The storage area is the gross storage area and includes the net storage area and internal roads, pipelines and/or conveyor belts and equipment rails at the storage yard.

Cargo type	Capacity range	Units
Conventional general cargo	4 – 6	ton/yr per m^2 storage area
Containers (Drewry, 2010)	0.75 – 5.5	TEU/yr per m^2 storage area
Coal import (Van Vianen et al., 2011)	15 – 75	ton/yr per m^2 storage area
Coal export (Van Vianen et al., 2011)	30 – 200	ton/yr per m^2 storage area
Iron ore import (Van Vianen et al., 2011)	30 – 80	ton/yr per m^2 storage area
Iron ore export (Van Vianen et al., 2011)	60 – 200	ton/yr per m^2 storage area
Crude oil	40 – 50	ton/yr per m^2 storage area

Table 2.2: Empirical throughput capacity ranges per unit storage area (Ligteringen, 2017).

The total terminal area, in addition, includes the other buildings and infrastructure in between quay and yard on the terminal. For a first-order estimate of the total terminal area, the following rules of thumb can be used:

Cargo type	Total terminal area [m^2]
Containers	total quay length [m] * 400 – 500 m
Dry bulk (Kox, 2017)	gross storage area [m^2] * 1.4 – 1.5 m
Liquid bulk (Kox, 2017)	gross storage area [m^2] * 1.6 – 1.7 m

Table 2.3: Total terminal area estimation (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

In the example elaborated in Part II – Chapter 4, a first-order terminal design was made for a container terminal with straddle carriers and an annual throughput of 2,460,000 TEU. The estimated quay length of 1,579.7 m translates to 1,557 TEU/yr per m^1 quay length. The estimated storage area of 494,000 m^2 for just the stacks and 593,787 m^2 including roads, translate to 5.0 and 4.1 TEU/yr per m^2 storage area respectively. The apron area of

129,533 m^2 plus the storage area including roads, divided by the quay length translates to a terminal depth of 458 m. It may be observed that these values fall well within the ranges shown in Table 2.1 to Table 2.3.

A rule of thumb often used in port development refers to the (non-linear) relationship between berth occupancy and waiting time. As soon as the occupancy factor exceeds 0.5, waiting times increase rapidly. This means that the number of berths needs to be increased in order to keep waiting times within the agreed bounds. Building or extending quays, however, is a costly and lengthy procedure, which means that studies and design procedures have to start at lower values of the occupancy factor.

Apart from capacities and cost-benefit ratios, safety is an important quality identifier for ports. Safety may concern the risk due to a hazardous event in a port, such as an explosion at an Liquified Natural Gas (LNG) terminal (Figure 2.3). Here risk is defined as the probability of occurrence of the event, times the impact in terms of material damage (material risk) or casualties and injuries (personal risk). The risk level calculation is to a large extent empirical, if it were only because the impact of a hazard is experience-based. While the risk contours are determined using models, the empirical factors depend on the degree of overpressure, toxicity or temperature increase causing a hazard.

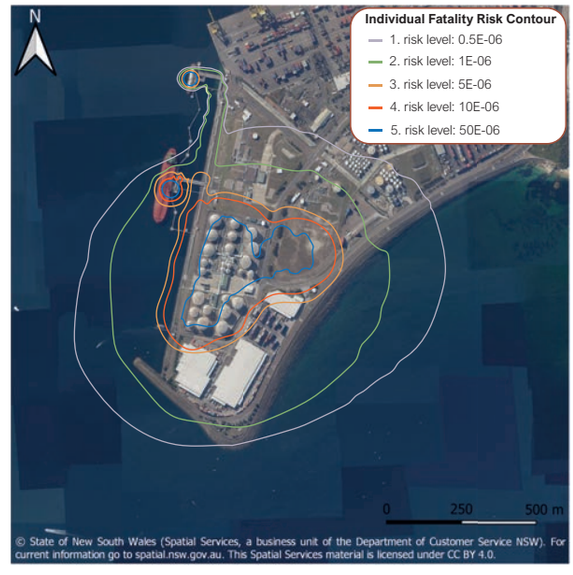

Figure 2.3: Individual fatality risk contours for an oil terminal, Port Botany, Australia (source: Spatial Service State of New South Wales, contours from Sherpa consulting, image by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0.

The LNG example concerns a case of impact reduction, as the risk contours are likely to be translated to exclusion zones if a hazard occurs. Prevention of hazardous situations is another type of measure to increase port safety. Moored vessel motions, for instance, can not only hamper (un)loading operations, but can also lead to hazardous situations, such as snapping mooring lines. Instead of computing mooring line forces from ship motions derived from wave conditions, one uses experience-based wave height limits to decide whether a vessel should leave berth (Table 2.4).

2.2.2 Waterway performance

Like in the case of port water bodies, many rules for waterway dimensions are empirically based (see, for instance RVW, 2020). Rules of thumb can also serve to estimate the capacity of open waterways, i.e. without delays due to locks or bridges.

We define the intensity, I, of the traffic on a waterway as the number of vessels (or the tonnage) that pass a fixed cross-section per unit time at a given point in time (Figure 2.4). The maximum possible value of this instantaneous quantity is the waterway's capacity, C. The ratio I/C is called the traffic load. The number of vessels per unit area at the same time is the density, D.

Vessel type	**Limiting wave height H_s [m]**	
	0° (ahead or astern)	**45–90° (abeam)**
General cargo	1.0	0.8
Container, Ro-Ro	0.5	
Dry bulk (30,000 – 100,000 DWT), loading	1.5	1.0
Dry bulk (30,000 – 100,000 DWT), unloading	1.0	0.8 – 1.0
Tankers (30,000 – 200,000 DWT), loading	1.5 – 2.5	1.0 – 1.2
tankers (> 200,000 DWT)	2.5 – 3.0	1.0 – 1.5

Table 2.4: Limiting wave heights for moored vessels (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

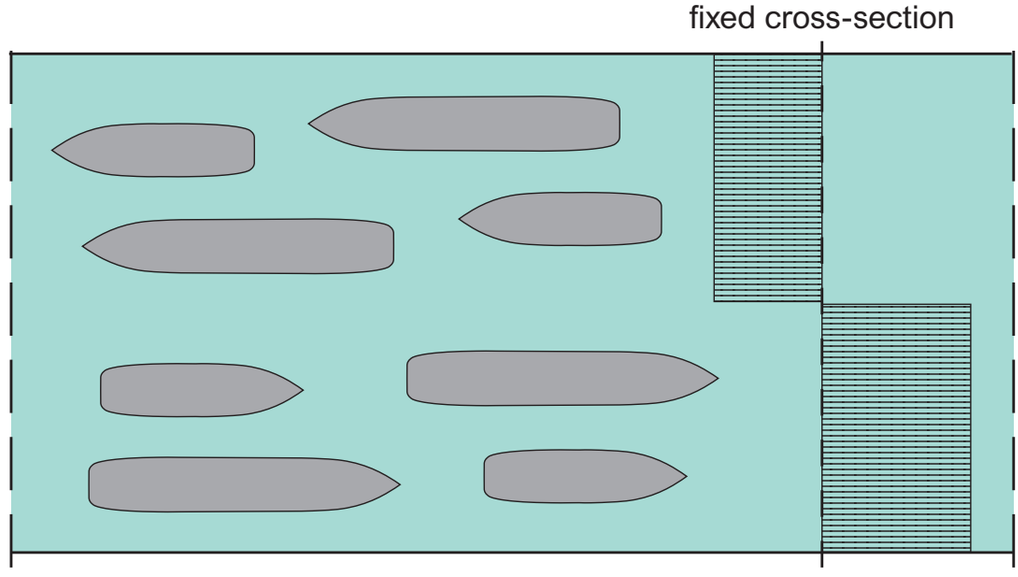

Figure 2.4: Traffic intensity on a waterway (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

On an open waterway, i.e. without any delays due to bridges or locks, this capacity is determined to a large extent by ship-waterway interactions (resistance, limit speed, economic speed) and ship-ship interactions (encounters, overtaking, blue-boarding, etc.).

Clearly, the vessel speed is a determining factor of a waterway's traffic intensity and capacity. This speed in influenced by the ship-waterway interaction, as well as by ship-ship interactions and economic considerations (fuel consumption). If we reduce the ship-ship interaction to a relationship between average vessel speed and traffic density, there is a complex interaction between the vessel speed V_s, the traffic density and the traffic intensity. This can be explained as follows (Figure 2.5):

- at very low traffic density, the vessels can sail without mutual interaction, at speed V_0;
- as the density increases, the velocity starts decreasing, but at such a rate that the traffic intensity still increases; the intensity in this situation is less than the waterway's capacity;
- this intensity increase continues until a maximum is reached; at that point, the velocity is V_{crit} and density D_{crit} ; the intensity is equal to the capacity here;
- as the density increases further, the ship speed decreases so strongly, that the intensity also starts decreasing, until it reaches zero at the density where sailing is no longer possible; in this situation the capacity decreases along with the intensity.

The relationship between vessel speed and traffic density can be determined either empirically or from model simulations. The intensity is the result of a calculation, viz. the product of vessel speed and traffic density.

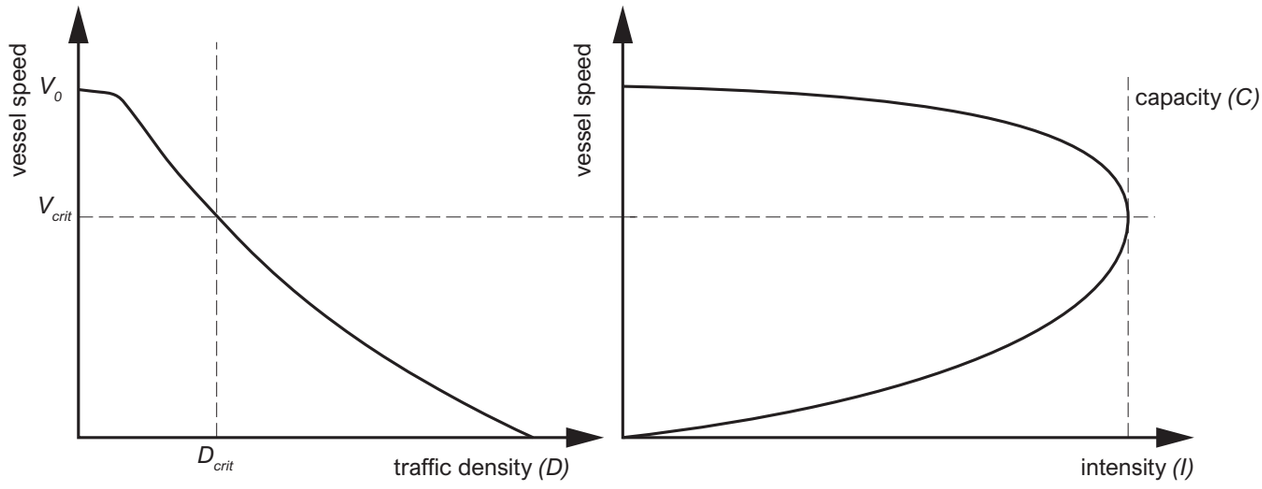

Figure 2.5: Relationship between vessel speed and traffic density, capacity and load (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

If we combine the curve for the traffic load with another largely empirical relationship, viz. between the vessel speed and the propulsion power required to sustain it (as a proxy of energy consumption and emissions), we can find what is called the economical speed (Figure 2.6). Note that this speed is larger than the critical speed, which corresponds with the highest throughput of the waterway. This means the operational optimum of an individual vessel differs from that of the waterway as a whole.

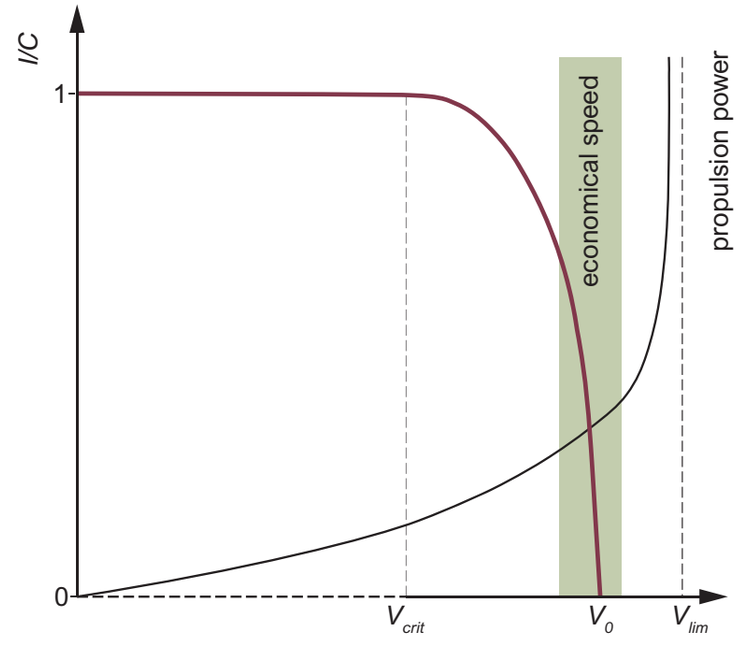

Figure 2.6: Economical vessel speed (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

The maximum allowable traffic density and the capacity of open waterways are determined by simulation, or by assuming a virtual space around each vessel, taking into account the stopping distance and safety margins for ship-ship interactions (Figure 2.7).

The dimensions of this virtual space are determined with the aid of photographs from a radar screen. For simplicity, its shape is taken as rectangle, with length L_v and width B_v. These dimensions depend on the vessel size, the situation (straight reach, bend) and whether the vessel is sailing upstream or downstream. In the early 1970s Rijkswaterstaat performed an extensive study on vessel traffic at the river Waal (RWS, 1976). It led to the

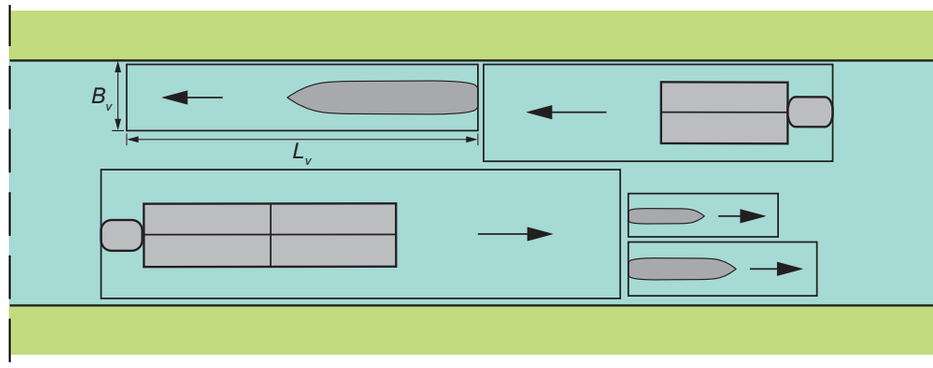

Figure 2.7: Virtual space to determine the waterway capacity (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

conclusion that L_v is proportional to the vessel length L_s, with a coefficient of proportionality of 1.45 if the vessel is sailing downstream, 1.05 when sailing upstream on a straight reach, and 1.25 when sailing upstream in a bend. Similarly, the width of the virtual area has been determined empirically for every vessel type, taking into account margins for safe navigation.

Safety margins like these, but also those included in waterway dimensions, bridge heights, etc., are by definition empirical, if only because the events against which they are supposed to protect are not exactly known. Yet, absolute safety cannot be guaranteed; there is always a residual risk.

So far, we have considered waterway performance from a fixed standpoint, looking at how many vessels or how much cargo can pass per unit time. Instead, one may also consider it from a skipper's standpoint: "How much time do I need to spend to sail from A to B?". This time spent (sometimes called the waterway's 'resistance') depends on the vessel's speed and waiting times at bridges and locks. The vessel's speed, in its turn, is influenced by the interaction with other vessels (see Figure 2.8 for an example), so the time spent is a function of the traffic. Given the traffic's variability and the complexity of the interactions, rules of thumb are the obvious way to take this into account.

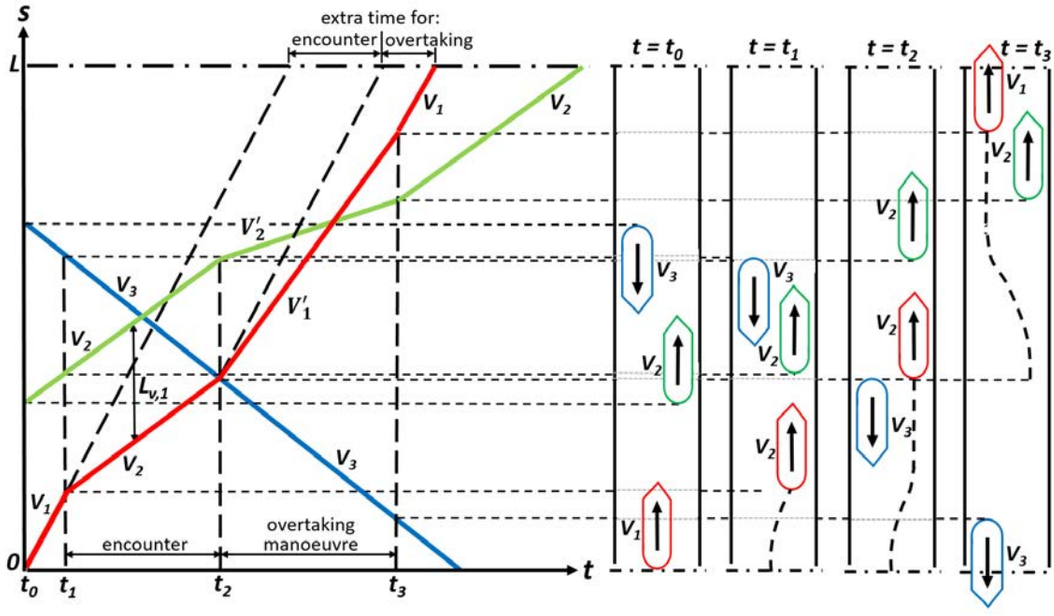

Figure 2.8: Vessel tracks during an encounter and an overtaking manoeuvre (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

The delaying effect of ship-ship interactions also explains why droughts hit a waterway's transport capacity so hard. Not only do they reduce the possible draught, and thus the load capacity of the vessels, they also lead to a higher traffic intensity (more ships needed to transport the same amount of cargo), hence a further capacity reduction. A rule of thumb often used per vessel in case of low water is the load capacity reduction in relation to the draught reduction, e.g. 'every decimetre of draught reduction means a load capacity loss of 100 ton for a Class IV inland vessel'. Further see Section 4.1.1.

In many waterways there are obstacles such as bridges, weirs and locks, or other delaying elements such as bends or (temporary) obstacles. They generally take extra time or cause delays, because vessels have to make complex manoeuvres to pass a bridge (Figure 2.9), have to wait for other vessels before being able to pass a bridge, wait for movable bridges to open, need time to pass a lock, need manoeuvring to navigate a bend, et cetera.

Figure 2.9: Passing a bridge may require careful manoeuvring and take time (image by Quistnix is licenced under CC BY-SA 2.5).

2.2.3 Lock performance

Locks, if present, are usually determining factors for a waterway's capacity. This means that locks have to meet certain capacity requirements in order to comply with the other parts of the waterway. An extra complication is that many locks are not only used by cargo ships, but also by pleasure craft (Figure 2.10). Although cargo ships have priority, locking the pleasure craft inevitably takes extra time, so causes extra transport delays.

Figure 2.10: Locks are not only used by cargo ships (image by https://beeldbank.rws.nl, Rijkswaterstaat).

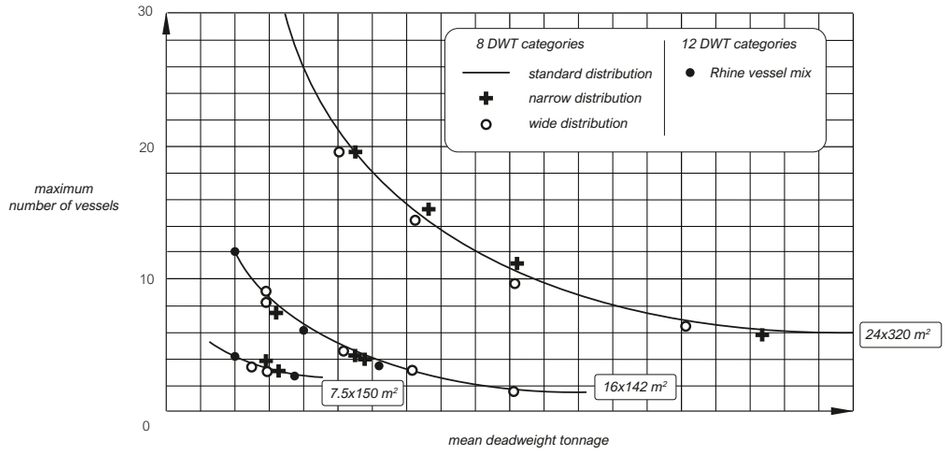

Figure 2.11: Maximum number of vessels in a lock chamber (reworked from Kooman and De Bruijn, 1975, by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

The investment involved in building a lock is too high for the design to be based on rules of thumb. Yet, empirical relationships can be of use in (pre)feasibility studies, to characterise existing lock systems, or for early warning of an upcoming capacity problem. A relevant design question is how many vessels fit into a lock chamber of a given size, given the vessel mix. As an example, Figure 2.11 shows observation-based relationships for various chamber sizes and various characteristic vessel mixes at the time.

Figure 2.12 shows an example of lock characterisation, for a lock in the river Maas near Heel (Netherlands). It shows the passage time as a function of the traffic load and for different arrival patterns. The underlying data have been produced with the Kooman model (Kooman and De Bruijn, 1975), which combines empirical rules with a look-up table from queueing theory. One may assume that these curves are characteristic of the waiting times at this lock as long as the vessel mix does not change.

If the traffic load is increasing, there comes a point where the maximum waiting time is exceeded. As this maximum corresponds with a rather low value of the traffic load, studying and planning of measures to increase the capacity have to start at a traffic load value where no problem may be suspected. Figure 2.13 shows the example of the Kreekrak locks in the Scheldt-Rhine Canal, which connects the port of Antwerp with the Rotterdam area. As a rule of thumb, capacity increase should start being considered at an I/C-value as low as 0.5.

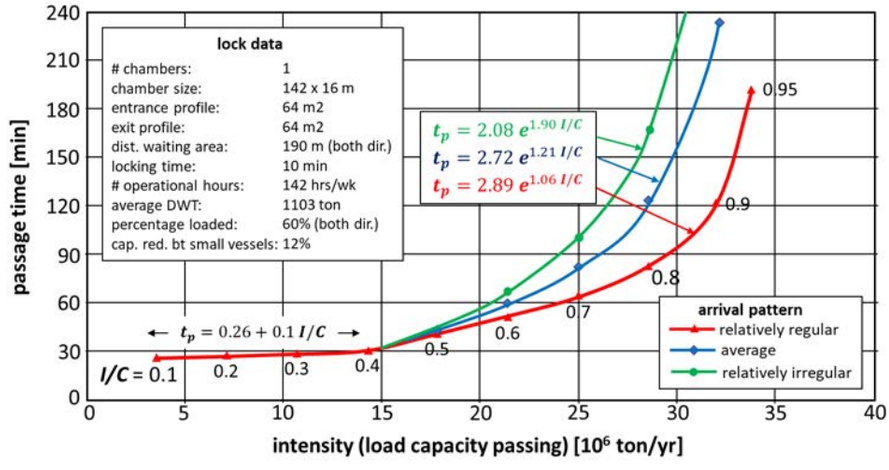

Figure 2.12: Passage times at the lock near Heel, NL (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

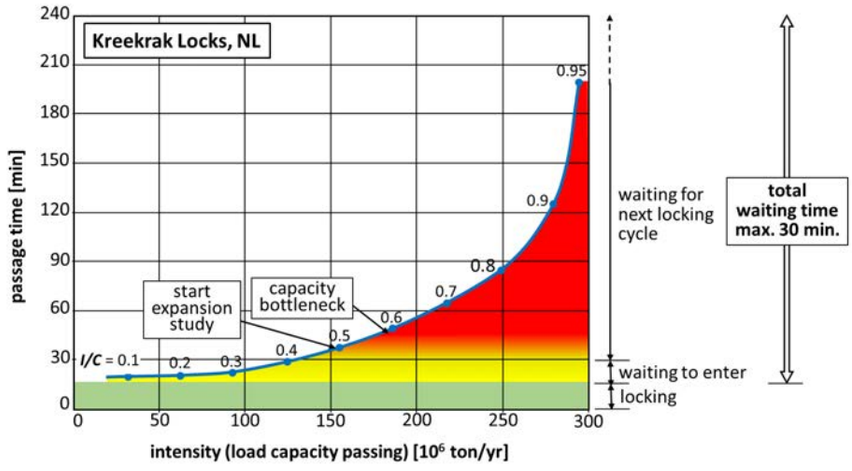

Figure 2.13: When to start considering capacity-increasing measures (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0)?

2.3 Analytical models

Analytical models are used in waterborne transport in a wide variety of forms. In general, they consist of a set of calculation rules or (differential) equations that can be handled analytically. The determination of the net present value of capital or operational expenditures (Part I – Section 2.2.4) is an example of a logic-based calculation rule, viz. the sum of a finite power series. Schijf's method to determine the water level set-down beside a ship sailing in confined water (Part III – Section 4.1) is an example of an analytical model based on conservation laws from classical physics. Simple forms of queueing theory can also be derived analytically, as we will show in this chapter.

A way of model classification is by the number of dimensions they involve. In the background of the NPV calculation there is a discrete representation of the dimension time (year by year). The equations underlying Schijf's model are one-dimensional in space and time, but by fixing the coordinate system to the moving vessel, the time dimension is eliminated. Time has also been eliminated from frequency-domain wave propagation models, which are one- or two-dimensional in space.

Models can be deterministic and stochastic. The NPV-calculation and Schijf's model are deterministic, while queueing theory starts from probability distributions and is therefore stochastic.

We have seen that waiting time limitations are of paramount importance to the design of ports and waterways. One of the ways to deal with this is queueing theory. Therefore, we will first focus on an analytical form of this often-used tool.

2.4 Queueing theory and Kendall's notation

The origin of queueing theory is attributed to the Danish engineer Agner Krarup Erlang (1878-1929). His work on modelling the capacity of telephone networks (Erlang, 1909) laid the theoretical foundation for modern telecommunication networks. Queueing theory focuses on systems with customers arriving at random (with a given probability distribution $Pr\{A\}$) and staying in the system until they have been served. These services also have a random character with a given probability distribution $Pr\{S\}$.

Typical aspects that are described by queueing models are:

- L_s: the long-term average number of customers present in the system,
- L_q: the long-term average number of customers waiting in the queue,
- W_s: the long-term average time present in the system, and
- W_q: the long-term average waiting time in the queue.

Queueuing systems can be characterised by the stochastic properties of arrivals, those of the services and the number of servers (e.g. berths on a quay). Further aspects that distinguish the systems we look at are the number of places in the system (finite or infinite), the size of the calling population (finite or infinite) and the queue discipline (often First In First Out, but other disciplines may apply).

Kendall (1953) defined a classification method to distinguish between different types of queueing systems, using the notation ($A/S/c$: $K/N/D$), with:

- A: the arrival process, with a random distribution of the inter-arrival time marked by a letter code,
- S: the service process, with a random distribution of the service time, marked by another letter code,
- c: the number of servers,
- K: the number of places in the system,
- N: the number of individuals in the calling population,
- D: the queue discipline.

The simplest queueing system assumes both arrivals and services to be Markovian (marking letter M), which means that their respective probabilities do not depend on past states (memoryless). Note, however, that queueing systems are not always Markovian, in which case more complex (numerical) computations are needed.

For a few queueing systems analytical solutions can be derived from first principles. We will show this for one of them, with Kendall notation ("$M/M/1$: ∞/∞/FIFO"), i.e.. Markovian arrival and service processes, one server, an infinite number of places, an infinite calling population and a First In First Out queue discipline.

2.4.1 Arrival, service and queueing processes

Arrival process

Vessels waiting to enter a port, form a good example where queueing theory can be applied (Figure 2.14).

Figure 2.14: Ships queueing for the Port of Singapore (image by shawnanggg is free to use under the Unsplash Licence).

In general, the arrival process of ships calling at port is of a stochastic nature and arrival times are stochastically independent, i.e. future arrivals don't depend on arrivals in the past. Such stochastic processes are called Poisson processes. The most convenient way of representing the arrival process is to look at the intervals between successive arrivals at the port, the interarrival times. Table 2.5 and Figure 2.16 give an example of interarrival times, taking hourly intervals and assuming all arrivals to take place within 8 hours.

inter arrival time	number	%	cumulative %
2h	3	5	5
3h	10	17	22
4h	12	20	42
5h	15	25	67
6h	14	23	90
7h	5	8	98
8h	1	2	100

Table 2.5: Example of interarrival times (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

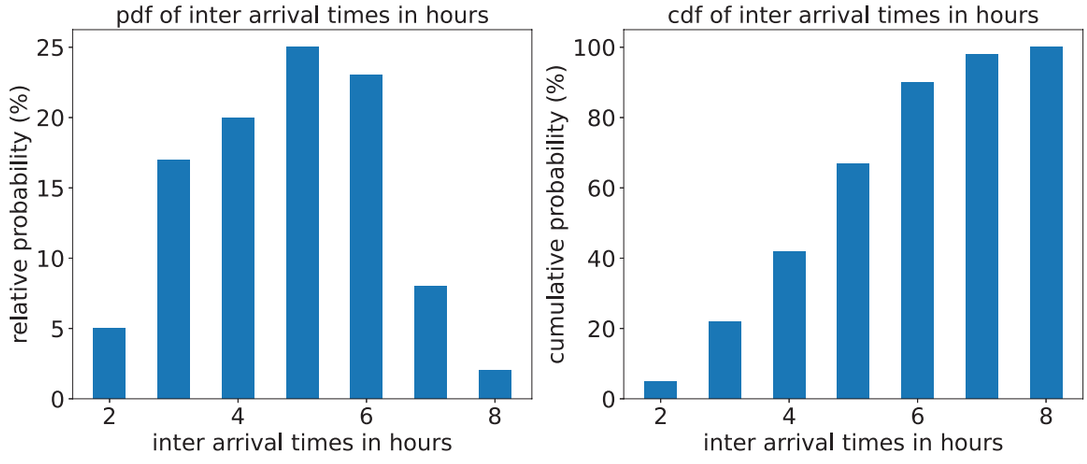

Figure 2.15: Relative probability and cumulative distribution of the interarrival times as in Table 2.5 (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

Probability distributions are usually given in the form of a continuous or discrete probability distribution function (pdf), which relates the relative probability of occurrence to an independent variable, in this case time. By definition, the time-integral of a continuous pdf, as well as the infinite sum of a discrete pdf, equals 1.

Interarrival times in a Poisson process are typically exponentially distributed (Markov-process). This means that the probability that the interarrival time T lies between s and t is given by

$$Pr\{s < T \leq t\} = e^{-\lambda s} - e^{-\lambda t} \tag{2.2}$$

where λ is the average arrival rate. i.e. the average number of arrivals per unit time. The corresponding pdf reads

$$f(t) = \lambda e^{-\lambda t} \ (t > 0) \tag{2.3}$$

The average interarrival time can be derived from

$$\overline{T} = \int_0^\infty t f(t) dt = \frac{1}{\lambda} \tag{2.4}$$

If N is the number of arrivals in a time interval of fixed length t, then this discrete random variable has the Poisson distribution and the probability that $N = k$ follows from:

$$Pr\{N = k\} = e^{\lambda t} \frac{(\lambda t)^k}{k!} \tag{2.5}$$

Figure 2.16 gives examples of continuous exponential and discrete Poisson distributions.

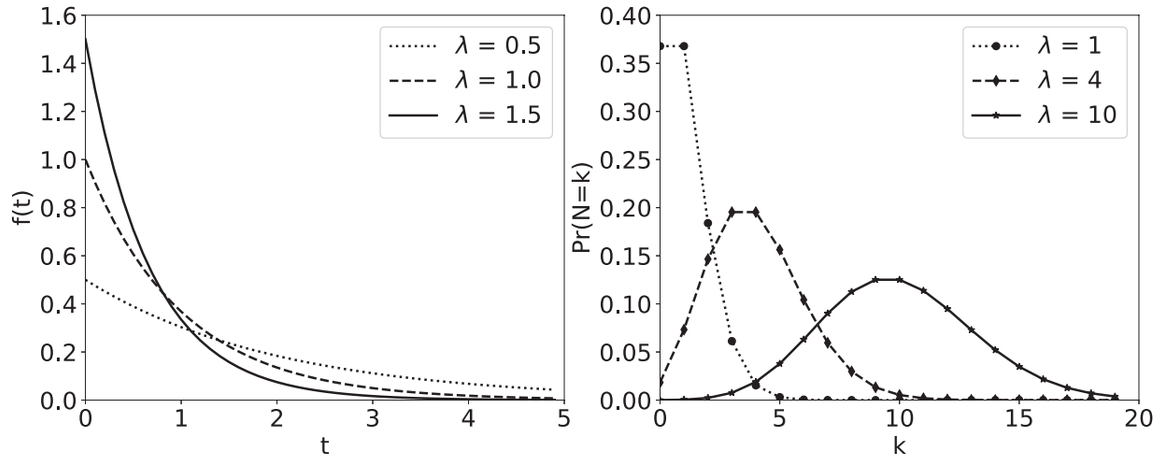

Figure 2.16: Continuous exponential distributions (left) and discrete Poisson distributions (right) (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

Service process

The time taken to serve ships on a berth obviously influences the length of the queue that may form. Even a system with sufficient berths to meet the average arrival rate of ships will still experience queue formation from time to time.

Considering service times as a random process, their probability distribution needs to be known before a study can be made. In port systems the total service time often consists of several different stages, each with its own probability distribution. If we assume each stage to have a negative exponential distribution with parameter $k\lambda$, the combination of stages is described by the Erlang-k distribution (Kendall's notation E_k):

$$f(t) = \frac{(k\mu)^k t^{k-1}}{(k-1)!} e^{-k\mu t} \text{ for:} t > 0 \tag{2.6}$$

in which μ is the average service rate per berth and $1/\mu$ the average total service time, i.e. the expected value of the sum of service times of the different stages. By changing the parameter k one can fit this distribution to observations (Figure 2.17). Taking $k = 1$ yields the negative exponential distribution of Equation 2.3.

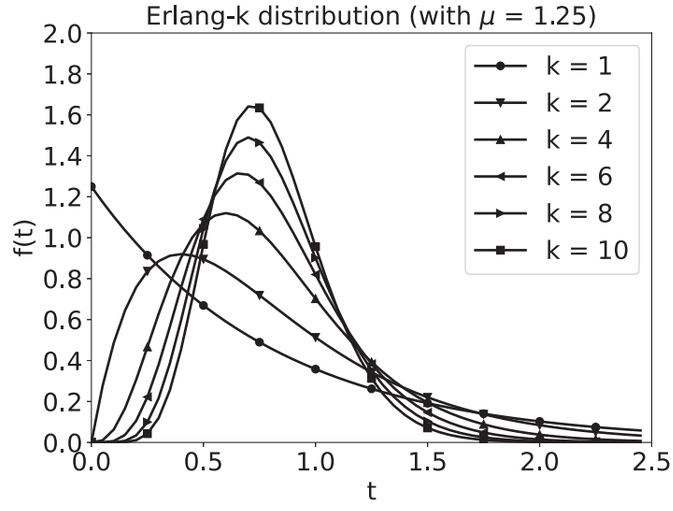

Figure 2.17: Erlang-k distributions for different values of k (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

Queueing

In situations where a queue of several customers has been formed there must be some way of deciding which customer (ship) is to be served next. Firstly, there are basically two different types of queues (Figure 2.18): in a row for each server (multiple queues), or in a single line for all servers (single queue). Here we will consider the latter form.

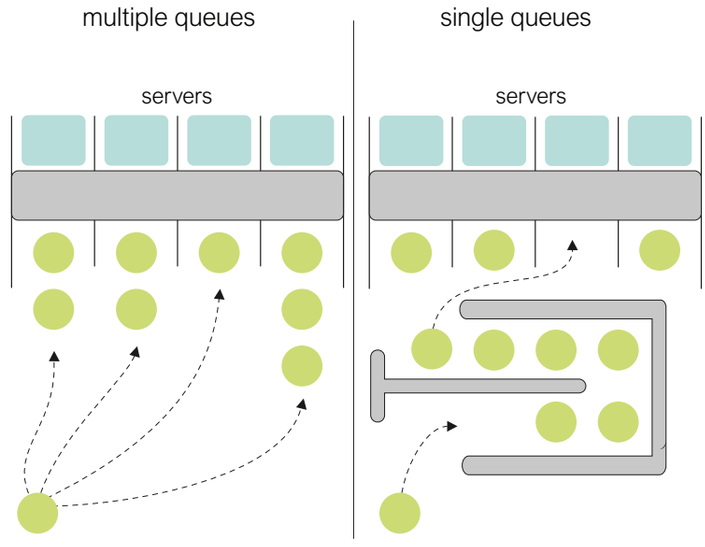

Figure 2.18: Two forms of queueing (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

In single-line queueing there are various rules (queue disciplines) to determine who is served next:

1. on the basis of arrival time,
   - First In First Out (FIFO) / First Come First Serve (FCFS),
   - Last In First Out (LIFO) / Last Come First Serve (LCFS),,
   - Service In Random Order (SIRO),
2. on the basis of service time requirement,
   - Shortest Processing Time First (SPTF),
3. on the basis of priority.
   - Priority Queueing (PQ),
   - et cetera.

Depending on the performance of a queueing system, customers may display various types of behaviour:

- *Balking* – when a customer perceives the waiting times to be too long, they may decide not to join the queue, and as a consequence not become part of the queueing system,
- *Jockeying* – when a customer decides to switch between queues in the anticipation that they will be served faster, and
- *Reneging*: when customers have already entered the queueing system, but decide to leave if they have waited too long.

Figure 2.19 shows the how the probability of the waiting time to exceed a certain value t varies with the queueing discipline. Clearly, the variance of a LIFO discipline is larger than that of a FIFO arrangement.

A port or terminal operator will try to organise the port/terminal operations in such a manner, that customers will perceive the service to be of satisfactory quality. Just like for retail stores, customers may look for alternatives when they feel that service levels are insufficient. Once customers have made the switch to an alternative terminal it can be hard to win them back. It is for this reason that queueing models are often used in the evaluation of (future) performance.

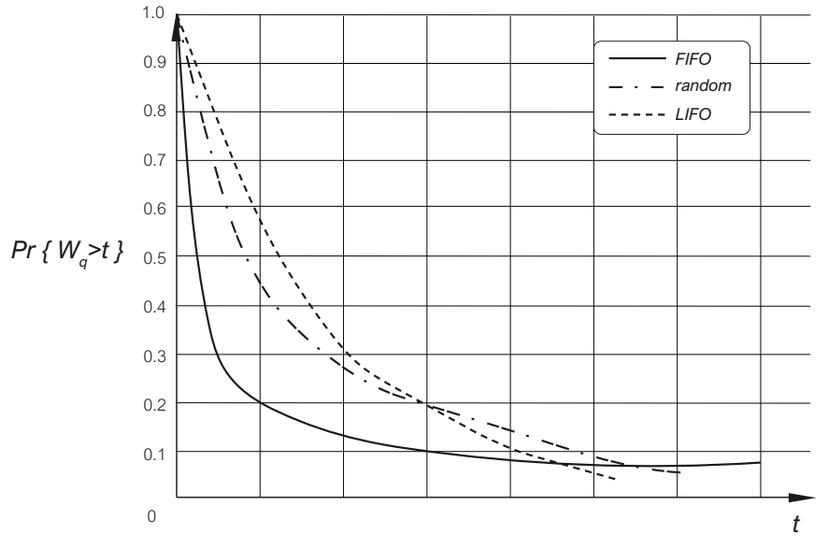

Figure 2.19: Effect of queue disciplines on waiting times (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

2.4.2 ("$M/M/1$: ∞/∞/FIFO") queueing theory

When a ship approaches a port, it is not obvious that it can be served immediately. The port operator has to negotiate between the capital and operational costs of the port and the waiting time for incoming ships. Since ships arrive generally at random, service without waiting times would be economically suboptimal. On the other hand, too long waiting times make the port less attractive to shipping lines. Clearly, these conflicting interests call for optimisation.

If a ship requires only a single service point and a single service before departing again, the factors determining the system's behaviour are (see Figure 2.20):

1. ship arrivals,
2. the service time per ship,
3. the service system (queue discipline, number of berths)

The queueing system requires specifying the statistics of the arrival and service processes and the number of berths in the system. For "$M/M/1$: ∞/∞/FIFO"-queues an analytical solution can be derived.

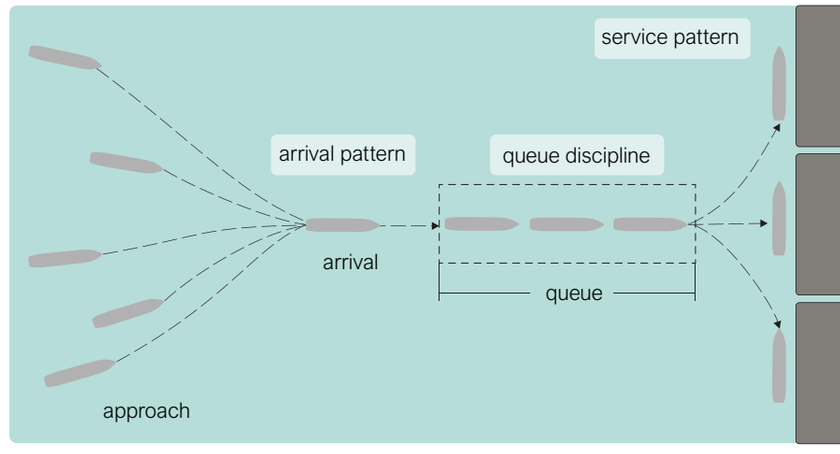

Figure 2.20: Schematic of a port service system (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

Box model

We start with the general box model formulation (see Figure 2.21 and Equation 2.7):

Figure 2.21: Box model of a queueing system (modified from image by Tsaitgaist, by TU Delft – Ports and Waterways is licenced under CC BY-SA 3.0).

$$\frac{\partial f}{\partial t} = F_{in} - F_{out} + P_f - D_f \tag{2.7}$$

When it comes to queueing systems, the abstract property f is taken to be the number of customers n present in the system. The flux of customers entering the system, F_{in}, is represented by the rate of arrivals (arrivals per hour). The flux of customers leaving the system, F_{out}, is represented by service rate μ (departures per hour). Assuming no customers are dissipated or produced inside the system, this would result in the following balance equation:

$$\frac{\partial n}{\partial t} = \lambda - \mu \tag{2.8}$$

Stochastic formulation of the balance equation

Equation 2.8 simply states that the change in the number of customers in the system is determined by the difference between the number of customers arriving and those leaving after having been served. This intuitive logic is easy to follow when the actual values of $n(t)$, λ, μ and the time step are known. In queueing theory, however, one of the challenges is that arrival rates and service times are mostly stochastic rather than deterministic. We will therefore rewrite Equation 2.8 using stochastic variables and look at the probability that at time t there are n customers in the system, $P_n(t)$. Using the box model approach we can now describe the system by:

$$\begin{aligned} P_n(t+h) = & \cdots \\ & P_{n-1}(t) \cdot P(arr.=1) \cdot P(dep.=0) + \cdots \\ & P_n(t) \cdot P(arr.=0) \cdot P(dep.=0) + \cdots \\ & P_n(t) \cdot P(arr.=1) \cdot P(dep.=1) + \cdots \\ & P_{n+1}(t) \cdot P(arr.=0) \cdot P(dep.=1) \end{aligned} \tag{2.9}$$

Basically, Equation 2.9 states that the probability that at time $t+h$ there are n customers in the system is equal to:

- the probability that $n-1$ people are present at t, times the probability that during time step h 1 arrival and 0 departures take place,
- plus the probability that n people are present at t, times the probability that during time step h 0 arrivals and 0 departures take place,
- plus the probability that n people are present at t, times the probability that during time step h 1 arrival and 1 departure take place,
- plus the probability that $n+1$ people are present at t, times the probability that during time step h 0 arrivals and 1 departure take place.

We further assume that time step h is so small that only one event can take place. As a consequence, the combination of 1 arrival *and* 1 service cannot take place during the same time step h. Note, furthermore, that the probability of 1 arrival during h is λh, whereas the probability of 0 arrivals is $(1 - \lambda h)$. Likewise, the probability of 1 service during h is μh, and the probability of 0 services is $(1 - \mu h)$. We can now rewrite Equation 2.9 to:

$$\begin{aligned} P_n(t+h) = & \cdots \\ & P_{n-1}(t) \cdot \lambda h \cdot (1 - \mu h) + \cdots \\ & P_n(t) \cdot (1 - \lambda h) \cdot (1 - \mu h) + \cdots \\ & P_{n+1}(t) \cdot (1 - \lambda h) \cdot \mu h \end{aligned} \tag{2.10}$$

Performing algebraic operations and neglecting higher order terms yields:

$$\begin{aligned} P_n(t+h) = & \cdots \\ & P_{n-1}(t)\lambda h - \underbrace{P_{n-1}(t)\lambda\mu h^2}_{neglect} + \cdots \\ & P_n(t) - P_n(t)\lambda h - P_n(t)\mu h + \underbrace{P_n(t)\lambda\mu h^2}_{neglect} + \cdots \\ & P_{n+1}(t)\mu h - \underbrace{P_{n+1}(t)\lambda\mu h^2}_{neglect} \end{aligned} \tag{2.11}$$

Rearranging yields:

$$\begin{aligned} \frac{P_n(t+h) - P_n(t)}{h} = & \cdots \\ & P_{n-1}(t)\lambda - P_n(t)(\lambda + \mu) + P_{n+1}(t)\mu \end{aligned} \tag{2.12}$$

For model aspects L_s, L_q, W_s and W_q we are interested in long-term averages, or steady states. We will therefore apply the steady state assumption, meaning that probabilities are no longer time dependent and $\frac{P_n(t+h)-P_n(t)}{h} = 0$. Applying this to Equation 2.12 yields the following equation:

$$P_n(\lambda + \mu) = P_{n-1}\lambda + P_{n+1}\mu \tag{2.13}$$

which relates the probability that n customers are in the system to the probabilities that $n-1$ and $n+1$ customers are in the system.

The limit case of zero customers in the system

We will now consider the limit case of zero customers in the system. First again in words:

$$\begin{aligned} P_0(t+h) = & \cdots \\ & P_1(t) \cdot P(arr. = 0) \cdot P(dep. = 1) + \cdots \\ & P_0(t) \cdot P(arr. = 0) \cdot P(dep. = 0) \end{aligned} \tag{2.14}$$

Converting this into mathematical formulations yields:

$$\begin{aligned} P_0(t+h) = & \cdots \\ & P_1(t) \cdot (1-\lambda h) \cdot \mu h + \cdots \\ & P_0(t) \cdot (1-\lambda h) \cdot 1 \end{aligned} \tag{2.15}$$

Logically, the probability of 'no service' must be 1 if there are no customers in the system. Performing algebraic operations and again neglecting second-order terms yields:

$$\begin{aligned} P_0(t+h) = & \cdots \\ & P_1(t)\mu h - \underbrace{P_1(t)\lambda\mu h^2}_{neglect} + \cdots \\ & P_0(t) - P_0(t)\lambda h \end{aligned} \tag{2.16}$$

After rearrangement this yields:

$$\begin{aligned} \frac{P_0(t+h) - P_0(t)}{h} = & \cdots \\ & P_1(t)\mu - P_0(t)\lambda \end{aligned} \tag{2.17}$$

Since in the steady state the time-variation of P_0 equals zero, we can rewrite Equation 2.17 into:

$$P_1\mu = P_0\lambda \tag{2.18}$$

whence:

$$P_1 = \frac{\lambda}{\mu} P_0 \tag{2.19}$$

Combining the zero-customer condition with the general equation

Resuming Equation 2.13

$$P_n(\lambda+\mu) = P_{n-1}\lambda + P_{n+1}\mu$$

and substituting $n = 1$ while making use of Equation 2.19 yields:

$$P_1(\lambda+\mu) = P_0\lambda + P_2\mu$$

This can be rewritten to:

$$P_1\lambda + P_1\mu = P_0\lambda + P_2\mu \tag{2.20}$$

If we substitute Equation 2.18 we get:

$$P_1\lambda + P_0\lambda = P_0\lambda + P_2\mu$$

which can be rewritten to:

$$P_2 = \frac{\lambda}{\mu} P_1 \tag{2.21}$$

With Equation 2.19 this can be rewritten to:

$$P_2 = \left(\frac{\lambda}{\mu}\right)^2 P_0 \tag{2.22}$$

The ratio λ/μ is commonly replaced by the symbol ρ. If we do so we can show that:

$$\begin{aligned} P_1 &= \rho P_0 \\ P_2 &= \rho P_1 = \rho^2 P_0 \\ P_3 &= \rho P_2 = \rho^3 P_0 \\ P_n &= \rho^n P_0 \end{aligned} \tag{2.23}$$

Finding the value of P_0

Equation 2.23 shows that the value of P_0 is the key to all other probabilities P_n. We can find P_0 by applying the integral property:

$$\begin{aligned} \sum_{n=0}^{\infty} P_n &= 1 \\ P_0 + P_1 + P_2 + \cdots \infty &= 1 \\ P_0 + \rho P_0 + \rho^2 P_0 + \cdots \infty &= 1 \\ P_0 \left[1 + \rho + \rho^2 + \rho^3 + \cdots \infty\right] &= 1 \end{aligned} \tag{2.24}$$

In this equation the expression:

$$\left[1 + \rho + \rho^2 + \rho^3 + \cdots \infty\right]$$

is a so-called infinite geometric series. In other words there is a fixed ratio between the successive terms into infinity; in this case that ratio is ρ. In general terms an infinite geometric series can be written as:

$$a + ar + ar^2 + ar^3 + \cdots \infty = \sum_{k=0}^{\infty} ar^k$$

The result of the infinite summation is:

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$$

as long as $|r| < 1$.

In Equation 2.24 the general variable r corresponds with ρ and the value of a is equal to 1. A fundamental requirement for the "M/M/1: ∞/∞/FIFO" type models, where the calling population size is infinitely large, is that the long-term average arrival rate should be smaller than the long-term average service rate, or in other words $\lambda/\mu = \rho < 1$. With this requirement satisfied, the infinite geometric series in Equation 2.24 can be replaced:

$$\left[1+\rho+\rho^2+\rho^3+\cdots\infty\right]=\frac{1}{1-\rho}$$

enabling us to rewrite Equation 2.24 to:

$$\begin{aligned}
&P_0\left[1+\rho+\rho^2+\rho^3+\cdots\infty\right]=1\\
&P_0\left[\frac{1}{1-\rho}\right]=1\\
&P_0=1-\rho
\end{aligned}\tag{2.25}$$

With this we can now expand and rewrite Equation 2.23 to:

$$\begin{aligned}
P_0&=1-\rho\\
P_1&=\rho P_0=\rho(1-\rho)\\
P_2&=\rho^2 P_0=\rho^2(1-\rho)\\
P_3&=\rho^3 P_0=\rho^3(1-\rho)\\
P_n&=\rho^n P_0=\rho^n(1-\rho)
\end{aligned}\tag{2.26}$$

This means that for an M/M/1 model in fact the value of ρ suffices to determine P_0, P_1 up to P_n.

Deriving expressions for queue length and residence times

The first aspect for which we will derive an expression is the number of customers n expected to be present in the system, L_s. L_s is defined as the sum for $n=1,2,\cdots\infty$ of the product nP_n. Rewriting in terms of P_0, bringing P_0 outside the summation, rewriting with a differential and reordering to reveal an infinite geometric series (!), rewriting, integrating and rewriting again, reveals that L_s can be expressed as a function of ρ:

$$\begin{aligned}
L_s=\sum_{n=0}^{\infty}nP_n&=\sum_{n=0}^{\infty}n\rho^nP_0\\
&=\rho P_0\sum_{n=0}^{\infty}n\rho^{n-1}\\
&=\rho P_0\sum_{n=0}^{\infty}\frac{d}{d\rho}\rho^n\\
&=\rho P_0\frac{d}{d\rho}\sum_{n=0}^{\infty}\rho^n\\
&=\rho P_0\frac{d}{d\rho}\left[1+\rho+\rho^2+\cdots\infty\right]\\
&=\rho P_0\frac{d}{d\rho}\frac{1}{1-\rho}\\
&=\rho P_0\frac{1}{(1-\rho)^2}\\
&=\frac{\rho(1-\rho)}{(1-\rho)^2}\\
&=\frac{\rho}{1-\rho}
\end{aligned}\tag{2.27}$$

The long-term average number of customers in the system, L_s, is equal to the long-term average of the number the queue plus the number being served. The probability that there are no customers in the system is P_0, so the probability that customers are being served is $1 - P_0 = \rho = \lambda/\mu$. Since only one customer at a time can be served, this means that the long-term average number of customers being served is also λ/μ. Hence:

$$L_s = L_q + \frac{\lambda}{\mu} \tag{2.28}$$

The long-term average number of customers in the system, L_s, can also be obtained by multiplying by the average customer arrival rate, λ, with the long-term average time that a customer spends in the system, W_s:

$$L_s = \lambda W_s \tag{2.29}$$

Similarly, the long-term average number of customers in the queue, L_q, can be obtained by multiplying by the average customer arrival rate, λ, with the long-term average time that a customer spends in the queue, W_q:

$$L_q = \lambda W_q \tag{2.30}$$

Equation 2.29 and Equation 2.30 are known as Little's equations and reflect Little's theorem. This states that the average number of customers in a system is equal to the average customer arrival rate times the average time that customers spend in the system. It is quite remarkable that this result does *not* depend on factors such as the distribution of arrivals and services, the size of the customer population, et cetera.

Summary

In summary, the following equations are of interest when working with "M/M/1: 1/1/FIFO" systems:

$$P_0 = 1 - \rho \tag{2.31}$$
$$P_n = \rho^n P_0 \tag{2.32}$$
$$L_s = \frac{\rho}{1 - \rho} \tag{2.33}$$
$$L_q = L_s - \frac{\lambda}{\mu} = \frac{\rho^2}{1 - \rho} \tag{2.34}$$
$$W_s = \frac{L_s}{\lambda} \tag{2.35}$$
$$W_q = \frac{L_q}{\lambda} \tag{2.36}$$

The probability that there are 0, 1 or n customers in the system can be used to establish utilisation rates. Note that the sum of all these probabilities is 1.

P_0 gives the probability that there are 0 customers in the system. The system's utilisation rate is determined by the probability that the number of customers is greater than 0:

$$\begin{aligned} P_{n>0} &= 1 - P_0 \\ &= 1 - (1 - \rho) \\ &= \rho \end{aligned} \tag{2.37}$$

The utilisation rate of the queue can be derived from the probability that there is at least 1 person in the queue, i.e. the probability that the number of customers in the system is greater than 1 (one being served):

$$\begin{aligned} P_{n>1} &= 1 - P_0 - P_1 \\ &= 1 - (1-\rho) - \rho(1-\rho) \\ &= \rho - \rho + \rho^2 \\ &= \rho^2 \end{aligned} \tag{2.38}$$

Example box 2.1: Example economic optimization

1. Case
A transhipment company owns one berth at a port.

Ships arrive for unloading every 12 hours on average, with a negative exponential distribution of interarrival times, so

$\lambda = 1/12$ hr^{-1}

Ship sizes vary widely, yielding negative exponential distribution of serves times T, with parameter $\mu = 1/T$ (to be determined)

The running costs RC of a berth are inversely proportional to the service time:

$RC = 10{,}000\ \mu$ \$/day

Delay costs are 1,000 \$ per ship per day.

2. Objectives
To determine:
- the most economical unloading time,
- the corresponding average berth occupancy, and
- the average delay per ship.

3. Solution
Daily running costs: $10{,}000\ \mu$

Average number of ships/day: 2

Average delay per ship: $\frac{\rho}{\mu(1-\rho)}$ hours, with $\rho = \frac{1}{12\mu}$

Delay costs per day: $2\frac{\rho}{\mu(1-\rho)}\frac{1000}{24}$ \$/day

Total daily costs: $10{,}000\ \mu + \frac{2{,}000}{24\mu(12\mu-1)}$\$/day $= \frac{10{,}000}{T} + \frac{2{,}000T^2}{24(12-T)}$

Economic optimum: $d(\text{costs})/dT = 0$

Hence: $T^4 - 24T^3 + 120T^2 - 2,889T + 17,280 = 0$

Only one root makes sense: $T = 6.163$ hr^{-1}

Correspondingly, $\mu = 0.163$ hr^{-1} and $\rho = 0.511$

Optimum berth occupancy: $100\rho = 51.1$ %

Long-term average delay per ship: $\frac{\rho}{\mu(1-\rho)} = \frac{0.511 \cdot 6.136}{1-0.511} = 6.29$ hrs $= 0.26$ days

Probability that on arriving, ship has no delay: $\mu(1-\rho) = \frac{0.489}{6.136} = 0.08$, so only 8%

Example box 2.1 – continued on next page

Example box 2.1 – continued from previous page

Average number of ships in the system: $\frac{\rho}{1-\rho} = \frac{0.511}{0.489} = 1.04$

Average number of ships in the queue: $\frac{\rho^2}{1-\rho} = 0.53$

The example shows that, although the average service time per ship is nearly half of the average interarrival time, the probability that a ship can be served immediately is only 8%. This is, of course, due to the random nature of the arrival and service processes.

2.4.3 More complex systems

The "M/M/1: ∞/∞/FIFO" system analysed in the previous section gives results that are uniquely related to ρ and λ. Analytical solutions are also possible for certain other combinations, such as "M/M/n: ∞/∞/FIFO" for a larger number of berths (see Table 2.6; note that for $n = 1$ the results agree with those above).

Berth occup. (ρ/n)	**Number of berths** n									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
10 %	0.1111	0.0101	0.0014	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
20 %	0.2500	0.0417	0.0103	0.0030	0.0010	0.0003	0.0001	0.0000	0.0000	0.0000
30 %	0.4286	0.0989	0.0333	0.0132	0.0058	0.0027	0.0013	0.0006	0.0003	0.0002
40 %	0.6667	0.1905	0.0784	0.0378	0.0199	0.0111	0.0064	0.0039	0.0024	0.0015
50 %	1.0000	0.3333	0.1579	0.0870	0.0521	0.0330	0.0218	0.0148	0.0102	0.0072
60 %	1.5000	0.5625	0.2956	0.1794	0.1181	0.0819	0.0589	0.0436	0.0330	0.0253
70 %	2.3333	0.9608	0.5470	0.3572	0.2519	0.1867	0.1432	0.1128	0.0906	0.0739
80 %	4.0000	1.7778	1.0787	0.7455	0.5541	0.4315	0.3471	0.2860	0.2401	0.2046
90 %	9.0000	4.2632	2.7235	1.9693	1.5250	1.2335	1.0285	0.8796	0.7606	0.6687

Table 2.6: Average waiting time/service time ratio in an "M/M/n: ∞/∞/FIFO" queueing system (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

The expressions resulting from these analyses, however, soon become rather complicated. Results are therefore often presented in the form of look-up tables or graphs (see the examples in Example box 2.1 and Figure 2.22).

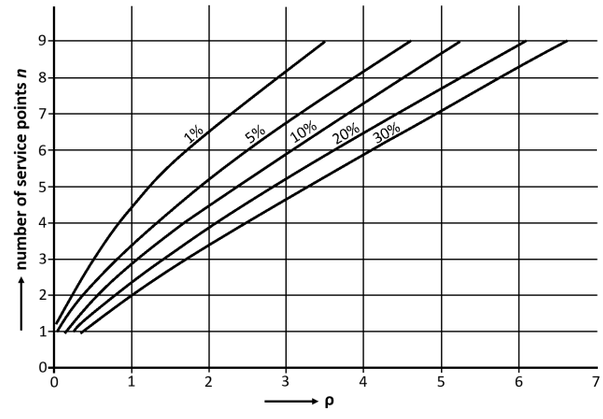

Figure 2.22: Probability that an arriving ship has to wait before being served (queueing system "M/M/n: ∞/∞/FIFO") (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

Example box 2.2: Summary of the M/M/n system

Summary of the M/M/n-system
Probability that the system is empty:

$$P(0) = \left[1 + \rho + \frac{\rho^2}{2!} + \cdots + \frac{\rho^{n-1}}{(n-1)!} + \frac{\rho^n}{n!}\frac{n}{n-\rho}\right]^{-1} = \left[\sum_{j=0}^{n-1}\frac{\rho^j}{j!} + \frac{\rho^n}{n!}\frac{n}{n-\rho}\right]^{-1}$$

Probability of j ships in the system:

$$\begin{aligned} P(j) &= P(0)\frac{\rho^j}{j!} \qquad \text{for } j < n \\ &= P(0)\frac{\rho^n}{n!}\left(\frac{\rho}{n}\right)^{j-n} \text{ for } j \geq n \end{aligned}$$

Probability that an arriving ship has to wait before being served:

$$S_q = P(0)\frac{\rho^n}{n!}\frac{n}{n-\rho} = C(n,p)$$

in which $C(n,p)$ is called Erlang's C-formula.

Long-term average number of ships in the system:

$$L_s = P(0)\rho\sum_{j=0}^{n-2}\frac{\rho^j}{j!} + C(n,p)\left[n + \frac{\rho}{n-\rho}\right]$$

Long-term average number of ships in the queue.

$$L_q = C(n,p)\frac{\rho}{n-\rho}$$

Long-term average time in the system:

$$W_s = \frac{L_s}{\lambda}$$

Long-term average waiting time:

$$W_q = \frac{L_q}{\lambda} = C(n,\rho)\frac{1}{n\mu}\frac{n}{n-\rho}$$

Utilisation rate:

$$u = \frac{\rho}{n}$$

Example box 2.3: Design of a break bulk terminal

1. Case
We are to design a break bulk terminal.

Negative exponential distribution of interarrival times, with parameter λ.

Negative exponential distribution of service times, with parameter μ

A working day consists of two 8-hour shifts. There are 6 working days per week, and 50 working weeks per year.

Cargo forecast: 600,000 ton/year

Example box 2.3 – continued on next page

Example box 2.3 – continued from previous page

Average number of gangs (groups of dock workers) employed per ship: 2.5

Production per gang: 12.5 ton/hr

Max acceptable waiting time - 0.25 times average service time

2. Objectives
To determine:
- To determine the minimum number of berths required to meet the waiting time requirements, and
- the corresponding average berth occupancy.

3. Solution
Queueing system: M/M/n: ∞/∞/FIFO

Operational hours: 2 * 8 * 6 * 50 = 4800 hr/yr

Arrival rate per berth: λ = 600,000 ton/yr

Service rate per berth: μ = 150,000 ton/yr

$\rho = \lambda/\mu = 4$

Number of berths with 100% occupancy: 4 (starting value)

Waiting time infinite → increase number of berths to 5

M/M/n-table for $n = 5$ and occupancy 0.8: $W/S = 0.5541$ → increase number of berths to 6

M/M/n-table for $n = 6$ and occupancy 0.67: $W/S = 0.152$ → meets requirement

4. Conclusion
Minimum 6 berths required, with an occupancy of 67%

2.5 Numerical approximations of queueing systems

It is not possible (or practical) to derive analytical solutions for all queueing systems; many require a numerical approach. A typical way to do this is simulation: interarrival and service times per customer are drawn at random from the given probability distributions. The other factors characterising the queueing system, viz. the number of service points, the size of the calling population, the number of places in the system and the queue discipline, are handled deterministically in the simulation software. Each customer is run through the system and waiting and service times are recorded for statistical analysis.

Table 2.7 shows an example output for 10 customers of an M/M/1 queueing system, with 8 arrivals and 9 services per hour on average. The first column gives the customer's identifier, in this case its arrival serial number. The second and third columns list the drawn values of the interarrival time (IAT) and the service time (ST), both in seconds. Time starts at $t = 0$ with no customers in the system. Customer 1 arrives after IAT(1) seconds and, since there are no other customers in the system, he or she is served immediately. So the arrival time AT(1) is equal to IAT(1) and so is the time TSB(1) that service begins. This service ends ST(1) seconds later, so at time TSE(1) = TSB(1) + ST(1). This is the moment that Customer 1 leaves the system. The time TCSS(1) the customer spent in the system is the interval between TSE(1) and AT(1). Since the system time was said to start at 0, the server was idle until the first customer arrived, causing and idle time server ITS(1) = AT(1).

c	IAT	ST	AT	TSB	TSE	TCSS	TCWQ	ITS	QL
1	770.058680	327.407626	770.058680	770.058680	1097.466305	327.407625	0.000000	770.058680	1
2	205.326085	90.992827	975.384765	1097.466305	1188.459132	213.074367	122.081540	0.000000	1
3	200.282222	90.603747	1175.666986	1188.459132	1279.062879	103.395892	12.792145	0.000000	0
4	127.819803	182.644607	1303.486789	1303.486789	1486.131396	182.644607	0.000000	24.423911	1
5	83.275849	282.955522	1386.762638	1486.131396	1769.086918	382.324280	99.368758	0.000000	1
6	192.753662	149.813260	1579.516300	1769.086918	1918.900177	339.383878	189.570618	0.000000	0
7	635.795797	395.303652	2215.312096	2215.312096	2610.615748	395.303652	0.000000	296.411919	3
8	141.942446	17.955685	2357.254543	2610.615748	2628.571433	271.316890	253.361205	0.000000	2
9	159.547619	1314.952441	2516.802161	2628.571433	3943.523874	1426.721713	111.769272	0.000000	8
10	12.377204	114.134531	2529.179365	3943.523874	4057.658405	1528.479039	1414.344509	0.000000	8

Table 2.7: Example output of an M/M/1 queueing system simulation (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

Before the service of Customer 1 has ended, Customer 2 arrives, so he/she has to wait in the queue until the service is available. This causes a queue length QL(1) of 1 during the presence of Customer 1 in the system. After departure of Customer 1 we follow the same steps for Customer 2 and all subsequent customers, thus filling in the table row by row. Once a sufficient number of customers has been dealt with, the simulation results can be analysed statistically, yielding quantities such as $P(0)$, $P(n)$, L_s, L_q, W_s and W_q.

A statistic that is often used in ports and waterways performance studies is the ratio of the long-term average of the waiting time over the long-term average of the service time:

$$\frac{\overline{TCWQ}}{\overline{ST}} \tag{2.39}$$

Each cell in Table 2.6, for example, is derived from $\overline{TCWQ}/\overline{ST}$ calculation on the output of a "M/M/n: ∞/∞/FIFO" calculation using $1 \cdot 10^6$ iterations. While more iterations do tend to lead to more stable results, it should be that the stochastic nature of the calculations means simulation results are not exactly identical. This is why tables found at various places in literature tend not to be exactly identical, unless they were generated with an analytical solution.

2.5.1 Application

In the example above, as well as in Part II – Chapter 4, we have seen how tables from queueing theory are applied in the design of a terminal, especially the determination of the required number of berths. But the operations on the terminal itself (supply, transport and storage of unloaded cargo, or collection, transport and pick-up of cargo to be loaded) are also supply chain processes with random components. Numerical simulations like the one shown in Table 2.7 can be useful to optimise these processes.

The examples also made clear how strongly the design is influenced by waiting time limitations and by the random character of the arrival and service processes. Waiting time limitations are usually based on economic arguments and/or competition with other ports. One may also try to influence the design, however, by service priority pricing (thus changing the queue discipline), or by traffic regulation (thus influencing the arrival statistics). The more vessels make long-planned roundtrips, the earlier and the better one may estimate arrival times at a port, thus enabling planning in advance of services.

Queueing also plays a role in inland waterways, especially near locks. Here, too, there is a random arrival process and a more or less random service process, which at least depends on the mix of vessels to be locked. The arrival process is not necessarily completely random, especially if there are other locks not far ahead in the same waterway. In that case, a narrow normal distribution is more suitable for the interarrival times than an M- or E_k-distribution.

For many years Rijkswaterstaat has used the Kooman method (Kooman and De Bruijn, 1975) to determine waiting and passage times at locks. It is partly based on empirical relationships (e.g. of the number of vessels that fit into a lock, given the average tonnage), partly on queueing theory. At the moment, this method is being replaced by computer simulations.

Locks are very common in waterways, natural or man-made. In rivers, weir and lock systems are used to make the river navigable, also at low water levels. An example is the river Maas, which in the Netherlands has seven weir/lock combinations (Figure 2.23).

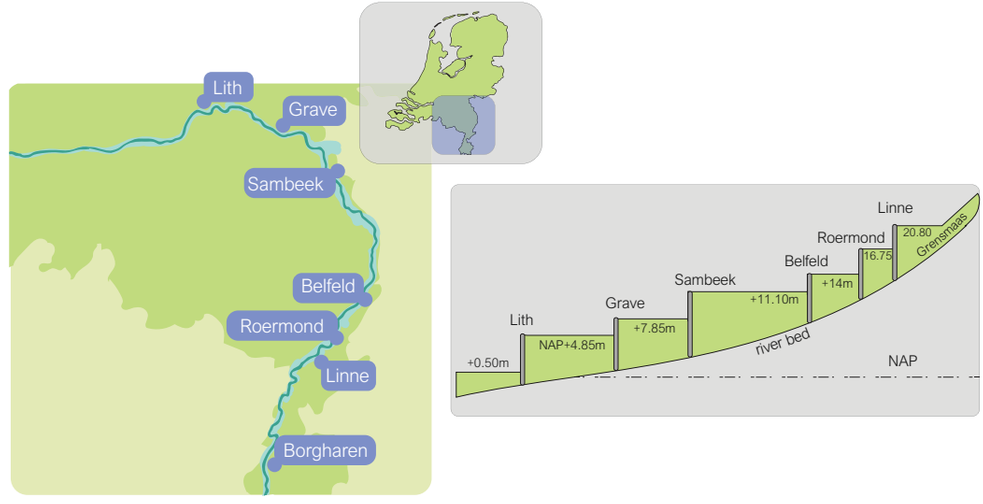

Figure 2.23: Weir/lock systems in the river Maas in the Netherlands (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

Another example is the Main-Donau Kanal, which needs as many as 16 locks to negotiate the drainage divide between the Rhine-Main and Danube basins (Figure 2.24).

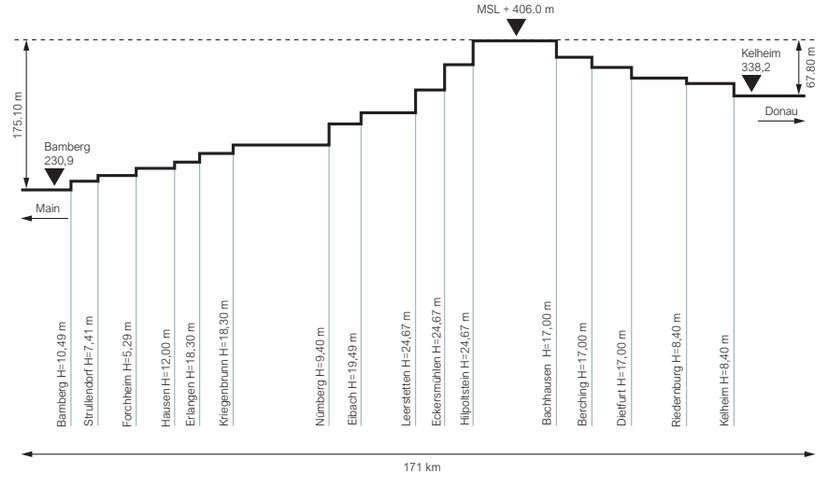

Figure 2.24: Locks in the Main-Donau Kanal (by TU Delft – Ports and Waterways is licenced under CC BY-NC-SA 4.0).

In such cases, the capacity of the waterway strongly depends on the capacity of the locks and the time spent sailing from one end to the other is dominated by the passage times of the locks. In Part III – Chapter 3 we have described how passage times and lock capacities can be determined. Note, however, that that capacity is not the only indicator of lock performance. Depending on the statistics of the interarrival time and the vessel mix in the lock, the maximum I/C ratio will be significantly smaller than 1, due to waiting time limitations (also see Figure 2.12).

2.6 Simulation models

2.6.1 National simulation models

Governments typically use information as described in the previous sections to analyse (1) whether the system under their control is or will be having problems (now or in the near future), and (2) what measures are likely going to be most effective to mitigate these problems.

As an example case we describe the modelling landscape set up and maintained by the Dutch Ministry of Public Works, to support policy making related to the waterway network. A chain of interconnected models that range from a national scale and a time horizon of 15 – 30 years, to a more local scale and and a time scale of minutes (see Figure 2.25).

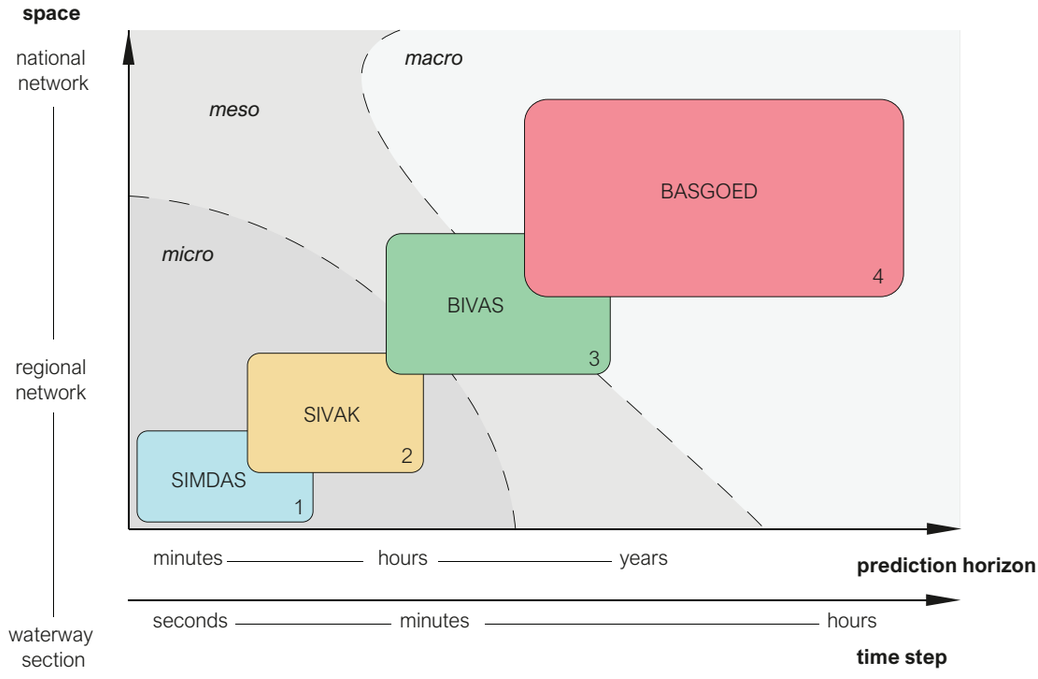

Figure 2.25: Waterways modelling landscape Rijkswaterstaat (modified from version by Tom van der Schelde, Rijkswaterstaat, by TU Delft – Ports and Waterways is licenced by CC BY-NC-SA 4.0).

BASGOED

BASGOED ('Basismodel voor Goederenvervoer') is the Netherlands' basic model to predict future transport of goods. Looking 15 – 30 years ahead, it estimates:

- the amount of goods that are produced and consumed,
- what part of this amount needs to be transported, and from which origin to which destination, and
- which transport modality will take care of what part of this transport demand.

BASGOED predicts the tons transported per cargo type and modality, on origin-destination pairs, based on sector-specific anticipated growth in combination with travel time and travel cost for the road, rail and IWT modalities.

For road transport the number of trips is also predicted. BASGOED provides crucial input for road transport: the National Model System (LMS) and the National Region Model (NRM); for rail transport: the Network Evaluation Model (NEMO); and for IWT: 'Binnenvaart Analyse Systeem' (BIVAS), 'Simulatiepakket voor Verkeersafwikkeling' (SIVAK) and 'Algemeen Scheepvaart Simulatiemodel' (SIMDAS).

BIVAS

BIVAS was developed to perform network analyses for IWT. The application provides support for answering policy-related questions on topics such as:

- traffic loads on waterways and infrastructure (locks, bridges),
- the effect of blockages in the network,
- the effect assessment of management strategies, and
- the effect assessment of maintenance scenarios.

Traffic flows are calculated based on the number of trips that need to be made on given origin-destination pairs (based on BASGOED). BIVAS checks if a trip can be made, and executes it as soon as it can. As a result BIVAS can predict the number of trips that make use of the network, but the individual trips are not simulated as such. Around key infrastructure, such as locks and bridges, the SIVAK model provides more detail. On waterway sections, in bends and on intersections, the SIMDAS model provides more detail.

SIVAK

SIVAK is a detailed simulation model for traffic flows near locks, bridges, waterway restrictions and combinations thereof. It allows for dynamic route selection and can calculate waiting times depending on arrival patterns and service rates.

Van Adrichem (2020) applied SIVAK as a benchmark, to investigate whether mathematical optimization of lock scheduling could significantly improve locking capacity.

SIMDAS

SIMDAS is a general shipping simulation model for waterway sections, bends and intersections. It can be used to calculate capacity and congestion effects on waterways. It resolves the position of vessels on the waterway and simulates interactions between vessels during encountering and overtaking manoeuvres. As such it complements SIVAK.

Verschuren (2020) investigated the effects of the low discharge extreme of 2018 on traffic flow on the river Waal. She used SIMDAS to investigate at which discharge levels the fairway would be so restricted that queue formation was likely to occur.

2.6.2 Research simulation models

The models described above have been developed with the clear purpose to support formal policy and decision making by Rijkswaterstaat and the Ministry of Public Works. Over the years much effort has gone into preparing and improving crucial inputs for the models, such as the careful specification of the waterway network (see also http://www.vaarweginformatie.nl) and the properties of the fleet. As a result, the modelling landscape forms the widely accepted platform that is commonly used to study the behaviour of the existing network and the effect of policies.

For new types of questions the robust quality of the modelling landscape can make the testing of new formulations and approaches a bit more cumbersome. Issues like calculating energy consumption and emissions on networks, or investigating the effect of low discharge extremes on the waterway capacity and fleet occupation, are typically easier investigated in more flexible systems. Once a new approach or formulation has been developed and properly tested, it may subsequently be implemented into the modelling landscape.

To facilitate the development of innovative approaches TU Delft developed a number of dedicated open source python packages that allow researchers to develop and test innovative approaches in a flexible and adaptable research environment.

We describe some of the more relevant packages here briefly.

OpenQTSim – Queueing Theory Simulation

For the numerical approximation of Kendall type queueing systems Van Koningsveld and Den Uijl (2020a) developed the python package OpenQTSim. It utilises the discrete event simulation and queue handling capabilities of the SimPy package, and the statistical functionality available in the SciPy package. Combined, this enables simulation of queueing system behaviour that emerges from inter-arrival and service times that follow predefined statistical distributions (see also Section 2.5).

The tool has been used to develop the waiting time as a function of service time tables presented in Part II – Table 3.9 and Table 3.10 and Table 2.6 in this chapter. Since generating these tables involves large amounts of simulations, OpenQTSim includes a multi-thread engine which allows the calculations to be run on multiple cores in parallel. The setup with SimPy furthermore enables the integration of additional restrictions that are normally not part of queueing simulators, such as limitations by available quay lengths, tidal windows and fairway restrictions.

OpenTNSim – Transport Network Simulation

Many studies of waterway systems aim to quantify traffic flows over the network, and the effects of interventions in the system. For safety studies detailed information about manoeuvring and ship-ship interactions may be needed, but for most capacity-related studies, as well as environmental effect studies, meso-scale traffic flows over the network are sufficient.

Apart from the national landscape of simulation models, numerous commercial packages are available to study the behaviour of agents on a network. Because for research purposes, policy-support and commercial software do not always provide the most flexible environment to test new algorithms, Van Koningsveld and Den Uijl (2020b) developed the OpenTNSim package.

OpenTNSim is a Python package for the investigation of meso-scopic traffic behaviour on networks to compare the consequences of different traffic scenarios and network configurations. The meso-scopic level (see also Figure 2.25) is of particular interest to problems that simultaneously require a large study area *and* more detailed engineering models to quantify specific aspects of the network or of the agents using it. Examples are the effects of network restrictions on traffic flows (e.g. due to maintenance, incidents, changing water levels), or of the assessment of the (environmental) cost and benefits of policy measures.

OpenTNSim uses the NetworkX package for transport graph construction and route selection. To investigate traffic flows and queueing patterns the discrete event simulation package SimPy is used. The Geospatial Data Abstraction Library (GDAL) is used to handle geospatial projections and operations. OpenTNSim contains a library that facilitates flexible specification of agents and various tools that facilitate the analysis of network behaviour.

Since OpenTNSim is built on widely used packages and libraries, it is quite robust and powerful. To facilitate use an automatic coupling is made to the Dutch Fairways Information Services (FIS), which is made available via http://www.waarweginformatie.nl. Fairway properties are automatically added to a graph, which subsequently can be used for routing questions and simulations. Interfaces are furthermore available to couple traffic simulations with the output of hydrodynamic models such as SOBEK (1D) or Delft3D (2DH).

Van der Does de Willebois (2019) used OpenTNSim to investigate the potential impact of quaywall maintenance on traffic flows over the canals of the city of Amsterdam. For this he characterised and implemented the behaviour of three types of agents that showed distinctly different behaviour. Groen (2020) expanded on this work looking at the impact of lock maintenance on the larger IWT network. For this he generated origin-destination information from AIS data. Vehmeijer (2019) used OpenTNSim to estimate CO_2 emission footprints of IWT traffic on on the Rotterdam – Antwerp corridor. For this she implemented simplified resistance estimation algorithms in combination with route selection algorithms to investigate the effect of policy measures and network modifications. Segers (2021) updated the resistance estimation algorithms based on new available literature, and added routines to estimate other emissions, such as SO_x, NO_x and fine dust. Kok (2021) expanded on this work by adding a fuel use estimator in order to develop a method to design suitable bunkering locations in case alternative energy carriers are used in IWT.

OpenCLSim – Complex Logistics Simulation

Apart from the performance of transport corridors, another category of interesting problems is the performance of (complex) waterborne supply chains. Where OpenTNSim focuses on network performance for a given traffic load, OpenCLSim aims to quantify the behaviour that follows from rule-based planning of (interdependent) cyclic activities, such as marine construction processes where one activity depends on the completion of another *and* each activity has its own sensitivity to currents and waves. To assess the complex behaviour of such interdependent supply chains we need to be able to simulate the behaviour of cyclic activities that are subject to logical rules. Examples are:

- the installation of an offshore wind park where first the monopiles have to be placed, before the wind turbine generators can be mounted,
- the construction of a submerged tunnel, where first a trench needs to be dug, before the prefabricated tunnel elements can be lowered into place,
- the realisation of a land reclamation, where first sand needs to be supplied until a sufficiently large area of land above water is created, before dry earth moving equipment can be mobilised to assist in the above water profiling,
- the execution of a dike reinforcement, where first base and filter layers need to be installed, before the toplayer of large rock can be installed,
- the supply of containers to a hinterland, where containers are first delivered to a sea port by very large ocean-going vessels, before a fleet of smaller inland container vessels can transport them to a series of inland ports.

The interdependence of activities in port and waterway settings generally translates into decisions on when to mobilise a vessel to avoid unnecessary (and typically expensive) waiting times, on how large certain storage facilities need to be to ensure that disruptions in one supply chain don't affect the others, et cetera. The interdependence of supply chains is already challenging, but it becomes even more complex and interesting when environmental and human restrictions play a role.

Van Koningsveld et al. (2020) developed OpenCLSim to study challenges like the ones mentioned above. Den Uijl (2018) investigated the potential of integrating engineering knowledge in logistical simulation of dredging projects. He showed in particular how application of suspended sediment spill restrictions would affect the overall project planning of dredging projects. Van der Bilt (2019) elaborated on this by implementing routines to estimate the power required during dredging operations, and subsequently the associated emissions. By analysing the execution of the same project, while using alternative execution methods, it becomes possible to compare the impact of alternative work methods. Kievits (2019) and Vinke et al. (2019, 2021) showed how the occurance of low discharge extremes affects not only the performance of key supply chains, but also impacts the overall fleet utilisation and ultimately the overall transport capacity of the system. They used the 1D hydrodynamics model SOBEK to translate discharge variations to water depths along the river. By subsequently translating available waterdepth into loading capacities, the impact of reduced water levels on the overall supply chain can be simulated. Van Halem (2019) expanded on this combination of OpenCLSim with a 1D flow model, by looking at the impact of time-varying 2DH flowfields. He showed that in open water with larger-scale currents, like in estuaries and coastal seas, the shortest route is not always the fastest. He furthermore demonstrated that in case of larger tidal motions and a complex bathymetry, it may even be beneficial to take on less cargo, when the reduced draught opens up a significantly shorter transport route.

OpenTISim – Terminal Investment Simulation

Where OpenTNSim and OpenCLSim are focused on waterways and logistics, the OpenTISim package (Van Koningsveld, 2020a) focuses on terminal development. In Part II – Chapter 3, 4 and 5 the design of terminals as a function of the desired annual throughput is discussed. Especially in functional designs, numerous trade-offs can and need to be investigated. The python package OpenTISim incorporates a range of design rules and investment triggers, for dry bulk, liquid bulk and container terminals, in order to investigate the impact of certain design choices on the overall business case of the terminal. Based on an anticipated development in demand, the model parametrically triggers additional investments in berths, quays, cranes, storage and hinterland facilities. By keep-

ing track of the costs associated with each added terminal element, the overall cashflows in terms of CAPital EXpenditures (CAPEX), OPerational EXpenditures (OPEX) and revenues can be derived.

Ijzermans (2019) applied OpenTISim to agribulk terminals. Where typically the available design guidelines provide suggested service levels, in terms of a maximum allowable waiting time as a factor service time, he explored the effect of 'playing' with this factor on the overall businesscase of the terminal in terms of NPV. It turned out that longer waiting times were often defensible, at least in terms of NPV of the overall project. Lanphen (2019) developed a liquid bulk module for OpenTISim to investigate the aspects involved in developing an import terminal for hydrogen. First, an analysis was made as to what hydrogen carrier would be preferential as a function of transport distance. Second, a functional design of hydrogen import terminals for four different energy carriers was made to illustrate which terminal elements were most critical in terms of the overall business case. Koster (2019) developed a container terminal module to analyse the effect of different yard equipment on the overall land requirements.

Stam (2020) combined OpenTISim with OpenCLSim to investigate trade-offs in port systems; i.e. the combination of an onshore port with an offshore port, connected by a barge link or a causeway. Through this combination it could be investigated, for example, how wave- related downtime of the barge link would affect the storage requirements in the offshore port, given the effect that such downtimes would have on dwell times. Abrahamse (2021) used a similar combination of OpenTISim and OpenCLSim to investigate the effects of alternative supply chain configurations on the cost of hydrogen at the end-use location.

In the next chapters we briefly illustrate typical port and waterway performance challenges as they may be encountered when looking at ports and terminals (Chapter 3), waterways and port water areas (Chapter 4) and port and waterway systems (Chapter 5).