

This chapter begins, in Sect. 2.1, with a brief overview of the use of models in the transportation sector, several types of models used in transportation planning, and the specific evaluation methods used. Thereafter, the theory of traffic flow is introduced which enables investigation of the dynamic properties of traffic on road sections with regard to the respective variables defined at each point in space and time. Based on the mathematical equations derived, different transportation system scenarios are investigated in Sect. 2.2. Section 2.3 examines queuing theory, the mathematical study of waiting in lines or queues. Transportation system models incorporate queuing theory to predict, for example, queuing lengths and waiting times. Section 2.4 analyzes transportation systems with regard to existing demand and the potential impact of changes resulting from transportation planning and development projects. Traffic management has become a critical issue as the number of vehicles in metropolitan areas is nearing the existing road capacity, resulting in traffic congestion. In some areas, the volume of vehicles has met and/or exceeded road capacity. The methodological background of congestion is described in Sect. 2.5. Graph theory is introduced in Sect. 2.6. It is widely used to model and study transportation networks. Section 2.7 focuses on shortages occurring in transportation systems, so-called bottlenecks. The main consequence of a bottleneck is an immediate reduction in the capacity of the transportation system infrastructure. Section 2.8 describes a ProModel-based case study for a four-arm road intersection. Section 2.9 contains comprehensive questions from the transportation model area of concentration, and the final section includes references and suggestions for further reading.

2.1 Introduction

A model can be introduced as a schematic description of a real-world system, theory, or phenomenon that accounts for its known or inferred properties used for further study of its characteristics or to predict or evaluate the intrinsic

dynamic behavior. In the transportation system sector, models or systems of models (so-called models of models (MOM)) are used to simulate traffic performance and traffic flow. Incorporating traffic requirements, as defined by technical and/or organizational constraints of real-world transportation systems, these models are used to predict impacts and/or to evaluate possible options for transportation planning and evaluation.

Transportation planning is defined as the process of making decisions about transportation resource needs, preferences, and values. Planning occurs at many different levels from day-to-day decisions to more general major decisions to strategic decisions with long-term impacts. Best practices in transportation planning can be achieved by coordinating short-term decisions in support of strategic long-term goals. An example of such comprehensive planning is transportation infrastructure planning with regard to land use, economic development, and social planning. Another example would be when manifold potential options exist to reduce traffic congestion, and some of these solutions may also help to overcome other traffic problems, such as finding parking spots and minimizing pollution emissions. A comprehensive transportation planning process will result in the prioritization of transportation activities and the efficient allocation of resources.

In the evaluation of transportation planning activities, the evaluation itself can be used as a method of determining the value of a potential planning option in order to support decision making. That is why evaluation in transportation planning is often applied when it comes to decision making (Small 1998; Litman 2006; USDOT 2003; CUTR 2007), and there are specific evaluation methods used, such as:

- *Cost-effectiveness (CE)*: This method compares the costs of different potential options for achieving a specific objective, such as building a particular highway or delivering a particular amount of airfreight, etc. The quantity of benefits (outputs) are held constant, so there is only one variable, the cost of inputs.
- *Cost-benefit analysis (CBA)*: This method compares the total incremental benefits with the total incremental costs for each of the potential options. This analysis is not limited to a single benefit or objective, such as potential highway routes which can differ in construction costs as well as quality of the services offered.
- *Lifecycle cost analysis (LCA)*: This method incorporates, in addition to CBA, the value of investments at the respective schedule, which allows a comparison of projects with regard to their cost and benefit milestones.
- *Multiple accounts evaluation (MAE)*: This method considers quantitative and qualitative evaluation criteria and can be used in cases where some impacts cannot be financially benchmarked. Using this evaluation method each potential option is rated for each potential criterion.

In general, transportation planning is involved with the evaluation, assessment, design, and siting of transportation facilities and is based on specific transportation planning models for which the respective environmental goals and objectives are defined. Problems which call for transportation planning or are identified during the

implementation of a solution require potential alternatives for development which have to be evaluated with regard to the existing budget. In this sense, the role of transportation planning is shifting from a purely technical analysis, including environmental aspects and sustainability, to a more integrated transportation framework which also embeds behavioral psychological aspects, e.g., persuading automobile drivers to use public transportation rather than their personal automobiles.

Several of the models used in transportation planning are the so-called travel demand models (TDMs) which have been developed to evaluate transportation demands in terms of the numbers of traveling individuals who may search for specific prices, transport services, modes, etc., in order to predict the corresponding traffic volumes and their potential impacts, such as congestion, pollution emissions, etc. Most TDMs are four-step models which follow these steps (TDM Encyclopedia 2013):

1. *Trip generation*: this approach predicts the total trips that start and end in a particular area of interest, the traffic analysis zone (TAZ), based on factors such as the zone's land use patterns; number of residents and jobs; demographic factors; transportation system features, such as number of roads, quality of transit service, etc.; and the distance between two zones.
2. *Trip distribution*: this approach focuses on trips that are distributed between pairs of zones, based on the distance between them.
3. *Mode split*: this approach focuses on trips that are allocated among the available travel modes.
4. *Route assignment*: this approach focuses on trips that are assigned to specific facilities included in the highway and transit transportation networks.

These models make use of travel surveys and census data to determine transportation demands, establish baseline conditions, and identify future trends. The trips used as a basis in these models are often predicted separately by purpose, i.e., work, shopping, etc., and thereafter aggregated into total trips on the respective network. This modeling approach allows the prediction of congestion problems because they mainly focus on measures of peak-period motor vehicle trips on major roadways. As a result of these predictions, a so-called level-of-service (LOS) roadway report is available with a letter grade from A (best) to F (worst) which indicates vehicle traffic speed and delay. As mentioned in TRB (2007), these models often incorporate several types of bias favoring automobile transport over other modes and undervaluing travel demand model (TDM) strategies. Because the travel surveys they are based on tend to ignore or undercount nonmotorized travel, they undervalue nonmotorized transportation improvements for achieving transportation planning objectives (Stopher and Greaves 2007). Moreover, they do not accurately account for the tendency of traffic to maintain equilibrium and the effects of traffic generated by roadway capacity expansion, thereby exaggerating future congestion problems and the benefits of roadway capacity expansion.

A number of recent studies have examined ways to better predict how smart growth locations and demand management programs can affect trip and parking

generation (Lee et al. 2012). Based on the assumption that a standard application of trip rates for an area with many smart growth characteristics will result in an overestimation of the number of trips generated, this study identifies eight available methodologies, five of which are candidate methods which are compared with the traditional trip generation method in a two-part assessment.

Economic models are used to evaluate and compare the value of particular transportation improvements, such as widening a roadway, improving public transit, or implementing a TDM strategy. The models compare the various categories of benefits and costs. They tend to consider a relatively limited set of benefits, since most of these models were originally developed to evaluate roadway improvement options. They generally assume that total vehicle mileage is constant and so is not well designed to evaluate the full benefits of TDM strategies that reduce automobile trips. For example, these models often ignore parking and vehicle ownership cost savings that result when travelers shift from automobile travel to alternative modes; and they generally ignore the safety benefits that result from reductions in total vehicle mileage (Ellis et al. 2012).

Integrated Transportation and Land Use Models are designed to predict how transportation improvements will affect land use patterns, e.g., the location and type of development that will occur if a highway or transit service is improved. They are often integrated with traffic models. These are considered the best tools for evaluating transportation policies and programs because they can measure accessibility rather than just mobility, but they are costly to develop, are complex, and may be difficult to apply, particularly for evaluating individual, small-scale projects (Dong et al. 2006). Some models predict how particular land use factors, such as density and mix, affect travel behavior and their impacts on congestion and pollution emissions (Donoso et al. 2006; Scheurer et al. 2009; Bartholomew and Ewing 2009). The Smart Growth Area Planning (SmartGAP) tool synthesizes households and firms in a region and determines their travel demand characteristics based on their built environment and transportation policies affecting their travel behavior (TRB 2012).

Transportation simulation models are a newer approach to modeling the behavior and needs of individual transport users (so-called agents), rather than aggregate groups. This improves the consideration of modes such as walking and cycling; the transport demands of nondrivers, cyclists, and the disabled; and the effects of factors such as parking supply and price, transit service quality, and local land use. Simulation models can provide a bridge between other types of models, since they can incorporate elements from the conventional traffic, economic, and land use models. Simulation models have been used for many years in individual projects and are increasingly used for area-wide analysis. Transportation simulation models allow traffic flow and network flow aspects to be combined for investigation of transportation systems with continuous services, such as road systems and transportation systems with discrete services, such as airplanes, buses, ships, and trains.

To conclude, the biases in current models tend to exaggerate the benefits of roadway capacity expansion and understate the value of alternative modes and TDM solutions. More accurate and comprehensive modeling is, therefore, a key

step in developing more optimal transport planning and the implementation of specific TDM strategies. Therefore, in TDM Encyclopedia (2013), the various problems common with current models and how they can be corrected are described. These deficiencies are not necessarily intrinsic; significant improvements can be made to existing models and how they are applied. For example, many problems could be reduced by simply educating planners and decisions makers about modeling assumptions, biases, and weaknesses so that they can take these factors into account.

2.2 Traffic Flow Models

The theory of traffic flow investigates the dynamic properties of traffic on road sections. Dynamic models of traffic flow date from the 1950s, representing traffic flow based on an analogy with lines of water flows in rivers, an approach that allows to treat individual vehicles as “continuous fluid.” Against this background, macroscopic traffic flow theory relates on variables declaring the dynamic properties of traffic which are:

- Density k
- Flow rate q
- Speed v

These result in the fundamental statement that flow q equals density k multiplied by speed v . These variables are defined at each point in space and time which means that the discrete nature of traffic is transferred into continuous variables. The evolution in time of these state variables can be modeled by partial differential equations (PDEs) comprising the conservation of mass (vehicles) and an experimental relation between flow rate q and density k . Using this approach, traffic flow models can be formulated for density k by the number of vehicles n at time t_0 occupying a given length x of a road or, more in general, on the location interval Δx of a roadway at a particular instant, as follows:

$$k = \frac{n}{\Delta x}. \quad (2.1)$$

The total space s of the n vehicles can be set equal to Δx , and thus we can write

$$k = \frac{n}{\sum_i s_i} = \frac{1}{\bar{s}}, \quad (2.2)$$

where the mean space occupancy in the interval s_i is defined as

$$\bar{s} = \frac{1}{n} \sum_n s_i. \quad (2.3)$$

From (2.3), it can be seen that density k depends on the designated roadway point x_0 , the time t_0 , and the measurement interval, defined as an area in the t - x space. As introduced in Immers and Logghe (2002), for a location x_1 , we can take the center of the measurement interval Δx . Thus, (2.1) can be rewritten in order to include these factors:

$$k(x_1, t_1, s_1) = \frac{n}{\Delta x}. \quad (2.4)$$

2.2.1 Uncongested Traffic Conditions

For uncongested traffic conditions, freeway traffic data suggests that desired speeds are relatively constant and chosen by the drivers. Under stationary conditions, the flow-rate-versus-density ratio can be expressed as mean speed v , which appears to be nearly constant for uncongested traffic flow. Introducing the flow-rate-versus-density ratio under congested conditions causes driver behavior to become an important factor. Assuming drivers can no longer choose free-flow speed under congestion, a simple classification can define driver types: aggressive drivers T_{AD} and nonaggressive drivers T_{NAD} . Assuming that each driver type drives at his/her desired speed, the uncongested flow-rate-versus-density relationship is a weighted average of the desired speeds. With regard to such behavior, a regression of traffic flow on total density interacts with proportions of distinct driver/vehicle types T_I with $I = AD$ or $I = NAD$. This results in estimates of free-flow speeds for these drivers, described for the uncongested flow rate by the following equation (Kockelman 2001):

$$q_u = \sum_{T_I} v_{free, T_I} p_{T_I} k, \quad (2.5)$$

where q_u is the total uncongested traffic flow rate, v_{free, T_I} is the free traffic flow speed of driver/vehicle type T_I , $p_{T_I} k$ is the density of driver/vehicle type T_I , and p_{T_I} is the proportion of vehicles on the road of driver/vehicles type T_I .

2.2.2 Congested Traffic Conditions

In the case of a congested condition, the driving situation is different because speed is no longer constant for tumescent densities. Drivers can no longer choose free-flow speeds because they have to be aware of the spacing at which they follow the car in front of them. For this situation, the behavioral assumption is of selected spacing d as a linear function of congested speed v_C . Since total vehicle density k is the inverse of average spacing of vehicles on the roadway and average spacing is a

proportion-weighted sum of type densities, one can solve for the total density k_T as a function of speed as shown in the following equation from Kockelman (2001):

$$k_T = \frac{1}{\sum_{T_i} p_{T_i} d_{T_i}} = \frac{1}{\sum_{T_i} p_{T_i} (a_{T_i} + b_{T_i} v)}, \quad (2.6)$$

where d_{T_i} stands for intervehicle spacing (front-to-front) of the i th driver type, v is mean speed, and a_{T_i} and b_{T_i} are constants defining the i th driver type behavior.

2.2.3 Flow-Density and Speed-Flow Graphs

Based on the foregoing specifications and definitions, the following graphs can be introduced (Muench 2004), showing the congested and uncongested flow rate q versus density k (Fig. 2.1) and the speed v versus flow rate q (Fig. 2.2). As indicated in Fig. 2.1, the optimal traffic flow capacity q_m correlates with the inflection point k_m at which the uncongested flow rate changes into the congested flow rate, meaning the more density k increases, the more traffic flow q decreases, which can be expressed by the equation of flow rate q with v_f as free space mean speed shown in Fig. 2.1.

In Fig. 2.2, it is shown that the optimal traffic flow capacity q_m correlates with the inflection point v_m at which the uncongested free-flow speed changes into the congested flow speed. In other words, the more the flow rate q increases, the more the mean speed v decreases, which results in the equation of flow rate q with v_f as free space mean speed shown in Fig. 2.2.

Flow rate q is the interaction of density k and mean speed (stationary traffic conditions) u . Thus, flow rate q represents the number of vehicles n passing some

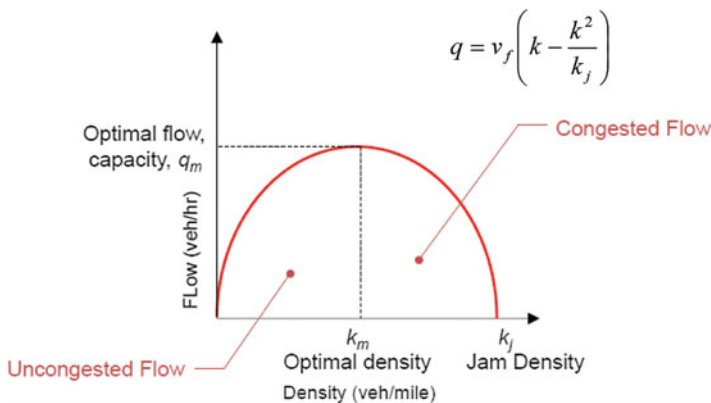


Fig. 2.1 Flow rate versus density graph (Muench 2004)

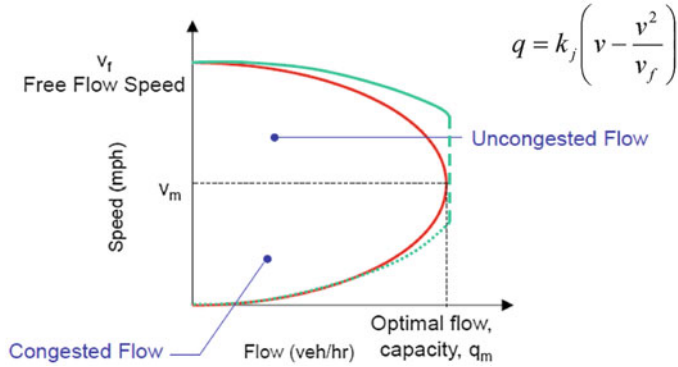


Fig. 2.2 Speed versus flow rate graph (Muench 2004)

designated roadway point x_0 in a given time interval Δt . For time interval Δt at any location x , such as measurement interval S , flow rate q be calculated as follows:

$$q(x, t, S) = \frac{n}{\Delta t}. \quad (2.7)$$

The time interval Δt is the sum of headways h between vehicles as their bumpers pass a given point x_0 :

$$\Delta t = \sum_{i=1}^n h_i. \quad (2.8)$$

Introducing a mean headway \bar{h} , we find the following expression for the traffic flow rate q :

$$q_u = \frac{n}{\sum_{i=1}^n h_i} = \frac{1}{\bar{h}}. \quad (2.9)$$

Mean speed v is the quotient of flow rate q and density k . Mean speed is a function of location x , time interval Δt , and measurement interval S which results in:

$$v(x, t, S) = \frac{q(x, t, S)}{k(x, t, S)}. \quad (2.10)$$

In another form, this definition of mean speed is also called the fundamental relation of traffic flow theory:

$$q = k \cdot v, \quad (2.11)$$

This relation links flow rate q , density k , and mean speed v . Knowing two of these variables immediately leads to the remaining third variable.

2.2.4 Traffic Flow Scenarios

Based on the mathematical equations above, we can work out some traffic scenarios as case study examples.

Scenario 1

Problem: Let us assume that a vehicle is traveling in uncongested conditions for a total distance D of 100 miles. For the first 60 miles of this distance D_1 , the vehicle travels $v_1 = 55$ mph; and for the next 40 miles D_2 of the total distance, it travels $v_2 = 65$ mph. For this scenario, the weighted average speed over the time spent traveling those 100 miles is of interest to know.

Solution: Intuitively, driving at 55 mph will take longer than driving at 65 mph. Hence, the weighted average speed v_{wa} for the entire trip is less than the arithmetic mean speed v_{am} of 60 mph. Thus, we will demonstrate that this is true by the following calculations:

- 60 miles at 55 mph $= t_1 = 65.45$ min
- 40 miles at 65 mph $= t_2 = 36.92$ min

Now we can calculate the weighted average speed:

$$v_{wa} = \frac{v_1 \cdot t_1 + v_2 \cdot t_2}{t_1 + t_2} = \frac{(55 \text{ mph} \cdot 65.45 \text{ min} + 65 \text{ mph} \cdot 36.92 \text{ min})}{65.45 \text{ min} + 36.92 \text{ min}} = 58.6 \text{ mph.}$$

Scenario 2

Problem: Let us assume that five vehicles with different driver types T_i are driving in uncongested conditions over a given distance D of 100 miles. For each vehicle, this distance requires a different time due to the different speeds chosen by the different driver types T_i as shown in Table 2.1.

Solution: To calculate the average speed v_a , we first have to calculate the average travel time t_a as follows:

$$t_a = \frac{t_1 + t_2 + t_3 + t_4 + t_5}{5 \text{ veh}} = \frac{533 \text{ min}}{5 \text{ veh}} = 106.6 \text{ min.}$$

Table 2.1 Driving time for a given distance by different driver types

Vehicle	Time required to drive 100 miles
1	$t_1 = 80$ min
2	$t_2 = 100$ min
3	$t_3 = 133$ min
4	$t_4 = 120$ min
5	$t_5 = 100$ min

Now we can calculate the average speed v_a based on the average travel time t_a

$$v_a = \frac{D}{t_a} = \frac{100 \text{ miles} \cdot 60 \text{ min}}{106.6 \text{ min} \cdot h} = \frac{6,000}{106.6} = 56.28 \text{ mph}.$$

Scenario 3

Problem: Let us assume that 20 vehicles pass a given point x_0 in 1 min and move a length of 1 mile. For this scenario flow rate q , density k , space mean speed, space headway h_s , and time headway h_t are of interest to know.

Solution: In general, the time (in seconds) between moving vehicles, as their front bumpers pass a given point x_0 , can be calculated based on (2.8) as follows:

$$t = \sum_{i=1}^n h_i.$$

And, therefore, flow rate can be calculated based on (2.9) as follows:

$$q = \frac{n}{\sum_{i=1}^n h_i} = \frac{20 \text{ veh} \cdot 60 \text{ min}}{1 \text{ min} \cdot h} = 1,200 \frac{\text{veh}}{h},$$

which results for density k in

$$k = \frac{v}{\Delta x} = \frac{20 \text{ veh}}{1 \text{ mile}} = 20 \frac{\text{veh}}{\text{mile}}.$$

Calculating space mean speed has to take into account that space mean speed of vehicles moving along and traversing a roadway segment of a known length l follows (2.11) assuming v is space mean speed:

$$v = \frac{q}{k} = \frac{1,200 \frac{\text{veh}}{h}}{20 \frac{\text{veh}}{h}} = 60 \frac{\text{mile}}{h}.$$

Now we can calculate the space headway h_s and the time headway h_t . Space headway h_s can be calculated in an idealized manner taking into account the result for density k as follows:

$$k = \frac{1}{\bar{h}_i}$$

$$\bar{h}_i = \frac{1}{k} = \frac{1}{40 \frac{\text{veh}}{\text{mile}}} = 0.025 \text{ mile}.$$

Time headway h_T can be calculated in an idealized manner taking into account the result for space headway h_S as follows:

$$\begin{aligned}\bar{h}_S &= v \cdot \bar{h}_T, \\ \bar{h}_T &= \frac{\bar{h}_S}{v} = \frac{0.025 \text{ mile}}{60 \frac{\text{mile}}{\text{h}}} = 1.5 \text{ s}.\end{aligned}$$

In general, traffic measurements are executed at a fixed location x_f which allows an easy measure of occupancy o . As introduced in Immers and Logghe (2002), the relative occupancy o_R of a vehicle in measurement interval S and time interval Δt can be calculated as follows:

$$o_R(x, t, S) = \frac{1}{\Delta t} n \sum o. \quad (2.12)$$

Assuming all vehicles have the same length l_V , then the relative occupancy o_R and density k can be given as follows (Immers and Logghe 2002):

$$o_R(x, t, S) = l_V k(x, t, S). \quad (2.13)$$

Scenario 4

Problem: Let a traffic stream have a mean speed v of 50 mph and a flow rate q of 1,000 vehicles/h. All vehicles are assumed to be 5 m in length l_V . What is the relative occupancy?

Solution: From (2.11) we receive

$$k = \frac{q}{v} = \frac{1,000 \frac{\text{vehicle}}{\text{h}}}{50 \text{ mph}} = 20 \frac{\text{vehicles}}{\text{mile}}.$$

Given that density k is 20 vehicles/mile means that k corresponds with space occupancy o_S of 80,465 m per vehicle. With an assumed vehicle length l_V of 5 m, the corresponding relative occupancy o_R is 6.21 %. Calculating the relative occupancy o_R by using (2.13) gives

$$o_R(x, t, S) = l_V k = 0.005 \text{ km} \cdot 0.62137 \frac{\text{mile}}{\text{km}} \cdot 20 \frac{\text{vehicles}}{\text{mile}} = 6.21 \%.$$

It should be noted that this formula cannot be used in real-world applications because a traffic stream is neither homogeneous nor stationary in reality. A possible solution calculating traffic density is to measure the traffic flow rate and traffic mean speed using the equations given in Immers and Logghe (2002) and then calculate traffic density by using the fundamental relation of traffic flow theory in (2.11).

2.2.5 Traffic Flow Behavior

With regard to the level of detail the models use to represent the traffic flow behavior, the models can be classified, as introduced in Chap. 1: macroscale as macroscopic traffic flow models, representing the traffic behavior at an aggregated level; microscale as microscopic traffic flow models, representing the movement of individual vehicles; and mesoscale models representing traffic flow at the level of detail of a single vehicle.

Macroscopic traffic flow models can be described as:

- *Space continuous models*: where state variables are defined at each point in space
- *Space discrete models*: where basic variables affecting link performance, such as density or speed, do not vary along the link

With regard to the control flow in the macroscale model, there is no consideration of detailed individual transportation units. Using the fundamental relation of traffic flow theory in (2.11) to describe the changes in time and location of the macroscopic variables along a road:

$$q(x, t) = k(x, t) \cdot v(x, t). \quad (2.14)$$

Let the road to be modeled be divided into cells with length Δx , and the density of cell i at time t_j is represented by $k(i, j)$; then the number of vehicles in cell i is $k(i, j) \cdot \Delta x$. Then, one time interval Δt later, at $t_j + 1$, density will change. Let us assume that a number n of vehicles have traveled from cell $i-1$ into cell i which results in a traffic inflow of

$$q(i-1, j) \cdot \Delta t, \quad (2.15)$$

and a number n of vehicles have traveled from cell i to cell $i+1$ which results in the traffic outflow

$$q(i, j) \cdot \Delta t. \quad (2.16)$$

Let us also assume there are branching and exit roads which will enable in- and outflows in the form of

$$z(i, j) \cdot \Delta x \cdot \Delta t, \quad (2.17)$$

where z is expressed per time and length of unit and is positive for an increase in the number n of vehicles.

Let the limit for time step Δt and cell length Δx approach zero, and we can write the partial differential equation representing the conservation law of traffic as follows:

$$z(x, t) = \frac{\partial k(x, t)}{\partial t} + \frac{\partial q(x, t)}{\partial x}. \quad (2.18)$$

Let the traffic flow be stochastic. This requires a stochastic model because the variables of the traffic flow cannot be described as a deterministic process but could be described as a stochastic process:

- Sequence of vehicle arrivals (arrival pattern)
- Sequence of service times at maintenance check of vehicles (service pattern)
- Queuing behavior

Let arrivals and services be independent, randomly distributed variables with time constant parameters. Let N be a random variable describing the queue length and n realizations of N . Let the queuing phenomena be defined by the following notation:

$$A/B/c(d, e)$$

where A denotes the type of arrival pattern variable describing time intervals between two successive arrivals, B denotes the type of service pattern, c is the number of service stations, d is the queue storage limit, and e denotes the queuing behavior, such as *FIFO* (First In First Out), *LIFO* (Last In First Out), etc., where d and e , if defined by ∞ (no constraint on maximum queue length) and by *FIFO*, are generally omitted. This representation in traffic flow models allows, with the help of queuing analysis, to determine how long it takes to complete a trip and/or how long it would have taken if there was no queuing, congestion, etc., which refers to the topic of Sect. 2.3, Queuing Models.

2.3 Queuing Models

Queuing theory is the mathematical study of waiting lines or queues. In queuing theory, a model is constructed so that queue lengths and waiting times can be predicted (Sundarapandian 2009). Queuing theory is generally considered a branch of operations research, a discipline which deals with the application of advanced analytical methods to help make better decisions, as the results are often used when making business decisions about the resources needed to provide a service. Thus, planning efficient transportation systems and networks is a crucial factor for urban insertion since it gives access to economic activity, facilitates family life, and more. Hence, the importance of transportation in human life and global economy cannot be overemphasized, and queuing theory can help to study traffic behavior near a certain section where demand exceeds available capacity. Queuing can be discovered in many common situations like boarding a bus or a train or a plane, freeway bottlenecks (see Sect. 2.7), etc. In transportation engineering, queuing can occur at red lights, stop signs, bottlenecks, or any traffic-based flow constriction. When not

dealt with properly, queues can result in severe network congestion or gridlock conditions, therefore making them important to be studied and understood by engineers. For example, based on the departure and arrival pair data, the delay of every individual vehicle can be determined. Using an input-output queuing diagram, it is possible to determine the delay for every individual vehicle: the delay of the i th vehicle is time of departure-time of arrival ($t_d - t_a$). Thus, the total delay is the sum of the delays of each vehicle.

Let us assume that the traffic flow q to and through an intersection is controlled by a traffic light sequencer. This can be accomplished by analyzing the cumulative flow of vehicles as a function of time. As it is known from traffic flow operation experience with traffic light sequencing, traffic lights change in the following sequence:

$$\text{green} \rightarrow \text{amber} \rightarrow \text{red} \rightarrow \text{amber} \rightarrow \text{green}$$

whenever, e.g., a person pushes a button. Let us assume the light is red and the traffic flow is stopped from time t_1 to t_2 during the red signal interval. At the start of the green interval (t_2), traffic begins to leave the intersection, with the so-called saturation traffic flow rate q_{sat} , and continues until the queue is exhausted. Thereafter, the departure rate $D(t)$ equals the arrival rate $A(t)$ until t_3 , which is the beginning of the next red signal. At this point, the process starts all over. The resulting type of traffic flow is called interrupted flow. Interrupted traffic flow is a flow regulated by an external means, such as a traffic signal. Under interrupted traffic flow conditions, vehicle-vehicle interactions and vehicle-roadway interactions play a secondary role in calculating the traffic flow. For interrupted traffic flow, the following impacts can be identified:

- Determining the optimal cycle length and phase length for traffic lights with regard to the daytime-dependent numbers of vehicles moving in different possible directions at the respective crossing
- Evaluating consequences, adding lanes, or changing the geometric configuration of an interstate highway on recurrent (peak period) and nonrecurrent (incident happens) delays
- Optimizing the frequency at which trucks should be dispatched along a route, taking cost of operation and service quality into account

Contrary to the interrupted traffic flow is the uninterrupted traffic flow which depends on vehicle-vehicle interactions and interactions between vehicles and the roadway situation. For example, vehicles traveling on an interstate highway are moving in an uninterrupted traffic flow so long as no congestion occurs as part of an accident on the interstate highway.

Thus, the dominant effect of queuing theory in transportation is the delay of a trip from an initial destination to a final destination, measured as

- Time in system
- Average speed
- Waiting time

Therefore, queuing analysis allows determining how long it will take to complete a trip assuming an uncongested situation and how long it would have taken if a queuing or congestion situation is assumed. For these cases, the performance measures, predicted with queuing models, are:

- Throughput rate at which vehicles proceed through the highway system, which means, in terms of transportation, how long a trip will take from the place of departure to the final destination.
- Crowding/congestion is the separation between or density of vehicles, which means, in terms of transportation, a specific number of cars and/or trucks are moving in their respective lanes on a highway from the place of departure to the final destination.
- Queue percentage refers to the number of vehicles that encounter a queue prior to traveling, which means, in terms of transportation, congestion happens where vehicles have to wait before they can drive on the highway from the place of congestion to the final destination.
- Transportation cost is the annual or per customer expense of providing transportation service, which means, in terms of transportation, the transportation ticket bill as part of public transportation from the place of departure to the final destination.
- Productivity of transportation depends on the amount of queuing and whether the transportation system is saturated, which means, in terms of transportation, a highway or roadway, which is congested every morning by vehicles commuting into a metropolitan area and leaving the same way in the evening to go back home, is saturated for peak traffic. The degree of saturation is:
 - Under saturated: $\lambda < \mu$.
 - Saturated: $\lambda = \mu$.
 - Oversaturated: $\lambda > \mu$.with the following notation:
 - Arrival rate (vehicles per unit time): λ .
 - Departure rate (vehicles per unit number): μ .

2.3.1 Little's Law

Let the average queue size (measured in vehicles) equal the arrival rate (vehicles per unit time) multiplied by the average waiting time (both delay time in queue and activity time (in units of time)); then the result is independent of particular arrival distributions, which is known as Little's Law (Little and Graves 2008). This law says that, under steady-state conditions, the average number of vehicles in a queuing system equals the average rate which vehicles arrive multiplied by the average time that a vehicle spends in the system. Letting

- L : average number of vehicles/customers in the queuing system
- W : average waiting time in the system for a vehicle/customer
- λ : average number of vehicles arriving per unit time

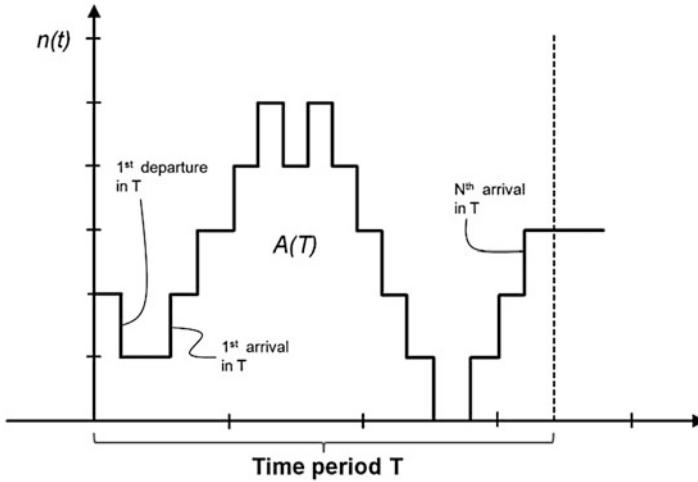


Fig. 2.3 Number of vehicles in a queuing system versus time (Little and Graves 2008)

the resulting law is called Little's Law:

$$L = \lambda W. \quad (2.19)$$

This equation is remarkably simple, extremely useful, and handy for back in the envelope calculations. The reason is that two of the terms in (2.19) may be easy to estimate but not the third. Thus, Little's Law provides the missing value (Little and Graves 2008).

In Fig. 2.3 we follow an example given in Little and Graves (2008) which shows a possible realization of a queuing system. With regard to Little's Law, one can make a heuristic argument interpreting the area under the curve in Fig. 2.3 in two different ways:

Let

- $n(t)$: number of vehicles in the queuing system at time t
- T : long period time
- $A(T)$: area under the curve $n(t)$ over the time period T
- $N(T)$: number of arrivals in the time period T

On the one hand, a vehicle in the queuing system is simply there. The number of items can be counted at any instant of time t to give $n(t)$. Its average value over T is the integral of $n(t)$ over T , meaning $A(T)$, divided by T . On the other hand, at time t , each of the vehicles is waiting and is accumulating waiting time. By integrating $n(t)$ over the time period T , we obtain a cumulative measure of the waiting time, again equal to $A(T)$. Furthermore, the arrivals are countable too and given by $N(T)$. Therefore, from Fig. 2.3, we can define

- $C = \frac{N(T)}{T}$: arrival rate during the time period T
- $L(T) = \frac{A(T)}{T}$: average queue length during time period T , indicating the number of customers in the system at time T
- $W(T) = \frac{A(T)}{N(T)}$: average waiting time in the system per arrival during T

A slight manipulation of Little's Law in (2.19) gives

$$L(T) = \lambda(T)W(T). \quad (2.20)$$

All of these quantities wiggle around a little as T increase because of the stochastic nature of the queuing process and because of end effects. End effects refer to the inclusion in $W(T)$ of some waiting by vehicles/customers which joined the system prior to the start of T and the exclusion of some waiting by vehicles/customers who arrived during T but have not left yet. As T increases, $L(T)$ and $\lambda(T)$ go up and down somewhat as vehicles/customers arrive and later leave.

Under appropriate mathematical assumptions about the stationarity of the underlying stochastic processes, the end effects at the start and finish of T become negligible compared to the main area under the curve. Thus, as T increases, these stochastic wiggles in $L(T)$, $\lambda(T)$, and $W(T)$ become smaller and smaller percentages of their eventual values so that $L(T)$, $\lambda(T)$, and $W(T)$ each go to a limit as T increase to infinity. Then, using the obvious symbols for the limits, we receive

$$\lim_{T \rightarrow \infty} L(T) = L; \quad \lim_{T \rightarrow \infty} \lambda(T) = \lambda; \quad \lim_{T \rightarrow \infty} W(T) = W$$

from which we get the desired result for (2.19). It is important to note the equation holds for each realization of the queuing system over time. This was argued by Little, in his original paper in 1961, noting that (2.19) held for each evolution of the time series of a particular queuing system (Little and Graves 2008).

2.3.2 Queuing Systems Attributes and Disciplines

Since the key elements of a transportation queuing system are vehicles and arrivals and/or departures, a queuing system can be described by the following attributes:

- Calling population, which represents the population of potential vehicles who may have called for an arrival and/or departure
- System capacity, which is the limit in numbers of vehicles that the queuing model can accommodate at any time
- Composition of arrivals and/or departures, which can occur at scheduled times or at random times

- Queuing discipline, which is the behavior of the queue in reaction to its current state
- Service mechanism, which means that service times may be constant or of some random duration.

Therefore, queuing models gain information about characteristic quantities that describe the workload of the transportation system or the time the activity needs to pass through the system. Against this background, the intention of using queuing models in transportation is to gain information about characteristic quantities that describe traffic flow, traffic density, etc., or the time a traffic flow needs to pass a distance, e.g., turnaround of an aircraft at an airport, across a flow interruption point, etc. The ways activities are processed through queues are based on specific queue disciplines which refer to the rule that a server uses to choose the next customer from the queue (if any) when the server completes the service of the current customer. Commonly used queue disciplines are:

- First come, first served (FCFS) or first in, first out (FIFO): means that customers (vehicles, passengers, etc.) are served one at a time and that the customer that has been waiting the longest is served first.
- Last come, first serve (LCFS) or last in, first out (LIFO): means it also serves customers (vehicles, passengers, etc.) one at a time; however, the customer with the shortest waiting time will be served first.
- Sharing: means activity capacity is shared equally between customers (vehicles, passengers, etc.).
- Priority: means customers (vehicles, passengers, etc.) with high priority are served first. Priority queues can be of two types: non-preemptive (activity in service cannot be interrupted) and preemptive (activity in service can be interrupted by a higher priority activity).
- Shortest activity first: means the next activity to be served is the one with the smallest size.
- Preemptive shortest activity first: means the next activity to be served is the one with the original smallest size.
- Shortest remaining processing time: means the next activity to be served is the one with the smallest remaining processing requirement.
- Round robin scheduling (RRS): means time slices are assigned to each activity in equal portions and in circular order, handling them all without priority; also known as cyclic executive.
- Multilevel feedback: means a scheduling algorithm which meets the following design requirements for multimode systems:
 - Gives preference to short activities.
 - Gives preference to I/O bound processes. This means it refers to a condition in which the time it takes to complete an activity is determined principally by the period spent waiting for input or output services to be completed.
 - Separates processes into categories based on their need for services.
- Service in random number: means random numbers are generated in a predictable fashion using a mathematical formula announcing the sequence of services.

Scenario 5

Problem: Let a transportation system have the following elements: a calling population, a waiting line, and services. Let calling population be infinite, i.e., if a vehicle leaves the calling population and joins the waiting line or enters service, there is no change in the arrival rate of other vehicles that may need service. Arrivals for service occur one at a time using a randomized schedule; once they join the waiting line, they are eventually served. In this transportation model, service times are assumed to be of some random length according to a probability distribution that does not change over time. Assume that system capacity has no limit, meaning that any number of vehicles can wait in line. Furthermore, the vehicles should be served in the order of their arrival by a single server, which results in the first come, first served (FCFS) or first in, first out (FIFO) service schedule.

Let arrivals and services be defined by the distribution of the time between arrivals and the distribution of the service times, respectively. For any simple transportation queue, the overall effective arrival time has to be less than the total service rate; otherwise, the waiting line will grow without bounds. If queues grow without bounds, they are called explosive or unstable. In cases where the arrival time will be for short terms greater than the service rate, there is a need for queuing networks with routing capabilities.

Queuing systems can be represented by terms such as state, event, simulation clock, etc. Hence, the state of the queuing system is represented by its number of vehicles as well as the state of the activity (server), which can be busy or idle. An event then represents a set of circumstances that causes an instantaneous change in the state of the system. There are only two possible events that can affect the state of the transportation system: the arrival event, which means the entry of a vehicle into the system, and the departure event, meaning the completion of an activity (service) on a vehicle. Furthermore, a simulation clock is used to track simulated time.

Solution: If a vehicle enters a discrete-event transportation system, the vehicle can find activity (server) either busy or idle, which results in two possible cases:

- 1. Vehicle begins with activity (service) immediately if the server is idle.
- 2. Vehicle enters queue for activity (server) immediately if server is busy.

It is not possible for the server to be idle and the queue to be empty, which can be interpreted as a third case. The results of which can be expressed in a matrix form for the potential unit actions upon arrival, as shown in Table 2.2.

After completing a service, as shown in Table 2.2, the server can become idle or remain busy with the next unit. The relationship of these two outcomes of the state of the queue is shown in Table 2.3. If the queue is not empty, another unit can enter

Table 2.2 Cases of unit actions upon arrival (for details see text)

Server status	Queue status	
	Not empty	Empty
Busy	2	2
Idle	3	1

Table 2.3 Server outcomes of Table 2.2 after service completion (for details, see text)

Server status	Queue status	
	Not empty	Empty
Busy	1 or 2	3
Idle	3	1 or 2

the server keeping him busy; or if the queue is empty, the server will be idle after a service is completed, which is indicated by the disjunctive indication of case 1 or 2. Again, it is impossible for the server to become busy if the queue is empty when a service is completed, which is indicated by case 3.

Simulating queuing systems requires the stipulation of an event list for determining what will be next. This event list tracks the future times at which different types of events occur. Hence the simulation system is able to calculate the respective simulation clock time, e.g., for arrivals and departures. If events occur at random times, the randomness needed can be realized through random numbers. A random number is a number generated by a process, whose outcome is unpredictable and which cannot be subsequentially reliably reproduced. This definition works fine provided that one has some kind of a black box that fulfills this task. Random numbers have the following properties:

- The set of random numbers is uniformly distributed between 0 and 1.
- Successive random numbers are independent.

When used without specific meaning, the word random usually means random with uniform distribution. A uniform distribution also known as a rectangular distribution is a distribution that has constant probability. A transformation which transforms from a two-dimensional continuous uniform distribution to a two-dimensional bivariate normal distribution or complex normal distribution is the Box-Muller transformation which allows pairs of uniform random numbers to be transformed to corresponding random numbers having a two-dimensional normal distribution. Random numbers can be generated with the respective queuing system simulation tools. When generating random numbers over some specified boundary, it is often necessary to normalize the distributions so that each differential area is equally populated.

Scenario 6

Problem: Let the transportation system in Scenario 5 have interarrival times and service times that can be generated from the distribution of random variables. Consider having seven vehicles with the interarrival times 0, 2, 6, 4, 3, 1, 2. Based on the interarrival times, the arrival times of the seven vehicles in the queuing systems result in 0, 2, 8, 12, 15, 16, 18.

Solution: Due to these boundaries, the first vehicle arrives at clock time 0, which sets the simulation clock in operation. The second vehicle arrives two time units later at clock time 2, the third vehicle arrives six time units later at clock time 8, etc. The second time values of interest in Scenario 5 are activity (service) times that are

generated at random from a distribution of activity (service) times. Let the possible activity (service) times be one, two, three, and four time units. Hence, we are able to mesh the interarrival times and the activity (service) times, simulating the simple transportation queuing system.

In this queuing model, the first vehicle arrives at clock time 0 and activity (service) starts immediately, which requires four time units. The second vehicle arrived at clock time 2, but activity (service) could not begin until clock time 4. This occurred because vehicle 1 did not finish activity (service) until clock time 4. The third vehicle arrives at clock time 8 and is finished at clock time 10, etc. The strategy that serves vehicles in Scenario 6 is based on the *first come, first served (FCFS)* or *first in, first out (FIFO)* basis, which keeps track of the clock time at which each event occurs.

Furthermore, the chronological ordering of events can be determined as records of the clock times of each arrival event and of each departure event, depending on the vehicle number. The chronological ordering of events is needed as a base concept for the realization of discrete-event simulation systems.

2.3.3 Queuing Systems Parameters and Performance Measures

Further interesting parameters for queuing systems are the:

- Workload, which represents the percentage of the simulation time a resource was working
- Throughput, which is the number of vehicles per time unit that leave the system
- Mean waiting time
- Mean time in system
- Queue length
- Mean number of waiting vehicles

Moreover, knowledge of the layout of the queuing networks is of importance for the use of discrete-event simulation systems. The layout depends on:

- Open-queuing systems, which have sources and sinks. The jobs pass through the queuing net and leave it when all demands are satisfied. Typical examples of open-queuing systems are production lines, where the jobs are the raw materials that have to be processed using certain operations and leave the system as ready-made products.
- Closed-queuing systems, which are identified by a closed loop in which the jobs move through the queuing net. The number of jobs is fixed for the whole simulation time. A typical example of a closed-queuing system is a multiuser system with n terminals and a single central processing unit (CPU). The jobs circle between the terminals and the CPU; their number stays constant during the simulation time.

Simulating queuing systems generally requires maintaining data specifying the dynamic behavior of the discrete-event system, which can be done using simulation tables designed for the problem being investigated. Hence, the content of the simulation table depends on the system and can give answers such as:

- The average waiting time of a vehicle is determined by the total time the vehicles wait in the queue divided by the total number of vehicles.
- The average time a vehicle spends in the queuing system is determined by the total time the vehicles spend in the queuing system divided by the total number of vehicles.
- The average service time is determined by the total service time divided by the total number of customers.
- The average time between arrivals is determined by the sum of all times between arrivals divided by the number of arrivals $- 1$.
- The probability a vehicle has to wait in the queue is determined by the number of vehicles who wait in queue divided by the total number of vehicles.
- The fraction of idle time of the server is determined by the total idle time of the server divided by the total runtime of the simulation.

Moreover, it has to be decided whether:

- It is possible to leave the queue without being served at all.
- The number of jobs in the queue is limited.
- There are priorities for the jobs (static and/or dynamic).
- It is possible for a job with high priority to interrupt the service for a low-priority job and to occupy the service station immediately when entering the queue.

There are measures of performance for queuing systems available, but, with regard to the complexity of the queuing system investigated, some of which are not well defined.

Let D be the delay in queue of the i th customer, $W_i = D_i + S_i$ be the waiting time in system of the i th customer, $Q(t)$ be the number of customers in queue at time t , and $S(t)$ be the number of customers in system at time t . Then the measures

$$d = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n D_i}{n}$$

and

$$w = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n W_i}{n}$$

are called steady-state average delay d and steady-state average waiting time w . Similarly the measures result in

$$Q = \lim_{T \rightarrow \infty} \frac{\int_0^T Q(t) dt}{T}$$

and

$$L = \lim_{T \rightarrow \infty} \frac{\int_0^T L(t) dt}{T}$$

and are called steady-state time-average number in queue Q and steady-state time-average number or queue length L .

The most important equations for queuing systems among others are

$$Q = \lambda D,$$

and

$$L = \lambda W. \quad (2.19)$$

These equations hold for every queuing system for which D and W exist. Equation 2.19 is the Little formula.

2.3.4 Kendall's Notation

Queuing systems offer a standard notation which can hold the following characteristics:

- s servers in parallel and one FIFO queue feeding all servers.
- A_1, A_2, \dots are random variables.
- S_1, S_2, \dots are random variables.
- A_i and S_i are independent.

Such a queue is called $GI/G/s$ queue, where GI (general independent) refers to the distribution of the $A_{i,s}$ and G (general) refers to the distribution $S_{i,s}$.

If specific distributions are given for the $A_{i,s}$ and $S_{i,s}$, symbols denoting these distributions are used in place of GI and G . Thus, e.g., symbol M is used for the exponential distribution because of the Markovian, i.e., memory loss, property of the exponential distribution, the symbol E_k for a k-Erlang distribution, and D for deterministic (or constant) times. For any $GI/G/s$ queue, the quantity

$\rho = \lambda / sw$ —with sw as service rate of the system when all servers are busy—is called utilization factor of the queuing system. Thus a single-server queuing system with exponential interarrival times and service times and a FIFO queue discipline is called M/M/1 queue, following Kendall's notation which was introduced to standardize the description of queuing models. Kendall introduced a notation for queuing systems, which includes information about the processes, such as job arrivals, and the distribution of the time that is needed in the server. This standard notation is based on a five-character code

$$A/B/c/N/k, \quad (2.21)$$

where A represents the interarrival time distribution, B is the service time distribution, c is the number of parallel servers of a station ($c \geq 1$), N represents the system capacity, and k is the size of the population.

The elements of queues and servers are represented in the term “station.” Hence, a station can be described using Kendall's notation as

$$A/B/c - \langle \text{strategy} \rangle [\text{pre} - \text{emptive}] [\text{maximal queue} - \text{length}] \quad (2.22)$$

The short forms for the mostly used distributions of queuing systems are:

- G : general (no limitation concerning the distribution)
- D : deterministic
- M : exponential distribution

It should be mentioned that the foregoing discussed performance measures can also be analytically computed for $A/B/c$ queues with $c \geq 1$.

Scenario 7

Let Kendall's notation be used as follows:

1. $M/D/1$: represents the simplest example, the FCFS/FIFO principle.
2. $M/G/2$: represents a so-called preemptive systems example, the LCFS/LIFO principle.
3. $MM/1/\infty/\infty$: indicates a single-server system with unlimited queue capacity and infinite calling population. Interarrival times and service times are exponentially distributed.

Queuing systems typically have two states of behavior, short-term or transient, followed by long-term or steady-state behavior. If a queuing system is started, it must operate for a period of time before reaching steady-state conditions. A discrete-event simulation model of a queuing system must run for a sufficiently long period of time to exceed the transient period before measures of steady-state performance can be determined, which results in a specific notation for queuing systems containing:

- Steady-state probability of having n vehicles in system
- Probability of n vehicles in system at time t

- Arrival state
- Effective arrival state
- Effective rate of one server
- Server utilization
- Interarrival time between vehicles $n-1$ and n

Based on this notation for the various classes of queuing system models, a performance analysis can be introduced based on steady-state parameters for:

1. $M/M/1$ queues
2. $M/G/1$ queues
3. $M/E_k/1$ queues
4. $M/D/1$ queues
5. $M/M/1/N$ queues

For the first three queues, the service times are exponentially distributed for M , generally distributed for G , and Erlang distributed for E . For the fourth case, D , the service times are constant. For $M/M/1/N$ queues, the system capacity is limited to N ; and for $M/M/c$ queues, the channels c operate in parallel.

The exponential distribution can be characterized as follows: Let X be an absolute continuous random variable. Let its support—the set of values that the random variable can take—be the set of positive real numbers:

$$R_x = [0, \infty).$$

Let $\lambda \in \mathfrak{R}_{++}$. We say that X has an exponential distribution with parameter λ if its probability density function is:

$$f_x(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } x \in \mathfrak{R}_x \\ 0 & \text{otherwise} \end{cases} \quad (2.23)$$

where parameter λ is called rate parameter.

A random variable having an exponential distribution is also called an exponential random variable.

The Erlang distribution is a continuous probability distribution which was developed to examine the number of telephone calls which might be made at the same time to the operators of the switching stations. This work on telephone traffic engineering has been expanded to consider waiting times in queuing systems. Erlang-distributed random numbers can be generated from uniform distribution random numbers ($U \in (0,1)$) using the following formula:

$$E(k, \lambda) \approx -\frac{1}{\lambda} \ln \prod_{i=1}^k U_i. \quad (2.24)$$

Simulation of queuing systems is often done manually, based on simulation tables. One has to decide, comparing the difference between possible analytical and

Table 2.4 Advances and limitations for analytical and simulation solutions

Solution	Advantages	Limitations
Analytical	Results which are general for use with all possible parameterizations	Preconditions, concerning the distribution of the interarrival times and time to be served
		Substantial problems, to handle queuing strategies
		Numerical efforts, to solve the state equations
		Results only for the steady state
Simulation based	No preconditions concerning the distributions	A single simulation run only corresponds to a single random sample, all simulation results are singular solutions for the given initial state, and they are not general results for the whole model
	Any strategy can be reproduced	
	Observation of the individual history for jobs and queue lengths possible	

simulative solutions, which of the two methods should be used. This comparison can be restricted, reflecting limitations and advances (see Table 2.4).

2.3.5 Inventory System

Another important class of simulation problems of queuing systems involves inventory systems. An inventory system has a periodic review of length at which time the inventory level is observed, and an order that is made to bring the inventory up to a specified level of amount in inventory. At the end of the review period, an order quantity is placed.

Problem: Let us consider an inventory problem that deals with the purchase and sale of parts. The part sellers may buy the parts for 30 US\$ each and sell them for 50 US\$ each. Parts not sold at the end of the month are sold as scrap for 5 US\$ each.

Solution: The problem to be solved with this inventory system is to determine the optimal number of parts the part seller should purchase, which can be done by simulating the demands for a month and recording the profits from sales each day. The profit P can easily be calculated as follows:

$$P = \left(\begin{matrix} \text{sales} \\ \text{revenue} \end{matrix} \right) - \left(\begin{matrix} \text{cost of} \\ \text{parts} \end{matrix} \right) - \left(\begin{matrix} \text{profit loss} \\ \text{excess demand} \end{matrix} \right) + \left(\begin{matrix} \text{salvage sale} \\ \text{scrap parts} \end{matrix} \right). \quad (2.25)$$

Based on the foregoing example, the primary measure of the effectiveness of inventory systems, which are total system costs, can be extracted. Contributing to total inventory cost are the following:

- Item cost which represents the actual costs of the Q items acquired.
- Order costs which are the costs of initiating a purchase or production setup.
- Holding costs which are the costs for maintaining items in inventory.
- Shortage costs represent the costs of failing to satisfy demand.

In general, inventory problems of the type discussed above are often easier to solve than queuing problems.

2.3.6 Simulation Languages

Furthermore, discrete-event simulation of queuing models is based on simulation languages, which use programming languages. Assume that a model consists of two events: customer arrival and service completion. The events can be modeled with event subroutines, which are *ARRIVE* and *DEPART*, respectively. These subroutines contain an *INCLUDE* statement and can be described with generalized statements as follows:

```

SUBROUTINE ARRIVE
INCLUDE 'mm1.dcl'
...
Schedule next arrival
....
IF (SERVER.EQ.BUSY) THEN
....
END
SUBROUTINE DEPART
INCLUDE 'mm1.dcl'
...
Check whether the queue is empty or not
....
IF (NIQ.EQ.0) THEN
....
SERVER = IDLE
....
ELSE
Queue is not empty
NIQ=NIQ+1
....
END

```

2.3.7 Probability in Queuing Systems

In simulating queuing systems, the modeler sees a probabilistic world. The time it takes a system to fail, e.g., a traffic light system at a road intersection, is a random variable, as is the time it takes maintenance to repair the road intersection traffic light system. Thus, modeling probabilistic problems requires skills in recognizing the random behavior of the various phenomena that must be incorporated into the model, analyzing the nature of these random processes, and providing appropriate mechanisms in the model to mimic the random processes.

If X is a variable that can assume any of several possible values over a range of such possible values, X is said to be a random variable.

Let X be a variable in which the range of possible values is finite or countable infinite. For x_1, x_2, \dots , the probability mass function of X is

$$p(x_i) = P(X = x_i) \quad (2.26)$$

$$\begin{aligned} p(x_i) &\geq 0 \text{ for all } i \\ \sum_i p(x_i) &= 1. \end{aligned} \quad (2.27)$$

Assume X is a continuous random variable in which the range of possible values is the set of real numbers $-\infty < x < \infty$. If $f(x)$ is the probability density function of X , then

$$P(a \leq X \leq b) = \int_a^b f(x) dx \quad (2.28)$$

$$\begin{aligned} f(x) &\geq 0 \text{ for all } x \text{ in } \Re \\ \int_{\Re_x} f(x) &= 1. \end{aligned} \quad (2.29)$$

The expected value of the random variable X is given by

$$E(X) = \sum_i x_i p(x_i) \quad \text{if } X \text{ is discrete,} \quad (2.30)$$

and by

$$E(X) = \int_{-\infty}^{\infty} x(x) dx, \quad (2.31)$$

if X is continuous. The expected value is also called the mean, denoted by μ . Defining the n th moment of X results in the variance of the random variable X

$$V(X) = E[(X - E(X))^2] = E[(X - \mu)^2]. \quad (2.32)$$

Random variables can be based on continuous distributions or discrete distributions that are used to describe random phenomena. The focus of using distribution functions is analyzing raw data and trying to fit the right distribution to that data by answering four basic questions about the data to help in its characterization:

1. First question: relates to whether the data can take on only discrete values or whether the data is continuous.
2. Second question: focuses at the symmetry of the data and if there is asymmetry, which direction it lies in. In other words, are positive and negative outliers equally likely or is one more likely than the other?
3. Third question: looks whether there are upper or lower limits on the data; there are some data items like revenues that cannot be lower than zero, whereas there are others like operating margins that cannot exceed a value (100 %).
4. Fourth question: relates to the likelihood of observing extreme values in the distribution; in some data, the extreme values occur very infrequently, whereas in others, they occur more often ([URL 1](#)).

For continuous distributions some of which one can use are the:

- Erlang distribution ***erlang(p)***: is a continuous probability with wide applicability primarily due to its relation to the exponential and Gamma distributions. The Erlang distribution was developed to examine the number of telephone calls which might be made at the same time to the operators of the switching stations, and has been expanded to consider waiting times in queuing systems.
- Exponential distribution ***expo(β)***: fit, evaluate, and generate random samples with regard to interarrival times of vehicles/customers to a system that occur at a constant rate and time to failure of a piece of a component. Parameter β is the scale parameter with $\beta > 0$.
- Uniform distribution (***U/a,b***): also known as rectangular distribution is a distribution that has constant probability. Can be used as a first model for a quantity that is assumed to be randomly varying between parameters a and b but about which little is known. Thus, the uniform distribution is essential in generating random values from all other distribution. Parameters a and b are real numbers with $a < b$; a is the location parameter, and $b - a$ is the scale parameter.
- Normal (or Gaussian) distribution ***N(μ, σ²)***: continuous probability distribution showing that the probability of any real observation will fall between any two real limits or real numbers as the graph approaches zero on either side. Normal distributions are very important in statistics and are often used for real-valued random variables whose distributions are not known. The parameter of the normal distribution is the location parameter $\mu \in (-\infty, \infty)$ and scale parameter σ with $\sigma > 0$.

- Weibull distribution **Weibull**(α, β): continuous probability distribution used as a rough model in the absence of data like time to failure of a component, or time to complete a task, or to describe a particle size distribution, etc. Parameters α and β are so-called shape parameters with $\alpha > 1$ and $\beta > 0$.

For discrete distributions some of which one can use are the:

- Bernoulli distribution **Bernoulli**(p): probability distribution of a random variable with two possible outcomes used to generate other discrete random variates, e.g., *binominal*, *geometric*, and *negative binominal*. Its outcomes can take value 1 with success probability p and value 0 with failure probability $q = 1 - p$. Thus, parameter p holds $p \in (0, 1)$.
- Binomial distribution **bin**(t, p): discrete probability distribution of the number of successes in t independent Bernoulli trials with probability p of success on each trial; number of defective components in a batch of size t , e.g., number of vehicles or passengers of a random size. Parameters are t and p whereby t is a positive integer, and p holds $p \in (0, 1)$.
- Geometric distribution **geom**(p): is either of two discrete probability distributions:
 - Probability distribution of number X of Bernoulli trials needed in finding one success, supported on the set $\{1, 2, 3, \dots\}$
 - Probability distribution of number $Y = X - 1$ of failures before first success, supported on the set $\{0, 1, 2, 3, \dots\}$
- Poisson distribution **Poisson**(λ): discrete probability distribution expressing the probability of a given number of events that occur in an interval of time when the events are occurring at a constant rate; number of components in a batch of random size. Parameter λ holds $\lambda > 0$.

Example 2.1

Assume the number X of defective assemblies in the sample n of manufactured assemblies is binomially distributed. Let $n = 30$ and the probability of defective assembly $p = 0.02$ results in

$$P(X \leq 2) = \sum_{x=0}^2 \binom{30}{x} (0.02)^x (0.98)^{30-x} = 0.5455 + 0.3340 + 0.0988 = 0.9783 \quad (2.33)$$

The mean number of defectives in the sample is

$$E(X) = n \cdot p = 30 \cdot 0.02 = 0.6. \quad (2.34)$$

The variance of defectives in the sample is

$$V(X) = n \cdot p \cdot q = 30 \cdot 0.02 \cdot 0.98 = 0.588. \quad (2.35)$$

Example 2.2

Assume a class of vehicle has a time to failure that follows the Weibull distribution with $\alpha = 200$ h, $\beta = 0.333$, and $\nu = 0$. The mean time to failure yields for the mean Weibull distribution:

$$E(X) = \nu + \alpha \Gamma\left(\frac{1}{\beta} + 1\right) = 200\Gamma(3 + 1) = 200(3!) = 1200 \text{ h}, \quad (2.36)$$

and for the variance Weibull distribution:

$$V(X) = \alpha^2 \Gamma\left(\frac{2}{\beta} + 1\right) - \left[\Gamma\left(\frac{1}{\beta} + 1\right)\right]^2. \quad (2.37)$$

The probability that a vehicle fails before 200 h can be calculated based on the cumulative distribution function of the Weibull distribution as follows:

$$F(x) = 1 - e^{-\left(\frac{x-\nu}{\alpha}\right)^\beta} = 1 - e^{-\left(\frac{200}{200}\right)^{0.333}} = 1 - e^{-2.15} = 0.884 \quad (2.38)$$

2.4 Traffic Demand Models

To analyze and design transportation systems, it is necessary to estimate the existing demand and predict the impact of changes which will result from the transportation planning and development projects considered. In doing so, traffic analysis incorporates a wide spectrum of topics as part of transportation planning and development activities. Thus, traffic analysis is conducted to assist decision makers in improving their transportation planning decisions. One strength of modern traffic demand forecasting is the ability to ask “what if” questions about proposed plans and policies. For this reason, a computerized travel demand forecasting model is used to estimate the relationship between travel demand flows and their characteristics and transportation supply systems and their characteristics.

Traffic demand flow is introduced as an aggregation of individual trips, whereby each trip can be the result of multiple choices made by the users of the transportation system. These users can be individual travelers in passenger transportation, ramp traffic controllers, and/or control tower operators at airports, freight transportation operators, etc. In Cascetta (2009), some classification criteria of travel demand models are introduced, as shown in Table 2.5.

Travel demand models have been designed to include a method for evaluating transport demands with regard to the amount of travel people may choose under specific conditions, e.g., price or transport services. This information is then used to predict roadway traffic volumes and impacts, such as congestion, pollution emissions, etc. Most of the models use a four-step approach (Virginiadot 2014) as shown in Table 2.6.

Table 2.5 Characteristics of travel demand models (Cascetta 2009)

Characteristic	Model
Type of choice	Mobility or context
	Travel
Sequence of choice	Trip-based demand
	Trip chaining
	Activity based
Level of detail	Disaggregate
	Aggregate
Basic assumptions	Behavioral
	Descriptive

Table 2.6 Four-step approach used in travel demand models

Step	Initial tasks
1	Trip generation (the number of trips to be made)
2	Trip distribution (where those trips go)
3	Mode choice (how the trips will be divided among the available modes of travel)
4	Trip assignment (predicting the route trips will take)

The constraints for the task steps in Table 2.6 are shown in Table 2.7.

Once the four steps have been completed, the travel demand forecasting model provides planners with data for existing travel patterns, which are validated and cross-checked to determine how well the model predicts current data, such as park-and-ride utilization, highway vehicle traffic counts, etc.

Besides actual travel surveys, travel demand models use census data to determine the transportation demands, establish baseline conditions, and identify trends. Thus, trips are often predicted separately by purpose (i.e., work, shopping, etc.) and then aggregated into total trips on the network. From this perspective, it can be concluded that travel demand models are designed primarily to identify congestion problems because they mainly measure peak-period motor vehicle trips on major roadways. They generally report roadway level of service (LOS), and a letter grade from A (best) to F (worst) indicates vehicle traffic speeds and delays.

As described in TDM Encyclopedia (2014), travel demand models often incorporate several types of bias favoring automobile transport over other modes and undervaluing travel demand modeling strategies (TRB 2007). The travel surveys they are based on tend to ignore or undercount nonmotorized travel and so undervalue nonmotorized transportation improvements for achieving transportation planning objectives (Stopher and Greaves 2007). Most do not accurately account for the tendency of traffic to maintain equilibrium (congestion causes travelers to shift time, route, mode, and destination) and the effects of generated traffic that result from roadway capacity expansion, and so tend to exaggerate future congestion problems and the benefits of expanding roadway capacity. They are not sensitive to the impacts many types of travel demand model strategies have on trip generation and traffic problems and so undervalue travel demand model benefits.

Table 2.7 Constraints of the four-step travel demand model approach

Step	Constraints
1	Takes into account area factors such as:
	Number and size of households
	Automobile ownership
	Types of activities (residential, commercial, industrial, etc.)
	Density of development, how much travel flows from or to a specific area within the region
	For simplicity, a geographic unit called a transportation analysis zone (TAZ) is used to create trip generation rates for the region
2	Takes into account a certain number of trips generated from each TAZ based on which trip distribution can be analyzed, leading to trip origin and destination points within the region and the number of trips between each pair of TAZ
3	Takes into account the mode of transportation used between trip origins and destinations, i.e., cars, carpools, public transportation, etc.
4	Determines the selected routes taken from origins to destination points, assuming a preference for the fastest route to a destination, based on all kinds of information, such as:
	Actual or predicted congestion and/or other incidents
	Road conditions
	Transit schedules and fares
	Traffic signal systems
	Uses a multicriteria approach to determine the optimal trip assignment

Different reports have been published summarizing information from numerous site surveys, such as:

- Trip and Parking Generation models (Lee et al. 2012).
- Economic Evaluation models (Ellis et al. 2012).
- Integrated Transportation and Land Use models (Dong et al. 2006; Donoso et al. 2006; Scheurer et al. 2009; Bartholomew and Ewing 2009; TRB 2012).
- Simulation models which model the behavior and needs of individual transport users (called agents), rather than aggregate groups, which improves consideration of modes such as walking and cycling; the transport demands of nondrivers, cyclists, and the disabled; and the effects of factors such as parking supply and price, transit service quality, and local land use accessibility factors. Simulation models can provide a bridge between other types of models, since they can incorporate elements from conventional traffic, economic, and land use models. Simulation models have been used for many years on individual projects and are increasingly used for area-wide analysis (TDM Encyclopedia 2014).
- Energy and Emission models (Litman 2013).

Assume a trip-based demand model predicts the average number of trips d with given characteristics executed for a given reference period resulting in the equation (Cascetta 2009)

$$d[C_1, C_2, \dots] = d(SEV, L : D). \quad (2.39)$$

where the average travel demand flow between two transportation analysis zones has the characteristics C_1, C_2, \dots, C_m , which can be expressed as a function of vector SEV , a socioeconomic variable, related to the activity system and/or the decision makers, and of a vector L , level-of-service attributes of the transportation supply system. Demand functions also involve a vector D of coefficients or parameters.

Trip characteristics that are considered relevant in trip-based demand models include (Cascetta 2009):

- u : user's class—category of socioeconomic characteristics
- o, d : zones of trip origin and destination
- p : trip purpose
- t : time period which is the time band in which trips are undertaken
- m : mode used during the trip
- tp : trip path, that is, the series of links connecting centroids o and d over the network, representing the transportation service providers by mode m

With demand flow denoted by $d_{o,d}^u[p, t, m, tp]$, the demand model can be expressed as

$$d_{o,d}^u[p, t, m, tp] = d(SEV, L). \quad (2.40)$$

It is difficult to incorporate freight information into transportation models because freight data is proprietary, and the release of that data is considered to be detrimental to the company's competitive position. Due to the difficulty in acquiring freight data, the inclusion of freight in most transportation plans and models has either been limited in scope or based upon limited sample sizes without knowledge of the contents. In Harris (2008), the Freight Analysis Framework Database, developed and distributed by the Federal Highway Administration, contains freight flows for 114 zones at the national level. This allows the formulation of a travel demand model for truck trips, with vehicles moving through counties of the so-called freight flow zones (TPC_i). These zones can be used to calculate the zonal truck counts for each county as follows:

$$TPC_i = TCZ_{ab} \frac{WF * FLC_i}{\sum FLC_{ij}}, \quad (2.41)$$

2.5 Congested Network Models

Traffic management has become a critical issue as the number of vehicles in metropolitan areas is nearing the existing road capacity, resulting in traffic congestion. In some areas, the volume of vehicles has met and/or exceeded road capacity. However, many roads are constructed with less space than is needed to accommodate the ever-increasing traffic flow, resulting in congestion. Traffic

congestion occurs when the volume of traffic generates a demand for space that is greater than the available road capacity, commonly termed saturation (see Sect. 2.3). There are a number of specific circumstances which cause congestion. The majority of those circumstances are the result of a reduction in road capacity at a given point due to roadwork, weather conditions, accidents, and/or other incidents or an increase in the number of vehicles required for a given transportation volume of people and/or freight. But traffic congestion in transportation is not limited to roads. It is also a problem at airports, at harbors, on railways, and for travelers on public transportation networks.

As introduced in Sect. 2.2, traffic congestion can be studied either at a microscopic level, where the motion of individual vehicles is tracked, or at a macroscopic level, where vehicles are treated as a fluidlike continuum. Therefore, both macroscopic and microscopic models are used to address various traffic flow and congestion phenomena, such as phase transitions, a phenomenon whereby free-flow traffic can spontaneously break down for no obvious reason and persist in a self-maintained congested state for long periods (Kerner and Rehborn 1997). The importance of modeling and controlling traffic congestion can also be seen by reviewing the projects funded by the European Research Council (ERC) for the period 2012–2017. Due to the manifold types of traffic phenomena, traffic flow modeling cannot fully predict under what conditions a traffic jam, defined as heavy but smoothly flowing traffic, may suddenly occur. The reason is that individual incident, such as accidents, a single car braking, an abrupt steering maneuver by a single vehicle, or a truck breakdown, in a previously smooth traffic flow may cause a so-called cascading failure. A cascading failure in a traffic flow system of vehicles means that the failure of one vehicle can trigger the failure of successive vehicles. A cascading failure usually begins when one vehicle of the traffic flow system fails and the effect spreads out and creates a sustained traffic jam. When this happens, nearby traffic nodes must then, if possible, take up the stagnancy caused by the traffic jam which can, in turn, overload those nodes, causing them to fail, resulting in serious congestion. As mentioned in Sect. 2.2, theoretical traffic flow models apply the rules of fluid dynamics to traffic flow, like a fluid flow in a pipe. In spite of the poor correlation of theoretical traffic flow models to actual traffic flow, empirical models have been chosen with the scope to forecast traffic flow. These traffic models use a combination of macro-, micro-, and mesoscopic modeling features, with the addition of entropy effects, by grouping vehicles and randomizing flow patterns within the node segments of the network. These models are then calibrated by measuring actual traffic flows on the links in the network, and the baseline traffic flows are adjusted accordingly (Lindsey and Verhoef 1999; Lindsey et al. 2012).

Traffic flow can be described by the variables density (k), speed (v), and flow (q), measured in vehicles per lane per mile, mile per hour, and vehicles per lane per hour. At the macroscopic level, these variables are defined under stationary conditions at each point in space and time and are expressed by the fundamental equation of traffic flow theory, given in (2.11). For safety reasons, speed usually

declines as density increases which means that the less vehicles per lane per hour, the nearer the vehicles are to driving at free-flow speed v_f . At higher densities, the flow in the flow-density graph (see Fig. 2.1), as well as the speed-flow graph (see Fig. 2.2), drops more rapidly, reaching zero at the congestion density, k_j , where speed and flow are both zero. Thus we can say that the uphill branch of the graph in Figs. 2.1 and 2.2 is referred to as uncongested, unrestricted free traffic flow; and the downhill branch of the graph in Figs. 2.1 and 2.2 is referred to as congested, restricted, or queued. Thus in general mathematical terms of transportation supply models, the speed-traffic flow graph in Fig. 2.2 can be used to formulate relationships among performance, cost, and flow. Hence, interpreting traffic flow as quantity of trips supplied by the road per unit of time, a trip cost curve $C(q)$ can be generated in the form of

$$C(q) = c_0 + \frac{u_c D}{v(q)}, \quad (2.42)$$

where c_0 denotes trip costs, u_c is the unit cost of travel time, D is trip distance, and $v(q)$ is speed expressed in terms of flow. Then the trip cost curve based on (2.42) shows a positively sloped portion corresponding to the congested branch of the speed-flow curve. Thus, $C(q)$ measures the cost of a trip taken by a vehicle. Therefore, the total cost of q trips is then

$$TC(q) = C(q)q, \quad (2.43)$$

and the cost of an additional trip is

$$AC(q) = \frac{\partial TC(q)}{\partial q} = C(q) + q \frac{\partial C(q)}{\partial q}, \quad (2.44)$$

Congestion simulations and real-time observations have shown that in heavy but free-flow traffic, jams can arise spontaneously, triggered by minor events, such as the so-called butterfly effect, e.g., an abrupt steering maneuver by a single motorist. The butterfly effect is the sensitive dependency on initial conditions in which a small change at one place in a deterministic nonlinear system can result in large differences in a later state.

A team of Massachusetts Institute of Technology (MIT) mathematicians (Flynn et al. 2009) has developed a model that describes the formation of “phantom jams,” in which small disturbances (a driver hitting the brake too hard or getting too close to another car) in heavy traffic can become amplified into a full-blown, self-sustaining traffic jam. Key to the study is the realization that the mathematics of such jams, which the researchers call “jamitons,” are strikingly similar to the equations that describe detonation waves produced by explosions, according to Aslan Kasimov, lecturer in MIT’s Department of Mathematics. That discovery enabled the team to solve traffic jam equations that were first theorized in the 1950s.

2.6 Graph Models

In mathematics graph theory, graphs which are mathematical structures used to model pairwise relations between objects are studied. A graph in this context consists of a nonempty set of vertices (or nodes) and a set E of links called “edges” that connect (pairs of) nodes. Each edge has either one or two vertices associated with it, called its “endpoints.” An edge is said to connect its endpoints. A graph can be undirected, meaning that there is no distinction between the two vertices associated with each edge, or its edges can be directed from one vertex to another, etc., which have different formal definitions, depending on what kinds of edges are allowed. In this context, a directed graph is defined as

$$G = (V, E) \quad (2.45)$$

where V consists of a nonempty set V of vertices (or nodes), each node represents a variable, and

$$E \subseteq V \times V. \quad (2.46)$$

where E is the set of directed edges (or arcs) and edges encode the dependencies. Each directed edge $(u, v) \in E$ has a start (tail) vertex u and an end (head) vertex v . Note: A directed graph $G = (V, E)$ is simply a set V together with a binary relation E on V .

In graph theory, the following terminology is of importance. In a simple graph, each edge connects two different vertices; and no two edges connect the same pair of vertices. Multigraphs can have multiple edges connecting the same two vertices. When m different edges connect vertices u and v , we say that $\{u, v\}$ is an edge of multiplicity m . An edge that connects a vertex to itself is called a loop. A pseudograph can include loops as well as multiple edges connecting the same pair of vertices.

For a set V , let $[V]k$ denote the set of k element subsets of V . Equivalently, $[V]k$ is the set of all k combinations of V .

An undirected graph, (2.45) consists of a nonempty set V of vertices (or nodes) and a set

$$E \subseteq [V]^2, \quad (2.47)$$

of undirected edges. Every edge $\{u, v\} \in E$ has two distinct vertices $u \neq v$ as endpoints, and vertices u and v are then said to be adjacent in graph G . Note: The above definitions allow for infinite graphs, where $|V| = 1$.

Table 2.8 shows the terminology of graphs.

Two undirected graphs, $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, are isomorphic if there is a bijection $f: V_1 \rightarrow V_2$ with the property that for all vertices $a, b \in V_1$

$$\{a, b\} \in E_1 \text{ if and only if } \{f(a), f(b)\} \in E_2.$$

Such a function f is called an isomorphism.

An arbitrary undirected graph can be introduced as encoding a set of independencies. As an example, the following rule states when two sets of variables

$$U_1, U_2 \subseteq V, U_1 \cap U_2 = \emptyset$$

are separated in an undirected graph. Let us denote separation by \perp and take it to mean independence in the joint distribution over V :

$$U_1 \perp U_2 \mid U_3 \iff \text{all paths between sets } U_1 \text{ and } U_2 \text{ pass through set } U_3$$

Let U_3 blocks the paths between X and Y ; which can be interpreted as blocking the flow of information. A consequence of this rule is the following Markov property for Markov networks, also called the local Markov property:

$$A \perp \text{everything else} \mid n(A)$$

where $n(A)$ are the neighbors of variable A .

Let a set of variables that separate node A from the rest of the graph be called a Markov blanket for A . Hence, set $n(A)$ is a Markov blanket, and it is the minimal Markov blanket of A . Adding a node to a Markov blanket preserves the Markov blanket property.

Graph theory is widely used to model and study transportation networks:

- Airline networks can be modeled using directed multigraphs where:
 - Airports are represented by vertices.
 - Each flight is represented by a directed edge from the vertex representing the departure airport to the vertex representing the destination airport.
- Road networks can be modeled using graphs where:
 - Vertices represent intersections.
 - Edges represent roads.
 - Unidirected edges represent two-way roads.
 - Directed edges represent one-way roads.

One of the most interesting and powerful features of graphs is their use in modeling structures. With this possibility, one can model relationships, flight schedules, etc. By building a graph model, we use the appropriate type of graph (see Table 2.8) to capture the important features of the application. In a graph-based airport network, vertices represent the airport destinations; and edges represent the airway links between the destinations, as shown in Fig. 2.4.

To model an airport network in which the number of links between the vertices (airports) is important, we can use a multigraph model, as shown in Fig. 2.5.

To model an airport network in which diagnostic links at the vertices (airports) is of importance, we can use pseudograph model where loops are allowed, as shown in Fig. 2.6.

Table 2.8 Graph terminology

Type	Edges	Multiple edges allowed	Loops allowed
Simple graph	Undirected	No	No
Multigraph	Undirected	Yes	No
Pseudograph	Undirected	Yes	Yes
Simple directed graph	Directed	No	No
Directed multigraph	Directed	Yes	Yes
Mixed graph	Directed and undirected	Yes	Yes



Fig. 2.4 Airport network



Fig. 2.5 Multigraph airport network

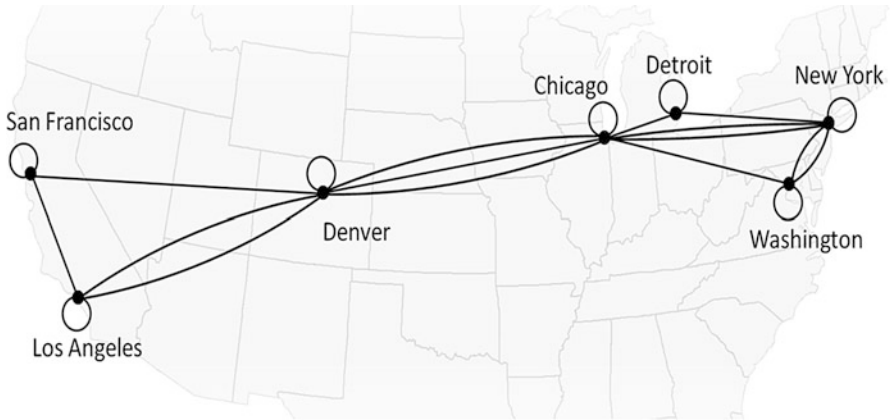


Fig. 2.6 Pseudograph airport network with loops

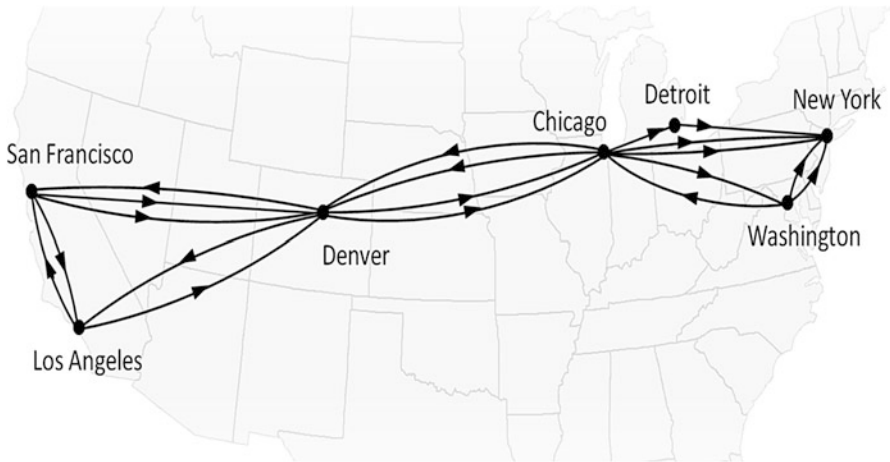


Fig. 2.7 Directed multigraph airport network

To model an airport network in which multiple one-way links between the vertices (airports) are of importance, we can use a directed multigraph model, as shown in Fig. 2.7. It should be noted we could also use a directed graph without multiple edges if we are only interested in whether there is at least one link from a vertex (airport) to another vertex (airport).

Simulations of the air transportation system with detailed models of terminal areas are often applied to optimize current concepts for managing air traffic and investigate future concepts. These simulations can include studies of operations at individual airports or wider regions covering several airports. All of them require detailed terminal and airspace models capable of representing both current and future operating conditions because no two airports are identical. To simulate the

movements of aircraft on the ramp, taxiways, and runways, or passengers in the terminal from check-in through security to gates, or freight to the aircraft, it is convenient to use vertex (node) link (edges) graph models. Modeling aircraft, passenger, or freight movements on a node-link model makes determining the separation between aircraft, passengers, and freight straightforward. In addition, various levels of detail can be modeled depending on the nature of the simulation; and the models can be adapted to existing simulation tools, such as the Airport and Airspace Delay Simulation Model used by the Federal Aviation Administration (FAA) or the Total Airspace and Airport Modeler (Lee and Romer 2011).

The FAA's Airport and Airspace Delay Simulation Model (SIMMOD) is an event-step simulation model which traces the movement of individual aircraft and simulated air traffic control actions required to ensure aircraft operate within procedural rules. The FAA's SIMMOD uses a node-link structure to represent the gate/taxiway and runway/airspace route system. Input parameters depend on the type of aircraft and include permissible airborne speed ranges for use by Air Traffic Control (ATC), runway occupancy times, safety separations, landing roll and declaration characteristics, taxi speeds, and runway/taxiway utilization. Gate utilization depends on the aircraft type and airline (URL 2). With SIMMOD, testing and analyzing the impact of various air traffic scenarios is possible. For this purpose, SIMMOD computes aircraft travel times and delay statistics. SIMMOD can be downloaded from (URL 3). An airfield, based on SIMMOD, is given in (URL 4), showing the runways, terminal buildings, and ramps.

With the Total Airport and Airspace Modeler (TAAM), airports and airspace can be modeled to facilitate planning, analysis, and decision making and to evaluate the impact of changes to infrastructure, operations, and schedules. TAAM is recognized as a standard in the aviation industry and is widely used by Air Navigation Service Providers (ANSP), Civil Aviation Authorities (CAA), airspace planners, airport operators, and major air carriers. TAAM allows modeling of the complete ground operation of an aircraft, preparing scenarios for the airport, creating simulation for a baseline airspace configuration, etc. (URL 4).

2.7 Bottleneck Analysis

The capacity of a transportation system can be modeled as a series of pipes of varying capacity, with the smallest diameter or capacity holding back the entire system. Figure 2.8 illustrates a five-pipe system with different capacities (diameters).

Pipe 2 in Fig. 2.8 represents a bottleneck in the transportation system with regard to capacity. At location *pipe 1* before the bottleneck *pipe 2*, the arrival of vehicles follows a regular traffic flow. If the bottleneck is absent, the departure rate of vehicles at location *pipe 2* is essentially the same as the arrival rate at *pipe 1* at some later time, free-flow travel time T_{FF} . However, due to the bottleneck, the system at location *pipe 2* is now only able to have a departure rate of μ (see Sect. 2.3). The vehicle's arrival at location *pipe 3* takes into account the delay caused by the bottleneck of *pipe 2*. The reason is that output from one pipe

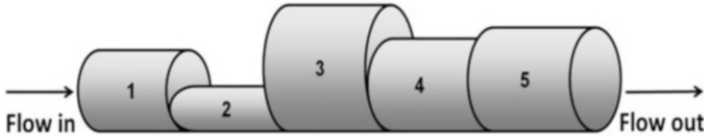
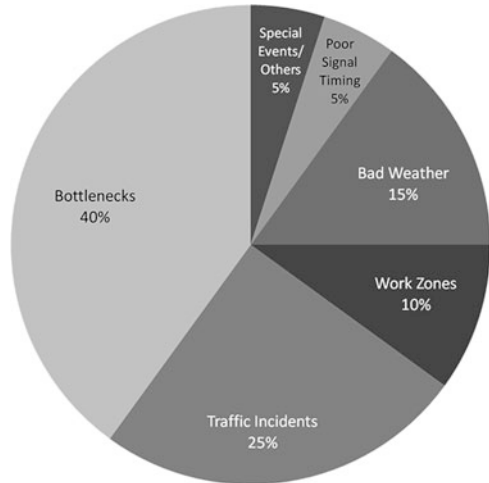


Fig. 2.8 Bottleneck in a transportation system pipe series

Fig. 2.9 Various causes of road congestion



becomes the input to the next until the transportation vehicle exits *pipe 5*. As shown in Fig. 2.9, *pipe 2* cannot handle the traffic flow that *pipe 1* can deliver; and, therefore, it restricts the traffic flow. Because of *pipe 2*'s limited capacities, it restricts the flow from upstream pipes and starves the downstream pipes. *Pipe 3*, *pipe 4*, and *pipe 5* can only work on what *pipe 2* delivers, meaning it determines the transportation system's capacity. Therefore, bottlenecks are important considerations because they impact the traffic flow and thereby the average speed of vehicles. The main consequence of a bottleneck is an immediate reduction in capacity of the transportation system roadway. The Federal Highway Authority has stated that 40 % of all congestion is from bottlenecks, as shown in Fig. 2.9.

Bottlenecks are characterized with regard to their characteristic features, which are stationary and moving bottlenecks. Stationary bottlenecks occur when a multi-lane road is reduced by one or more lanes, which causes the vehicular traffic in the ending lanes to merge into the other lanes.

Let us assume that at a certain location x_0 , the highway narrows to one lane. Thus, the maximum traffic flow rate is now limited to q_{c1l} , since only one lane of the two is available. The traffic flow rate is shared by q_{c1l} and q_c , but its vehicle density k_{c1l} is higher.

As described for Fig. 2.9, we can state that before the first vehicles reach location x_0 , the traffic flow is unimpeded. However, downstream of x_0 , the roadway narrows, reducing the capacity by half. Thus, vehicles begin queuing upstream of x_0 which



Fig. 2.10 Slow tractor creating a moving bottleneck

results in a slower mean space speed v_{qu} of the vehicles compared with the free-flow speed v_f of vehicles. The vehicles driving in the one-lane queue will begin to clear, and the traffic jam can dissipate. But the free space mean speed of the vehicles driving on the now one-lane capacity road will be slower than the vehicles moving at free-flow speed v_f .

Moving bottlenecks are those caused due to slow-moving vehicles, such as trucks, that disrupt the traffic flow. Moving bottlenecks can be active or inactive bottlenecks. If the reduced traffic flow rate caused by a moving bottleneck is greater than the actual traffic flow rate downstream of the vehicle, then this bottleneck is said to be an active bottleneck. In Fig. 2.10, a moving bottleneck is represented by a slow-moving heavy tractor with a mean space speed v_t approaching a downstream location. If the reduced traffic flow rate of the tractor is less than the downstream traffic flow rate, then the tractor becomes an inactive bottleneck.

An analytical expression for capacity reductions caused by a moving bottleneck where each lane has an underperforming flow of traffic can be described in terms of the desired traffic flow rate and can be modeled by the flow conservation equation. This occurs when vehicles pass an observer moving with speed v when traffic is in a steady flow rate-density state (see Fig. 2.1 in Sect. 2.2):

$$q_r = q - kv \quad (2.48)$$

If v , q , and k are given, then q_r is the vertical separation between the corresponding steady-state point on the flow rate versus the density graph in Fig. 2.1. Equation 2.48 applies to an observer that either trails or precedes a moving bottleneck by a substantial but fixed distance, which means the steady-state traffic flow is on either side of the bottleneck. Such a bottleneck is said to be active when,

as a result of its presence, the steady states upstream and downstream of it are different. This occurs in practice when the bottleneck holds back a queue, i.e., when a queue is detected behind it but no queue exists for a long sector of road downstream. Equation 2.48 implies that if a stable passing traffic flow rate q_r exists when an active bottleneck moves at speed v , then the two steady states before and after it must be somewhere on the red line of Fig. 2.1. Therefore, q_r will exclusively denote the passing rate when a bottleneck is active.

Identifying a bottleneck in a transportation system is critical; therefore, the importance of bottleneck analysis cannot be overstated because the results are used not only in determining capacity but also in planning and scheduling traffic flows.

Different methods for bottleneck analysis are known and applied in transportation analysis, such as:

- Capacity utilization
- Queuing time
- Elapsed time
- Shifting shortage

The capacity utilization method refers to the utilization of different resources and calculates the resource with the highest capacity utilization as shortage, which can be calculated after Wang et al. (2005) as follows:

$$B = \{i | p_i = \max(p_1, p_2, \dots, p_n)\} \quad (2.49)$$

with p_i as capacity utilization of the i th resource. The advantage of this method is the intrinsic simplicity making it ideal for transportation planning applications, such as roadway capacity planning and design, congestion management, traffic impact studies, etc. The intersection capacity utilization method is also defined as the sum of ratios of the approach volume divided by the approach capacity for each part of the intersection which controls the overall traffic signal timing plus an allowance for clearance times (Crommelin 1974). Hence, it can be predicted how much reserve capacity is available and how much the intersection is over capacity but does not predict delay. Moreover, the capacity utilization method can be used to predict how often a roadway intersection will cause congestion. But for this, the method requires a specific set of data to be collected which includes traffic volume, number of lanes, saturated traffic flow rates, signal timings, reference cycle length, and lost times for an intersection. Then the method can sum the amount of time required to serve all movements at a saturation rate for a given cycle length and divide it by the reference cycle length. This means that the method is similar to summing critical volume to saturation flow ratios which allow consideration of minimum timings. Moreover, the concept of level of services (LOS) is used whereby LOS reports on the amount of reserve capacity or capacity deficits.

In order to calculate the LOS for intersection capacity utilization, the intersection capacity utilization (*ICU*) must be computed first, which can be achieved as follows:

$$ICU = \left(\max \left(t_{\text{Min}}, \frac{v}{s_i} \right) * RCL + \frac{t_{Li}}{RCL} \right) \quad (2.50)$$

with t_{Min} as minimum green time, critical movement i , v/s_i as volume to saturation flow rate, RCL as reference cycle length, and t_{Li} as lost time for critical movement i (Husch 2003).

Once the *ICU* is fully calculated for an intersection, the *ICU* Level of Service for that intersection can be calculated based on the following criteria (Husch 2003):

- A. If *ICU* is less than or equal to 55%.
- B. If *ICU* is greater than 55% but less than 64%.
- C. If *ICU* is greater than 64% but less than 73%.
- D. If *ICU* is greater than 73% but less than 82%.
- E. If *ICU* is greater than 82% but less than 91%.
- F. If *ICU* is greater than 91% but less than 100%.
- F. If *ICU* is greater than 100% but less than 109%.
- H. If *ICU* is greater than 109%.

This grading criterion shows some specific details about the specific intersection (Husch 2003):

- A. Intersection has no congestion.
- B. Intersection has very little congestion.
- C. Intersection has no major congestion.
- D. Intersection normally has no congestion.
- E. Intersection is on the verge of congested conditions.
- F. Intersection is over capacity and likely to experience congestion periods of 15 to 60 consecutive minutes.
- G. Intersection is 9% over capacity and likely to experiences congestion periods of 60 to 120 consecutive minutes.
- H. The intersection is 9% or greater over capacity and could experience congestion periods of over 120 minutes per day.

To achieve an intersection capacity utilization level of service E or better is not always easy and, therefore, much care is given to the signal timings and geometric bottlenecks, such as lane drops, hard curves, hills, etc., in order to get the LOS to be better than E.

The queuing time method determines the shortage (bottleneck) in relation to the queuing time of the resources before loading and uploading containers for

transportation within the supply chain, which can be calculated after Tan and Bowden (2004) as follows:

$$B = \{i | W_i = \max(W_1, W_2, \dots, W_n)\} \quad (2.51)$$

with W_i as queuing time utilization of the i th resource. The advantage of the method is its easy implementation.

The elapsed time method is a traffic flow scheduling problem in which processing time is associated with their respective probabilities including the transportation time. Finding a good traffic flow schedule for a given set of activities helps transportation managers to effectively control traffic flows and provide solutions for activity sequencing. A traffic flow activity scheduling problem consists when determining the processing sequence for n vehicles on a road network. Therefore, the objective of the elapsed time method can be to minimize the time required to pass a bottleneck. The notation of the elapsed time method contains sequences, vehicles, activity processing times, and the probability associated with activity processing times which finally results in the expected processing time of an activity. The calculation of an elapsed time approach is based on the following criteria which have been mentioned previously:

S Sequence of activities 1, 2, 3, ... n

V_j Vehicle j , $j = 1, 2, \dots, k$

A_1 Processing time of i th activity for vehicle V_1

A_2 Processing time of i th activity for vehicle V_2

A_2 Processing time of i th activity for vehicle V_3

...

P_1 Probability associated to processing time A_1 of i th activity for vehicle V_1

P_2 Probability associated to processing time A_2 of i th activity for vehicle V_2

P_3 Probability associated to processing time A_3 of i th activity for vehicle V_3

...

T_1 Transportation time of i th activity from vehicle 1 to destination D_1

T_2 Transportation time of i th activity from vehicle 2 to destination D_2

T_3 Transportation time of i th activity from vehicle 3 to destination D_1

...

PT_1 Expected processing time of i th activity on vehicle V_1

PT_2 Expected processing time of i th activity on vehicle V_2

PT_3 Expected processing time of i th activity on vehicle V_3

...

The sequence of activities is processed on the vehicles in the order V_i , $i = 1, 2, 3, \dots$ with A_1 , A_2 , and A_3 as processing time of each activity on vehicle V_1 , V_2 , and V_3 ,

respectively, assuming their respective probabilities P_i , $i = 1, 2, 3, \dots$ such that $0 \leq P_i \leq 1$. T_i is the transportation time of the i th activities from vehicles V_i to destination D_i , $i = 1, 2, 3, \dots$, respectively. The algorithm of the given problem is (shown only in part):

- Step 1: Define expected processing time PT_i on vehicle V_i , $i = 1, 2, 3$, as follows:

$$PT_i = A_i \times P_i \quad (2.52)$$

- Step 2: Compute processing time by creating two fictitious vehicles, G and H , with their processing times G_i and H_i , respectively.
- Step 3: Define new reduced problem with processing times G_i and H_i as defined in Step 2.
- Step 4: Find the optimal sequence for two vehicles G and H with processing times of G_i and H_i obtained in Step 2.
- Step 5: Compute the in-out graph for the sequence obtained in Step 4 (Gupta et al. 2013).

The shifting shortage method, in contrast with the sole shortage approach, requested average active time steps of shortages based on which it will be possible to estimate the timeliness shortages are shifting. This allows identification of nonshortages too. This method differentiates between the probability of the existence of shortages and the existence of nonshortages. Moreover, this method allows separation between primary and secondary shortages due to the average shortage over time (Lima et al. 2008). But the primary methodological problem of this method is its implementation and the computing time required. Figure 2.11 shows the principle of moving shortages (with the shortages designated as S1 and S2). As shown in Fig. 2.11, at a specific time step, the shortage is caused by the active periods of tasks which may have the longest runtime. Therefore, the shifting shortage is based on the overlap of shortages.

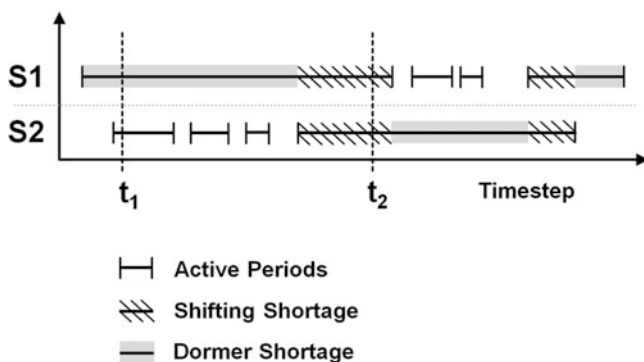


Fig. 2.11 Shifting shortage method

Primary delays as a result of shortages corroborate a belief in so-called distributions of:

- Shipping time.
- Arrival time.
- Quay time for uploading/loading.
- Accomplishable delay compensation through optimization of an objective function to maximize it; for each criterion a higher value will be preferred opposite a lower value.
- Accomplished improvement to compensate for delays.

In general, distribution assumptions can be summarized in a model which allows statistical data analysis (Stahl 2002). However, problems occur with secondary delays, the so-called domino effect, because they start with the distribution assumptions of the primary delays. Secondary delays in maritime transportation, as a consequence of primary delays, can, for example, result in delayed arrival of the following for uploading and loading the containers:

- Trucks
- Trains
- Feeders

The consequences of connection delays can be estimated using mathematical models which allow statistical calculations. The outcome is throughput estimation as a result of the delay, which can be compared with the original assumptions to show the implications of the delay from a general perspective, as well as for a single case study.

Based on the distribution graphs composed, shortages can be identified and rectified through a representative selection of objective functions following the multicriteria approach for simulation runs, which finally results in appropriate adjustments.

As can be seen in Fig. 2.11, short time delays are dominant for the maritime probability transportation chains model. For the shortage analysis, it is important to identify if the resources allocated for the transportation chains can be used without shortages. This means that the transportation job will be operated in an optimal manner. Otherwise, it has to be proven whether the transportation job can be operated with a restricted number of alternatives, meaning a nonempty set of alternatives.

2.8 ProModel Case Study: Road Intersection

Graph theory can be used to model road intersections with running and waiting links. Therefore, vertices (nodes) are usually located at the intersection between road segments included in a model for continuous service in transportation, such as

an urban road network. Typical examples of road intersections in urban areas are individual modes, such as cars, buses, pedestrians, etc., using a road intersection network represented by nodes. Thus, links correspond to connections between nodes to constitute the urban road network. Figure 2.12 shows an example of an intersection in an urban road system in Clausthal-Zellerfeld, Germany, embedded in ProModel.

In this case study, two distinct types of links are considered: running links, which represent a vehicle’s real movements as it moves along the road in the urban road section, and waiting or queuing links, representing queuing at intersections in the city road system, as shown in Fig. 2.13.

The level of detail of the road system depends on the purpose of the model. In this case, the road intersection to be studied is represented by nodes, where the accessed links converge into a four-arm road intersection. The graph model representation for this road intersection is shown in Fig. 2.14 for single node

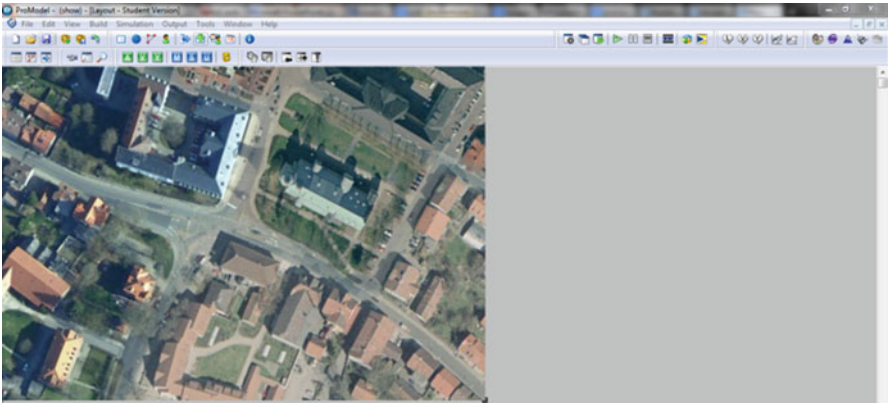


Fig. 2.12 Screenshot of the four-arm road intersection in downtown Clausthal-Zellerfeld, Germany



Fig. 2.13 Representation of road intersection with running and waiting links

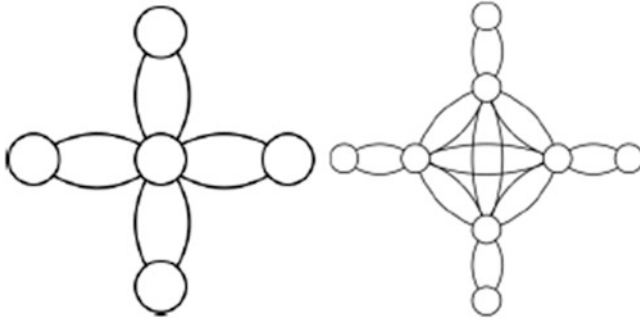


Fig. 2.14 Single node (*left*) model of road intersection in Fig. 2.13 and detailed representation (*right*). For details, see text

representation (left) and a detailed representation, including options for a driving maneuver in the different driving directions of the four-arm intersection (right).

In the single node representation of Fig. 2.13, a left turn can only be achieved if allowed. Moreover, different waiting times cannot be assigned to maneuvers with green-phase duration such as a right turn, for example, in the single node model. This will require a more detailed representation, as given in the right part of Fig. 2.14. In case one wants to expand the model with a parking supply representation, this type of an expanded network representation can be found in Cascetta (2009).

Link performance of a road intersection can be expressed by a cost function introducing the respective performance attributes, which can be travel time along a section, waiting time at the intersection, and monetary cost. In this case, a cost function can be obtained as the sum of the aforementioned performance functions which results in (Cascetta 2009)

$$c(f) = \alpha_1 tr_a(f) + \alpha_2 tw_a(f) + \alpha_3 tm_a(f) \quad (2.53)$$

with

- $tr_a(f)$: function relating to the running time on link a to the flow vector
- $tw_a(f)$: function relating to the waiting time on link a to the flow vector
- $tm_a(f)$: function relating to the monetary cost on link a to the flow vector

and α_i , $i = 1, 2, 3$, are weighting factors.

A more detailed diagram of the node in Fig. 2.14 has been developed by I. A. Jehle for the four-arm intersection in downtown Clausthal-Zellerfeld, as shown in Fig. 2.15:

In Fig. 2.15 the icon has the following meaning:

- Round and elliptic nodes represent the directions from which cars arrive at the junction \rightarrow arriving points.
- Rectangular nodes show the traffic lights (or other traffic flow regulations) for different driving directions.

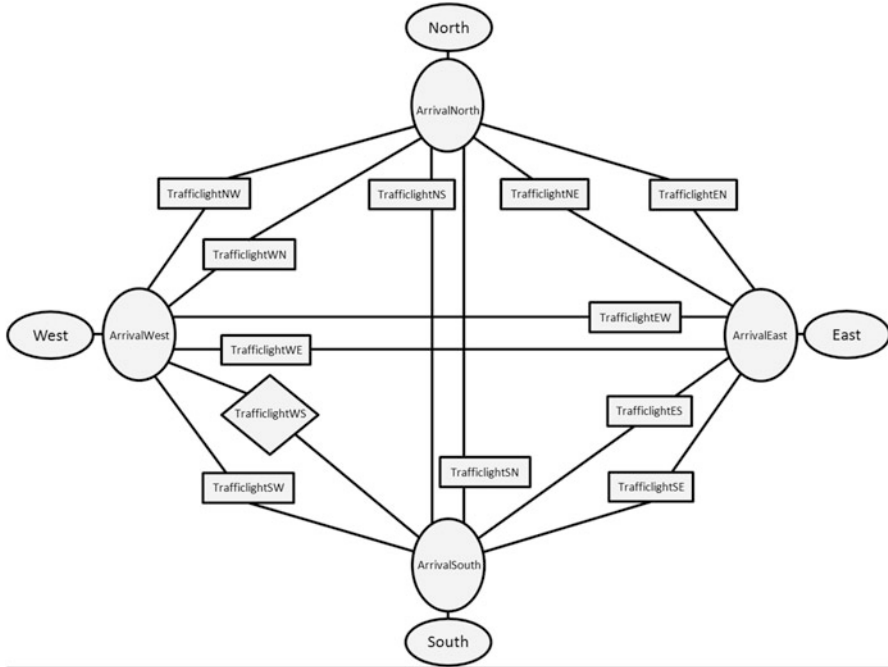


Fig. 2.15 Detailed four-arm road intersection graph model

To simulate the traffic movement at the four-arm intersection in downtown Clausthal-Zellerfeld, the detailed network in Fig. 2.15 has to be transferred into the basic logic components of the ProModel model. In Fig. 2.16, a bird's-eye view of the four-arm intersection is shown which serves as a background for the simulated scenarios in ProModel.

The representation in Fig. 2.13 is the background for a vehicle's movement simulation with ProModel. It refers to:

- Cars arriving at the road intersection from different cardinal directions represented by an icon in the simulation, as shown in Fig. 2.17 for entities
- Entities having a name to identify them and their speed

The node representation in Fig. 2.14 requires, in ProModel, the allocation of the respective locations which have specific attributes, as shown in Fig. 2.18:

- Locations have a name, a capacity describing the number of entities allowed at the location at one time, and rules by which entities enter the location.
- Units describe the number of locations with the same behavior.

The outermost nodes, North, East, South, and West, represent the link with other junctions from which cars arrive at the first simulated location of this direction. The



Fig. 2.16 Bird’s-eye view of the four-arm intersection nodes in downtown Clausthal-Zellerfeld

Entities			
Icon	Name	Speed (mpm)	State
	CarNorth	50	Time Series
	CarEast	50	Time Series
	CarSouth	50	Time Series
	CarWest	50	Time Series

Fig. 2.17 Representation of *entities* of the four-arm intersection

Locations						
Icon	Name	Cap.	Units	DTs...	State	Rules...
	TrafficlightNE	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightNS	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightNW	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightSE	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightES	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightEN	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightSN	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightSE	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightSW	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightWN	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightWE	INF	1	None	Time Series	Oldest, FIFO
	TrafficlightWS	INF	1	None	Time Series	Oldest, FIFO
	North	INF	1	None	Time Series	Oldest
	ArrivalNorth	INF	1	None	Time Series	Oldest
	East	INF	1	None	Time Series	Oldest
	ArrivalEast	INF	1	None	Time Series	Oldest
	South	INF	1	None	Time Series	Oldest
	ArrivalSouth	INF	1	None	Time Series	Oldest
	West	INF	1	None	Time Series	Oldest
	ArrivalWest	INF	1	None	Time Series	Oldest

Fig. 2.18 Representation of *locations* of the four-arm intersection

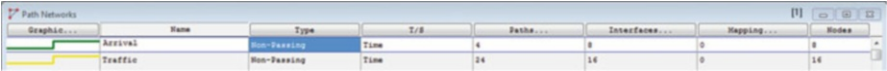


Fig. 2.19 ProModel path network

Arrivals					
Entity...	Location...	Qty Each...	First Time...	Occurrences	Frequency
CarNorth	North	1	0	INF	30
CarEast	East	1	5	INF	30
CarSouth	South	1	10	INF	30
CarWest	West	1	15	INF	30

Fig. 2.20 Arrivals of entities at locations

Table 2.9 Global variables used in ProModel

ID	Type	Initial value	Stats
Drive 1	Integer	0	Time series, time
Drive 2			
Drive 3			
Drive 4			

nodes, *ArrivalNorth*, *ArrivalEast*, *ArrivalSouth*, and *ArrivalWest*, together with the underlying network, represent a complete digraph. To each of the edges of this graph, one node, representing a traffic light, is added. These traffic light nodes are also locations in the ProModel model. All locations of this node model have infinite capacity and process each entity as one unit.

The entities in Fig. 2.17 move on two different path networks, shown in Fig. 2.19, when running the simulation. The path network, *Arrival*, brings entities from the arrival locations to the junction nodes, *ArrivalNorth*, *ArrivalEast*, *ArrivalSouth*, and *ArrivalWest*. The inner network, *Traffic*, shows the graph structure.

The entities arrive at the locations, North, East, South, and West, individually at different times. The first entity which starts the simulation arrives at time 0:00 at the northern point. After 5 s, the next entity enters the traffic junction at the location East, and five and ten seconds later, entities at South and West arrive. These arrivals repeat themselves every 30 s until the end of the simulation. This scenario is shown in Fig. 2.20.

The global variables are shown in Table 2.9.

In this case study, driving phases for a green signal last 60 s. During this 60-s window, the traffic light for one direction shows red-yellow for 1 s, then green for 56 s, and yellow for 3 s. The signals for the other directions show a red light.

The yellow traffic light is divided into two subphases. The first one lasts 1 s and still allows cars to cross the intersection. The next phase, lasting 2 s, requires stopping at the traffic light.

The main algorithm for handling the traffic light signals consists of four different cases realized by three “if” conditions and one “else” case (Jehle 2014):

```

IF (CLOCK (SEC) /x@n=M) AND ( (CLOCK (SEC) @x>0) AND (CLOCK (SEC) @x< (x=y
+1)))
                                //each direction has a green phase of x seconds
                                //signal for direction M green

THEN {DriveM=1}
ELSE {
IF (CLOCK (SEC) /x@n=M) AND (CLOCK (SEC) @x=0)
THEN {DriveM=4 DriveMAlt=22}
                                //signal for direction M red-yellow
                                //and yellow for the former M

ELSE {
IF (CLOCK (SEC) /x@n=M) AND (CLOCK (SEC) @x> (x-y))
THEN {DriveM=21}              //signal for direction M yellow
ELSE {DriveM=3}               //signal for direction M red
}
}

```

The algorithm of the above form can be used for every length x of the green phase and different numbers of driving directions. In this case study, there are four sets of driving directions. Each category combines some directions, which can be used without the need to cross the path of another car, as shown in Table 2.10.

The driving directions shown in Table 2.10 allow the logic macros of the ProModel simulation model to be specified as shown in Table 2.11.

The processing decides, for each entity, the way through the path network and the behavior during the simulation. The following algorithm shows the link between traffic light signals and the movement of the simulated cars for direction M.

```

IF (DriveM=1) OR (DriveM=21)      //drive if signal is green or yellow
THEN {MOVE ON Traffic}
ELSE {WAIT UNTIL (Drive1=1) OR (Drive1=21)}
                                //else wait until signal changes

```

This logic was implemented together with the macros in the processing table, showing the parts of the four entities in Tables 2.12, 2.13, 2.14, and 2.15.

Table 2.10 Driving directions

M	Driving directions
1	From North to South and West; from South to North and East
2	From North to East; from South to West
3	From East to West and North; from West to East and South
4	From West to North; from East to South

Table 2.11 Macros of the ProModel simulation model

ID	Macros	Operations
OpLogic1	IF (CLOCK(SEC)/60@4=1)AND	None
	((CLOCK(SEC)@60>0)AND(CLOCK(SEC)@60<58))	
	THEN {Dr ive1=1}	
	ELSE {	
	IF (CLOCK(SEC)/60@4=1)AND(CLOCK(SEC)@60=0)	
	THEN {Dr ive1=4 Dr ive4=22}	
	ELSE {	
	IF (CLOCK(SEC)/60@4=1)AND(CLOCK(SEC)@60>57)	
	THEN {Dr ive1=21}	
	ELSE {Dr ive1=3}	
OpLogic2	}	None
	}	
	IF (CLOCK(SEC)/60@4=2)AND	
	((CLOCK(SEC)@60>0)AND(CLOCK(SEC)@60<58))	
	THEN {Dr ive2=1}	
	ELSE {	
	IF (CLOCK(SEC)/60@4=2)AND(CLOCK(SEC)@60=0)	
	THEN {Dr ive2=4 Dr ive1=22}	
	ELSE {	
	IF (CLOCK(SEC)/60@4=2)AND(CLOCK(SEC)@60>57)	
OpLogic3	THEN {Dr ive2=21}	None
	ELSE {Dr ive2=3}	
	}	
	}	
	IF (CLOCK(SEC)/60@4=3)AND	
	((CLOCK(SEC)@60>0)AND(CLOCK(SEC)@60<58))	
	THEN {Dr ive3=1}	
	ELSE {	
	IF (CLOCK(SEC)/60@4=3)AND(CLOCK(SEC)@60=0)	
	THEN {Dr ive3=4 Dr ive2=22}	
OpLogic4	ELSE {	None
	IF (CLOCK(SEC)/60@4=3)AND(CLOCK(SEC)@60>57)	
	THEN {Dr ive3=21}	
	ELSE {Dr ive3=3}	
	}	
	}	
	IF (CLOCK(SEC)/60@4=0)AND	
	((CLOCK(SEC)@60>0)AND(CLOCK(SEC)@60<58))	
	THEN {Dr ive4=1}	
	ELSE {	
	IF(CLOCK(SEC)/60@4=0)AND(CLOCK(SEC)@60=0)	
	THEN {Dr ive4=4 Dr ive3=22}	
	ELSE {	
	IF (CLOCK(SEC)/60@4=0)AND(CLOCK(SEC)@60>57)	
	THEN {Dr ive4=21}	
	ELSE {Dr ive4=3}	
	}	
	}	

Table 2.12 Processing of entity CarNorth

Entity	Location	Operation	Output	Destination	Rule	MoveLogic
CarNorth	TrafficlightNE	OpLogic1 OpLogic2	CarNorth	ArrivalEast		IF (Drive2=1) OR (Drive2=21) THEN {MOVE ON Traffic} ELSE {WAIT UNTIL (Drive2=1) OR (Drive2=21)}
	TrafficlightNS	OpLogic3 OpLogic4		ArrivalSouth		
	TrafficlightNW			ArrivalWest		
	ArrivalNorth			TrafficlightNE	0.4	MOVE ON Traffic
				TrafficlightNS	0.3	
				TrafficlightNW	0.31	
	North			ArrivalNorth	First	MOVE ON Arrival
	ArrivalEast			EXIT	First	
	ArrivalSouth					
	ArrivalWest					

Table 2.13 Processing of entity CarEast

Entity	Location	Operation	Output	Destination	Rule	MoveLogic
CarEast	TrafficlightEN	OpLogic1	CarEast	ArrivalNorth		IF (Drive3=1) OR (Drive3=21) THEN {MOVE ON Traffic} ELSE {WAIT UNTIL (Drive3=1) OR (Drive3=21)}
	TrafficlightES	OpLogic3		ArrivalSouth		
	TrafficlightEW	OpLogic4		ArrivalWest		
	ArrivalNorth			EXIT		
	ArrivalEast			TrafficlightEN	0.4	MOVE ON Traffic
				TrafficlightES	0.3 1	
				TrafficlightEW	0.3	
	East			ArrivalEast	First1	MOVE ON Arrival
	ArrivalSouth			EXIT		
	ArrivalWest					

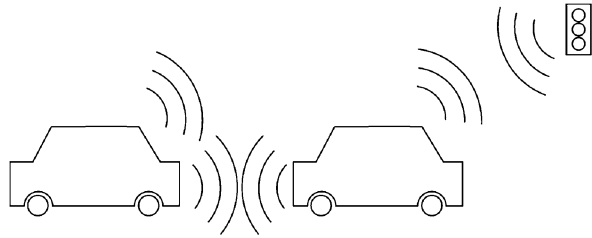
Table 2.14 Processing of entity CarSouth

Entity	Location	Operation	Output	Destination	Rule	MoveLogic
CarSouth	TrafficlightSN	OpLogic1 OpLogic2	CarSouth	ArrivalNorth		IF (Drive1=1) OR (Drive1=21) THEN {MOVE ON Traffic} ELSE {WAIT UNTIL (Drive1=1) OR (Drive1=21)}
	TrafficlightSE	OpLogic3 OpLogic4		ArrivalEast		
	TrafficlightSW			ArrivalWest		
	ArrivalNorth			EXIT	First1	
	ArrivalEast					
	ArrivalSouth			TrafficlightSN	0.3 1	MOVE ON Traffic
				TrafficlightSE	0.4	
				TrafficlightSW	0.3	
	South			ArrivalSouth	First1	MOVE ON Arrival
				EXIT	First1	

Table 2.15 Processing of entity CarWest

Entity	Location	Operation	Output	Destination	Rule	MoveLogic
CarWest	TrafficlightWN	OpLogic1 OpLogic2	CarWest	ArrivalNorth		IF (Drive3=1) OR (Drive3=21) THEN (MOVE ON Traffic) ELSE { WAIT UNTIL (Drive3=1) OR (Drive3=21) }
	TrafficlightWE	OpLogic3 OpLogic4		ArrivalEast		
	TrafficlightWS			ArrivalSouth		
	ArrivalNorth			EXIT	FIRST 1	
	ArrivalEast					
	ArrivalSouth					
	ArrivalWest			TrafficlightWN	0.3 1	MOVE ON Traffic
				TrafficlightWE	0.3	
				TrafficlightWS	0.4	
	West			ArrivalWest	FIRST 1	MOVE ON Arrival

Fig. 2.21 Introducing the Internet of Things paradigm into transportation system analysis



With regard to the advent of Internet and communication technologies and especially the new paradigm of the Internet of Things, vehicles and traffic lights at road intersections will be equipped with radio frequency identification (RFID) technology in the near future, enabling them to communicate wirelessly with each other and with traffic light systems. With the Internet of Things, it will be possible to send wireless radio signals from a traffic light system to vehicles approaching the traffic light system at a road intersection. By monitoring the actual speed of the vehicles, their arrival time at the road intersection traffic light system will be calculated. Based on this calculation, the vehicle will arrive at the traffic light system when it has changed from red to green. Thus the traffic light management system will send information to the vehicle's engine management system to slow down the vehicle's speed to ensure that the vehicle arrives on time at the next green phase of the road intersection traffic light, as illustrated in Fig. 2.21. This will help to avoid unnecessary accelerating and braking actions, which waste energy, and reduce CO₂ emissions and pollution. Moreover, it would be possible for vehicles in the near future to adjust their speed and distance via intercar communication through the Internet of Things (Moeller et al. 2013).

There is already some statics technology available for speed optimization through intercommunication with traffic light systems in a certain distance from a traffic light road intersection with the impact of different advisory speeds according to the time to the next signal change at the road intersection.

2.9 Exercises

1. Explain what is meant by the term cost-effectiveness.
2. Give an example of cost-effectiveness in transportation.
3. Explain what is meant by the term cost-benefit analysis.
4. Give an example of a cost-benefit approach in transportation.
5. Explain what is meant by the term lifecycle cost analysis.
6. Give an example of a lifecycle cost analysis in transportation.
7. Explain what is meant by the term trip generation.
8. Give an example of trip generation in transportation.
9. Explain what is meant by the term trip distribution.

10. Give an example of trip distribution in transportation.
11. Explain what is meant by the term mode split.
12. Give an example of mode split in transportation.
13. Explain what is meant by the term route assignment.
14. Give an example of route assignment in transportation.
15. Explain what is meant by the term level of service.
16. Give an example of a level of service in transportation.
17. Explain what is meant by economic transportation and land use models.
18. List and define the main characteristic statements.
19. To what specific approaches do traffic flow models refer?
20. Describe the two approaches for flow rate versus density and speed versus flow rate.
21. Explain what is meant by the term free-flow speed.
22. Describe the mathematical equation for free-flow speed in detail.
23. Explain what is meant by the term macroscopic traffic flow model.
24. Give an example of a macroscopic traffic flow model.
25. Explain what is meant by the term queuing model.
26. Describe how queuing analysis allows determining the time needed under congested conditions.
27. Explain what is meant by Little's formula.
28. Describe the mathematical background for Little's formula.
29. Explain what is meant by Kendall's notation.
30. Describe a case study example by using Kendall's notation.
31. Explain what is meant by the term FCFS or FIFO.
32. Give an example of FCFS and FIFO in transportation.
33. Explain what is meant by the term preemptive shortest activity first.
34. Give an example of preemptive shortest activity first in transportation.
35. Explain what is meant by the term Erlang distribution.
36. Describe the mathematical equation for the Erlang distribution.
37. Explain what is meant by the term traffic demand model.
38. Give an example of a traffic demand model.
39. Explain what is meant by the congestion network model.
40. Describe the results shown in Fig. 2.10 in your own words.
41. Explain what is meant by the term graph model.
42. Give an example of a multigraph in transportation.
43. Explain what is meant by the term bottleneck analysis.
44. Give an example of a bottleneck in transportation.
45. Explain what is meant by the road intersection.
46. Give an example of a four-arm road intersection.
47. What is meant by the term radio frequency identification?
48. Give an example of a radio frequency identification application in the transportation system sector.
49. What is meant by Internet of Things?
50. Give an example of an Internet of Things application in the transportation system sector.

References and Further Readings

- Bando M, Hasebe K, Nakayama A, Shibata A, Sugiyama Y (1995) Dynamical model of traffic congestion and numerical simulation. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 51(2):1035–1042
- Bartholomew K, Ewing R (2009) Land use-transportation scenarios and future vehicle travel and land consumption: a meta-analysis. *J Am Plann Assoc* 75(1):13–27
- Cascetta E (2009) *Transportation systems analysis: models and applications*. Springer, New York
- Crommelin RW (1974) Employing intersection capacity utilization values to estimate overall level of service. *Trans Res Board* 44(10):11–14, Transportation Research Record, Washington, DC
- CUTR (2007) Economics of travel demand management: comparative cost effectiveness and public investment, Center for Urban Transportation Research, at www.nctr.usf.edu/pdf/77704.pdf
- D'Agostino RB, Stephens MA (1986) Eds.: *Goodness-of-fit tests*. Marcel Dekker Publ.
- Dagpunar J (1988) *Principles of random variate generation*. Clarendon, Oxford
- DeGroot MH (1975) *Probability and statistics*. Addison Wesley Publ., Reading
- Doboszczarek S, Forstall V (2013) Mathematical modeling by differential equations. <http://www.norbertwiener.umd.edu/Education/m3cdocs/Presentation2.pdf>
- Dong X, Ben-Akiva M, Bowman J, Walker J (2006) Moving from trip-based to activity-based measures of accessibility. *Transp Res A* 40(2):163–180
- Donoso P, Martinez F, Zegras C (2006) The Kyoto protocol and sustainable cities: potential use of clean-development mechanism in structuring cities for carbon-efficient transportation. *Transp Res Rec* 1983:158–166
- Ellis D, Glover B, Norboge N (2012) Refining a methodology for determining the economic impacts of transportation improvements, University Transportation Center for Mobility at Texas A&M University. http://utcm.tamu.edu/publications/final_reports/Ellis_11-00-68.pdf (ERC) <http://transport.epfl.ch/modelling-controlling-traffic-congestion>
- Evans M, Hastings N, Peacock B (2000) *Statistical distributions*. Wiley Publ., New York
- Fishman GS (1973) Statistical analysis for queuing simulations. *Management Science* Vol. 20:363–369
- Fishman GS (2006) *Monte carlo: concepts, algorithms, and applications*. Springer Publ., New York
- Flynn MR, Kasimov AR, Nave J-C, Rosales RR, Seibold B (2009) Self-sustained nonlinear waves in traffic flow. *Phys Rev E* 79(5):056113. doi:10.1103/PhysRevE.79.056113, DOI:10.1103/PhysRevE.79.056113#_blank
- Gupta D, Singla S, Singla P, Singh S (2013) Minimization of elapsed time in $N \times 3$ flow shop scheduling problem, the processing time associated with probabilities including transportation time. *Int J Innovat Eng Technol Special Issue ICAECE-2013*, pp 38–42
- Harris GA (2008) Bridging the data & information gap, project report no. AL-26-7262-01
- Hogg RV, Craig AF (1995) *Introduction to mathematical statistics*. Prentice-Hall Publ., Englewood Cliffs
- Husch D (2003) *Intersection capacity utilization*. Trafficware, Albany
- Immers LH, Logghe S (2002) *Traffic flow theory*, course H 111. Catholic University Leuven
- Jehle IA (2014) Simulation of a traffic light junction, student project work, TU Clausthal, Germany
- Kerner BS, Rehborn H (1997) Experimental properties of phase transitions in traffic flow. *Phys Rev Lett* 79:4030–4033
- Kleijnen JPC (1974) *Statistical techniques in simulation*. Marcel Dekker Publ, New York
- Kockelman KM (2001) Modeling traffic's flow-density relation: accommodation of multiple flow regimes and traveler types. *Transportation* 29(4):363–373
- Lee H-T, Romer TF (2011) Automating the process of terminal area node-link model generation. *J Guid Control Dyn* 34(4):1228–1237
- Lee R, Niemeier D, Parker T, Handy S (2012) Evaluation of operation and accuracy of available smart growth trip generation methodologies for use in California, Transportation Research Record. *J Transport Res Board, Serial Issue Number* 2307:120–131

- Lima E, Chwif L, Baretto M (2008) Methodology for selecting the best suitable bottleneck detection method. In: Proceedings SCS 32nd conference on winter simulation. Society for Computer Simulation International, San Diego, pp 740–754
- Lindsey CR, Verhoef ET (1999) Congestion modelling, Series Tinbergen Institute discussion papers, Number 99–091/3
- Lindsey CR, van den Berg VAV, Verhoef ET (2012) Step tolling with bottleneck queuing congestion. *J Urb Econ* 72(1):46–59
- Little JDC, Graves SC (2008) Little's law, Chapter 5. In: Chhajed D, Loewe TJ (eds) Building intuition: insight from basis operations management models and principles. Springer Publ., pp 81–100
- Litman T (2006) What's it worth? Economic evaluation for transportation decision making, transportation association of Canada, (www.tac-atc.ca); at www.vtpi.org/worth.pdf
- Litman T (2013) Congestion costing critique: critical evaluation of the urban mobility report. In: VTPI, (www.vtpi.org); at www.vtpi.org/UMR_critique.pdf
- Moeller DPF, Haas R, Vakilzadian H (2013) Ubiquitous learning: teaching modeling and simulation (M&S) with technology. In: Vakilzadain H, Crosbie R, Huntsinger R, Cooper K (eds) Proceedings of the 2013 summer simulation multiconference GCMS 2013. Curran Publication, Red Hook, pp 125–132
- Muench S (2004) Traffic Concepts, http://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDEQFjAA&url=http%3A%2F%2Fcourses.washington.edu%2Fcee320w%2Flectures%2FTraffic%2520Concepts.ppt&ei=EkpRU82uC4e1tAajiYGYCw&usg=AFQjCNFOFglgi-iDYswasEvo91YzMZ_OKg&bvm=bv.65058239,d.Yms
- Scheurer J, Horan E, Bajwa S (2009) Benchmarking public transport and land use integration in Melbourne and Hamburg: hints for policy makers. Presentation at 23rd AESOP 2009 congress, Liverpool
- Small KA (1998) Project evaluation, Chapter 5. In: Gómez-Ibáñez JA, Tye W, Winston C (eds) Transportation policy and economics: a handbook in honor of John R. Meyer. www.brookings.edu, www.uctc.net/papers/379.pdf
- Stahl WA (2002) Statistical data analysis (in German). Vieweg Publ. Wiesbaden
- Stopher PR, Greaves SP (2007) Household travel surveys: where are we going? *Transp Res A* 41(5):367–381
- Sundarapandian V (2009) Queuing theory, Chapter 7. In: Probability, statistics and queuing theory. PHI Learning, New Delhi
- Tan AC, Bowden RO (2004) The Virtual Transport System (VITS)—final report. Monograph, Transportation Research Board, Washington, DC
- TDM Encyclopedia (2013) Transport model improvements: improving methods for evaluating the effects and value of transportation system changes; see <http://www.vtpi.org/tdm/tdm125.htm>
- TDM Encyclopedia (2014) Victoria Transport Policy Institute, Victoria
- Transportation Research Board, see <http://onlinepubs.trb.org/onlinepubs/sr/sr288.pdf>
- TRB (2007) Metropolitan Travel Forecasting: Current Practice and Future Direction, Special Report 288, Transportation Research Board (www.trb.org); at <http://onlinepubs.trb.org/onlinepubs/sr/sr288.pdf>
- TRB (2012) The effect of smart growth policies on travel demand, capacity project C16, Strategic Highway Research Program (SHRP 2), (www.trb.org); <http://onlinepubs.trb.org/onlinepubs/shrp2/SHRP2prepubC16.pdf>
- USDOT (2003) US Department of Transportation, Economic Analysis Primer, Office of Asset Management, FHWA, USDOT, 2003. www.fhwa.dot.gov/infrastructure/asstmgt/primer.pdf
- Virginadot (2014) http://www.virginadot.org/projects/resources/vtm/what_is_travel_demand_modeling.pdf
- Wang Y, Zhao Q, Zheng D (2005) Bottlenecks in production networks: an overview. *J Syst Sci Syst Eng* 14:18ff
- Yücesan E, Schruben LW (1998) Complexity of simulation models: a graph theoretic approach. *Journal of Comput* 10:94–108

Links

(URL 1) http://pages.stern.nyu.edu/~adamodar/New_Home_Page/StatFile/statdistns.htm

(URL 2) <http://www.tc.faa.gov/acb300/aasw.asp>

(URL 3) http://www.tc.faa.gov/acb300/more_simmod.asp

(URL 4) [http://ww1.jeppesen.com/support/technical_support_details.jsp?prodNameTxt2=Total
%20Airspace%20and%20Airport%20Modeler%20%28TAAM%29](http://ww1.jeppesen.com/support/technical_support_details.jsp?prodNameTxt2=Total%20Airspace%20and%20Airport%20Modeler%20%28TAAM%29)

Introduction to Transportation Analysis, Modeling and
Simulation

Computational Foundations and Multimodal
Applications

Möller, D.P.F.

2014, XIV, 343 p. 131 illus., 67 illus. in color.,

ISBN: 978-1-4471-5637-6