This introduction describes an analytical process that consists of predicting the output or performance of a transportation system as a function of specified inputs. Such an analytical process can take the form of a very simple equation relating the average number of vehicles on a roadway to the mean speed of these vehicles, or can take the form of a complex simulation model. It is clear that for a transportation analyst, knowledge of the system at hand - i.e. the flow characteristics and the analytical techniques - is essential to correctly predict its performance.

Before providing an in-depth description of the analytical framework, let us at this point refer to the chapter IV of this reader, which handles the application of (simulation) models in the analysis of traffic systems in general, and in particular to the assessment of modifications.

Figure 1.1 (page 5) depicts a flowchart from [36], illustrating the analytical process. The flowchart emphasizes the importance of knowing fundamental flow characteristics and basic analytical techniques. Although the process can be represented in many ways, the most common approach is to predict the performance of the traffic system as a function of traffic demand, transport supply, traffic control, and environment information. The predicted performance may include both performance measures relating to the user (travel times, comfort levels), to the system (throughput), or to society (pollution, noise).

The traffic demand is generally specified in terms of demand flow rates, or time headway distributions for the location and time period selected for analysis. The transport supply features are generally converted in capacity values or minimum time headway values. The resulting operations include predictions of speeds, operating speeds, distance headway distributions, and/or density levels. The control elements must be inserted into the supply – demand process and are used to modify predicted performance. Without an understanding of the flow characteristics, these tasks cannot be performed.

The applicability of different analytical techniques is determined by the complexity of the situation and the problem at hand. In some cases, we can tackle the problems using relatively simple models, in other situations, more involved methods are required. In this respect, the analyst should always make a prudent trade-off between research cost and benefits, given the research aims and constraints. Without knowledge of traffic analytical techniques, and the ability to select the most appropriate microscopic or macroscopic method for the problem at hand, an analyst cannot solve the problems in the traffic system adequately.

Chapter 1

Demand-supply analysis

In principle, all analytical techniques of traffic systems are structured in a demand-supply framework. This framework can be at the microscopic level of analysis, in which individual traffic units are studied, or at the macroscopic level of analysis, in which attention is given to groups of traffic units in aggregate form. These demand-supply analytical techniques vary and the analyst must not only learn these various techniques but must develop skills in selecting the most appropriate technique for the problem at hand. Several analytical techniques are described in this reader and include capacity analysis, shock wave analysis, traffic flow models, queuing analysis, and simulation. In part 2 of these notes, only macroscopic characteristics and techniques are considered in detail; part 3 considers microscopic analysis.

This introduction chapter is devoted to presenting an analytical framework for supply-demand analysis. The major elements of the framework are described and their interactions are identified and discussed briefly[1].

1.1 Analytical framework

The analytical framework consist of two processes: an initial process and one or more feedback processes. Fig. 1.1 schematizes the analytical framework. The input of the analytical process is used to predict the performance of the traffic system. Once the initial performance is predicted, the input is modified as a feedback process. These modifications can be user-selected or part of an optimization and/or demand-related process. The remaining sections of this chapter discuss both the initial analysis and the feedback processes.

[1]This chapter is an abbreviated version of chapter 8 of [36].

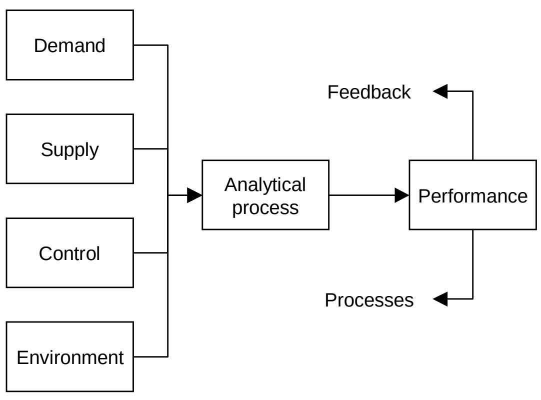

Figure 1.1: An analytical framework for demand-supply analysis of traffic systems

1.2 Initial analysis

The *demand input* represent the number of units that would like to be served. Clearly, the demand is always larger than the number of units that can be served. In microscopic analysis, the demand input is expressed as the arrival times (or equivalently, the headways) of the individual units, while in macroscopic analysis the demand input is expressed as an arrival rate (unit per time interval). It is important to note that in oversaturated conditions, i.e. when the demand exceeds the capacity, more multiple time periods need to be considered. In this way, excess demands in earlier time periods can be transferred and served in later, undersaturated time periods. The time frame for the demand can be for a current point in time, for some future year, or simply a demand modification due to supply or control changes (latent demand).

The *supply input* represents the maximum number of units that can be served. In microscopic analysis, the supply input is expressed as the minimum time headway between units to be served, while in macroscopic analysis the supply input is expressed as the maximum flow rate in unit per time interval. This maximum flow rate is often referred to as the capacity and generally is assumed to remain constant over time periods or traffic states. However, as we have seen, capacity can change between time periods or traffic states due to changes in the demand patterns, vehicle mix, driver characteristics, control states, and / or environment. Often, the analyst begins with the actual physical dimensions of the traffic facility and must translate these features into capacity values, which can be a rather involved process.

The *control input* consists of a set of rules as to how the units interact with each other and with the system. The control can be constant over time, such as car-following rules, rules of service (e.g. FIFO), traffic restrictions, vehicle performance capabilities, and so on. The control can be also time-dependent, such as traffic signals, or during peak periods. The control can be traffic responsive, such as railroad crossings, ramp control, traffic signals, or when drivers modify their behavior due to impatience or frustration. Finally, it must be noted that 'no control' is also a control state that can be encountered.

The *environment input* generally serves to modify the other inputs; namely demand, supply and control. Situations encountered include visibility, weather conditions, pavement conditions, unusual distractions, and others. In practice the performance may be overestimated because the traffic system is assumed to be under 'normal' environmental conditions, that is, under near-perfect conditions.

The analytical technique processes the above-mentioned input data to predict the performance of the traffic system. In the microscopic level of analysis (part 3 of this reader), analytical techniques may include car-following theories, time-space diagrams, stochastic queueing analysis and microscopic simulation. At the macroscopic level of analysis discussed in part 2 of this syllabus, these techniques may include capacity analysis, speed-flow-density relations, shock wave analysis, deterministic queueing analysis, and macroscopic simulation. These analytical techniques can be accomplished manually or using a computer, can utilized mathematical expressions or simulation, and can lead to solutions of a deterministic or a stochastic form.

The objective of demand-supply analysis is to *predict the performance of the traffic system from the viewpoint of the users and from the perspective of the total system.* The users will be interested in their travel times, incurred delays, queueing, comfort, risks, and energy consumption. The system manager will be concerned with system performance, level-of-service, air pollution, noise generation, accident rates, and total transport costs. The formulation often takes the form of a multi-objective function with a set of constraints. For example, at an isolated signalized intersection the objective may be to minimize a combination of delays and stops with the constraint of a maximum individual delay. Generally, as the size and the complexity of the traffic system increases, the formulation becomes more comprehensive and more complex as well.

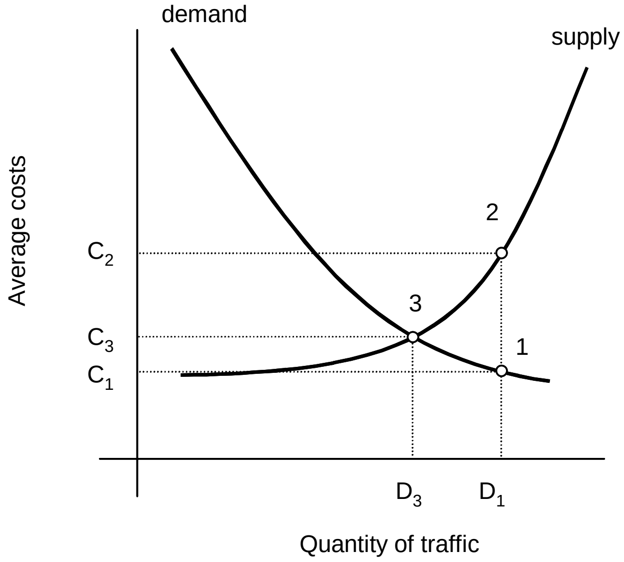

Figure 1.2: Demand-supply equilibrium

1.3 Feedback processes

Once the initial analysis is completed and the performance is predicted, a feedback process (or an iterative process) is generally desired or even required. These feedback processes take various forms and include determining demand-performance equilibrium, answering "what-if" statements, conducting sensitivity analysis, optimizing an objective function, and / or investigating traveler responses due to system changes.

Often the inputs to the analysis, such as demand and supply, are developed independently without knowledge as to how the traffic system will perform. When the initial analysis is completed, the anticipated demand level may not be compatible or appropriate for the resulting predicted performance. In illustration, consider a very simple single-link traffic system with demand-supply cost functions shown in Fig. 1.2. An initial demand level of D_1 was assumed with an anticipated user-cost of C_1. However, once the initial process is completed, a user cost of C_2 (rather than C_1) was predicted, with $C_2 > C_1$. The feedback process requires modifying the demand input until the demand and the specified supply are in equilibrium as denoted in point 3 in Fig. 1.2, where the demand level is reduced to D_3 with an associated average cost of C_3.

A second type of feedback process would be answering "what if" statements. For example, in investigating a rural road, the analyst might want to estimate the change in predicted performance *if* a 500-meter passing lane is added at the midpoint of the upgrade. The analyst would perform the initial process and then in the feedback process, the supply would be modified to include the 500-meter passing lane.

Another type of feedback process would be for *sensitivity analysis*. The analyst would generally supply input modifications either manually or by mathematical formulation. For example, the previous two-lane rural road could be addressed to find the effect of the location of the passing lane from the bottom of the upgrade to the top of the upgrade.

The optimization of an objective function is another form of the feedback process. The analyst normally employs some type of mathematical programming technique. This technique may vary from rigid linear programming approach to a brute-force branch-and-bound approach.

A final feedback process is the investigation of traveller responses to the system changes. Initially, there is equilibrium between demand and supply. However, over time the input to the demand-supply analysis is modified due to demand growth, supply changes, and / or control

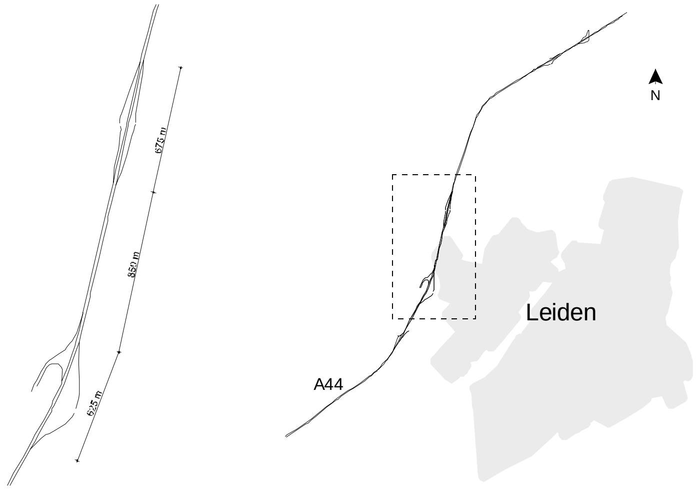

Figure 1.3: A44 near the Dutch city of Leiden

implementation. These input modifications change the predicted performance of the traffic system. The users then may respond by spatial, temporal, modal, or total travel changes. Users may change their paths of travel through the network, resulting in a spatial response. Users may change their time of travel, resulting in temporal responses. Users may change their mode of travel (car pool, bus, etc.), resulting in a model response. Finally, users may eliminate, combine, or create trips, resulting in total travel responses. As users respond in various ways, the input to the demand-supply analysis changes, and correspondingly, the predicted performance changes. This process of user responses and system performance changes becomes an iterative process and requires special procedures in order to reach equilibrium. An added complexity is that changes in travel patterns may modify the previously optimized control strategy. For instance, if a signal timing plan has been optimized, the system users may modify their paths of travel through the arterial network. Users will select cheaper cost routes, and flows change from some links to other links. If a sufficient number of users modify their paths through the network, the previously optimized signal timing plan may not longer be optimal.

1.4 Demand-supply analysis example

In this example, we aim to illustrate the application of the analytical framework for the demand-supply analysis for a motorway corridor. Since we have not yet discussed neither the knowledge, nor the analytical techniques that may be required to solve this problem, the approach is clearly oversimplified.

We will consider three situations: the reference case (or base case), the future situations without modifications (the so-called null-alternative) and an alternative treatment.

1.4.1 Base conditions

The considered roadway section is based on a part of a real motorway: the 2×2 motorway A44 near the Dutch city of Leiden (Fig. 1.3). The figure shows how two connections are situated very near each other. It was decided to consider only the Eastern roadway (see Fig. 1.4). The real roadway geometry has been used; the traffic demands are fictional.

Let us assume that the Table 1.1 describes the average traffic demands during the morning peak hour. We assume that there is a truck percentage of 10%. Furthermore, we assume that under ideal conditions, the capacity of a motorway lane equals 2000 veh/h. As we will see in the ensuing of this reader, the presence of (heavy) trucks will reduce the traffic supply (capacity), or equivalently, increase the traffic flow in person-car equivalents PCE. Assuming a PCE value

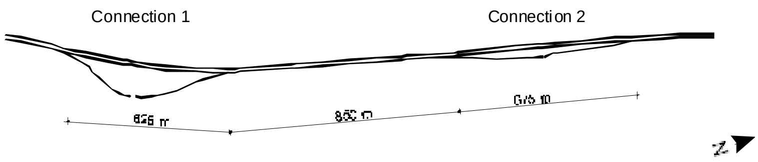

Figure 1.4: The Eastern roadway of the A44 near Leiden

	Destination		
Origin	Main road	Off-ramp 1	Off-ramp 2
Main road	2000	200	500
On-ramp 1	500	-	200
On-ramp 2	900	-	-

Table 1.1: Origin-destination demands (veh/h)

of 2 (which is rather high), the effective capacity per lane (expressed in veh/h) becomes 1900 veh/h.

1.4.2 Initial analysis

The first question that has to be answered is on which part of the motorway problems can be expected. This is the part where the volume-to-capacity ratio is highest (either on the main road, on an on-ramp or an off-ramp). It turns out that the traffic demand on the main road after the second connection equals 3400 veh/h. Since this is lower than the capacity (3800 veh/h), no real problems are expected. If traffic demands would increase equally for all origin-destination pairs, capacity problems are expected at the second on-ramp. The queue that would occur there would spill back, possibly reaching the off-ramp 2. In the chapter discussing shock-wave analysis, we will see that when this occurs, the speed at which the queue propagates increases, and the speed at which the queue spills back increases substantially. Even without shock-wave analysis, this can be understood, as will be seen in the remainder..

1.4.3 Feedback processes

We have now analyzed the reference case in which no problems are expected. In most cases, the traffic analyst would compare the modelling outcomes with empirical data (calibration and validation) to ensure that the representation of the current situation is indeed correct.

In the remainder of this example, we will consider several "what if" scenarios. First, let us assume that the traffic demands grow with 15% for all origin-destination pairs. In this case, the demands at the second connection will be 3910 veh/h (main road: 2875 veh/h; on-ramp 2: 1035), which is higher than the capacity of 3800 veh/h. From experience, we know that congestion mainly occurs on the main road, since the merging flow on the on-ramp appears to be able to merge without much delay. Thus, the 'effective capacity' that is available for the demand on the main road equals $3800 - 1035 = 2765$ (capacity bottle-neck – demand on-ramp 2). As a result, a queue starts building upstream of on-ramp 2, propagating upstream. The rate at which the number of vehicles in the queue increases equals $2875 - 2765 = 110$ veh/h. Assuming that vehicles in a queue take up approximately 10 m (per lane!), after 1 hour the length of the queue is $10 \cdot 110/2 = 550$ m. It will then be a matter of minutes before the queue spills back on the off-ramp. More precisely, the queue will spill back to the off-ramp when there will be $2 \cdot 675/10 = 135$ veh in the queue, which would take 135/110 hour. For that point onwards, problems will become more severe since now vehicles that do not have to pass the bottle-neck (on-ramp 2) will have to wait in the queue that is caused by it. How the delays

	Destination		
Origin	Main road	Off-ramp 1	Off-ramp 2
Main road	2000	900	500
On-ramp 1	500	-	1000
On-ramp 2	1400	-	-

Table 1.2: Origin-destination demands (veh/h) for future conditions

of with destination 'main road' relate to the delays of vehicles with destination 'off-ramp 2' compare is not obvious, since it may depend on complicated microscopic processes. However, if we assume that the speeds are equals irrespective of the destination, all vehicles will occur the same delay before arriving at off-ramp 2. How can this problem be solved?

Remark 1 *This situation has been analyzed with the microscopic simulation program FOSIM. It turns out that capacity problems will not always occur at on-ramp 2, but also at the motorway section between on-ramp 1 and off-ramp 2. From the micro-simulation it can be concluded that the inefficiency of the merging process as well as the use of the motorway lanes are the main cause for these problems. Vehicles with destination off-ramp 2 will merge to the right lane before off-ramp 1, causing traffic demands on the lane - on which the vehicles entering via on-ramp 1 need to merge - to exceed the capacity of the lane. In other words, the effective capacity of the merge is lower than the 3800 veh/h due to merging effects.*

Let us now assume that near connection 2, a new residential area is build. As a result, traffic demands on the on-ramp increase during the morning peak hour, and become 500 veh/h. At the same time, near connection 1 a business area is developed. It turns out that the employees working there mostly briefly use the A44, after which they use the underlying network. As a consequence, the traffic demands will increase significantly. The traffic demands from on-ramp 1 to off-ramp 2 increases to 800 veh/h. During the morning peak hour, the business area also attracts a lot of traffic. The traffic demands increases with 700 veh/h.

The highest traffic demand is now attained at the motorway right after the first on-ramp. This means that the upstream section will possibly run into congestion. The total demand equals 4000 veh/h, which is higher than the estimated capacity. Hence, we may expect that congestion will occur on this location. Congestion will mainly occur on the main road (which is usually the case, since merging traffic generally finds their way onto the main road). Then, the occurring queue will spill back upstream, possibly over the first off-ramp.

The number of vehicles that are stored depend on the duration of the oversaturated conditions. Assuming that all merging vehicles are able to flow onto the main road, yields that the remaining capacity for the traffic on the main road equals $3800 - 1500 = 2300$ veh/h. Since the demand equals 2500 veh/h, the queue will grow at a rate of 200 veh/h. Assuming that a vehicle in a queue requires 8 m of roadway space – in practise, more space will be needed since the vehicles in the queue are driving – a queue of 200 veh will have a length of $200{\cdot}8/2{=}800$ m. Hence, the queue will spill back to the off-ramp 1 with an hour. The chapter on queuing analysis will go into detail on how to compute the dynamics of the queue.

One might think that not only the first on-ramp, but also the second on-ramp will pose a problem. According to the traffic demands, the total demand at the second on-ramp equations 3900 veh/h, which clearly also exceeds the capacity. However, since the actual flow at the section after the first on-ramp is less than the traffic demand (namely equal to the capacity of 3800 veh/h), the actual downstream demand will be less than 3800 veh/h, and no congestion will occur.

1.4.4 Ramp control

A possible solution to solve the foreseen queuing problems is the use of ramp-control (ramp-metering). Assume that the objective of the control is to keep traffic on the main road from becoming congested. For the first situation, the flow on on-ramp 2 that is allowed to enter the main road equals 925 veh/h, and hence $1035 - 925 = 110$ veh/h need to be held back. The ramp-controller will thus let 925 veh/h merge onto the main road. We can now compute the queue lengths on on-ramp 2 after 1 hour. Note that since the on-ramp is a one-lane road, it has less storage capacity; the physical length of the queue will increase two times as quick.

For the second example (new residential area), a similar ramp-metering strategy can be proposed. In this case however, traffic on the metered on-ramp 1 may decide to use an alternative route, e.g. use the underlying network and drive to the second on-ramp (rat-running). When this occurs, on-ramp 2 will become oversaturated and congestion will occur there. In that case, it will also be prudent to meter the second on-ramp.

Part II

Traffic Flow Characteristics