Part A Transportation Systems Analysis and Modeling Process

Chapter 1

Introduction to Transportation Systems

1.1 Definitions

1.1.1 Preliminary Concepts of Transportation Systems

Transportation systems consist not only of the physical and organizational elements that interact with each other to produce transportation opportunities, but also of the demand that takes advantage of such opportunities to travel from one place to another. This travel demand, in turn, is the result of interactions among the various economic and social activities located in a given area. Mathematical models of transportation systems represent, for a real or hypothetical transportation system, the demand flows, the functioning of the physical and organizational elements, the interactions between them, and their effects on the external world. Mathematical models and the methods involved in their application to real, largescale systems are thus fundamental tools for evaluating and/or designing actions affecting the physical elements (e.g., a new railway) and/or organizational components (e.g., a new timetable) of transportation systems.

A transportation system can be defined as a set of elements and the interactions between them that produce both the demand for travel within a given area and the provision of transportation services to satisfy this demand. Almost all of the components of a social and economic system in a given geographical area interact at some level of intensity. However, in practice it is impossible to take into account every interacting element when addressing a given transportation engineering problem. The general approach of systems engineering is to isolate the elements most relevant to a problem at hand, and to group these elements and the relationships between them within the analysis system. The remaining elements are assigned to the external environment; they are taken into account only in terms of their interactions with the analysis system. In general, the analysis system includes the elements and interactions that an action under consideration may significantly affect. Hence there is a strong interdependence between the identification of the analysis system and the problem to be solved. The transportation system of a given area can also be seen as a subsystem of a wider territorial system with which it strongly interacts. The details of the specific problem determine the extent to which these interactions are included either in the analysis system or the external environment.

1.1.2 Components of Transportation Systems

An urban area could be considered as consisting of a set of households, workplaces, services, transportation facilities, government organizations, regulations, and so on. This system has a hierarchical structure and, within it, several subsystems can be identified (Figure 1.1).

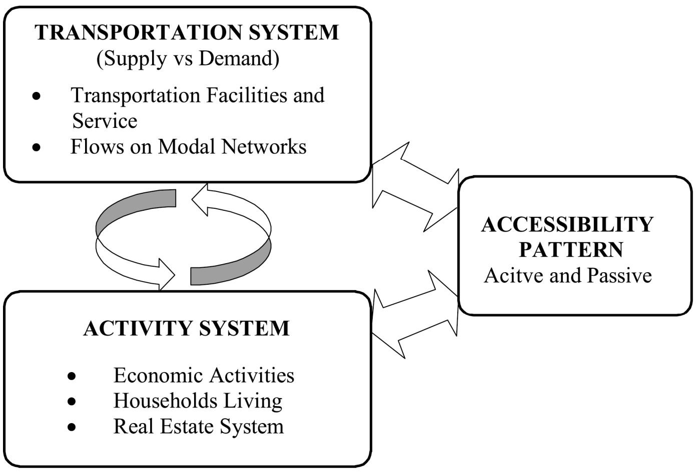


Figure 1.1 Relationships Between the Transportation System and The Activity System

One of the subsystems – **the activity system** – represents the set of individual, social, and economic behaviors and interactions that give rise to travel demand. To describe the geographic distribution of activity system features, the urban area is typically subdivided into geographic units called zones. The activity system can be further broken down into three subsystems consisting of:

- The households living in each zone, categorized by factors such as income level, lifecycle, composition, and the like.
- The economic activities located in each zone, categorized by a variety of socioeconomic indicators (e.g., sector of activity; value added; number of employees).
- The real estate system, characterized by the floor space available in each zone for various uses (industrial production, offices, building areas, etc.) and the associated market prices.

The different components of the activity system interact in many ways. For example, the number and types of households living in the various zones depend in part on employment opportunities and their distribution, and therefore on the economic activity subsystem. Furthermore, the location of some types of economic activities (retail, social services such as education and welfare, etc.) depends on the geographic distribution of the households. Finally, the number of households and the intensity of economic activities in each zone depend on the availability of specific types of floor space (houses, shops, etc.) and on their relative prices.

Another subsystem – **the transportation system** – consists of two main components: demand and supply.

Travel demand derives from the need to access urban functions and services in different places and is determined by the distribution of households and activities in the area. Household members make longterm “mobility choices” (holding a driving license, owning a car, etc.) and shortterm “travel choices” (trip frequency, time, destination, mode, path, etc.), and use the transportation network and services so that they can undertake different activities (work, study, shopping, etc.) in different locations. These choices result in travel demand flows, that is, the trips made by people between the different zones of the city, for different purposes, in different periods of the day, by means of the different available transportation modes. Similarly, economic activities require the transportation of goods that are consumed by other activities or by households. Goods are moved between production plants, retail locations, and houses or other "final consumption" sites. Their movements make up freight travel demand and corresponding flows.

Both mobility and travel choices are influenced by the characteristics of the transportation services offered by the available travel modes (such as private vehicles, transit, walking). These characteristics are known as level of service or performance attributes; they include travel times, monetary costs, service reliability, riding comfort, and the like. For instance, the choice of destination may be influenced by the travel time and cost needed to reach each alternative destination; the choice of departure time depends on the travel time to the destination and the desired arrival time; and the choice of transportation mode is influenced by the time, cost and reliability of the available modes.

The **transportation supply** component is made up of the facilities (roads, parking spaces, railway lines, etc.), services (transit lines and timetables), regulations (road circulation and parking regulations), and prices (transit fares, parking prices, road tolls, etc.) that produce travel opportunities. Travel from one location to another frequently involves the successive use of several connected facilities or services. Transportation facilities generally have a finite capacity, that is, a maximum number of units that may use them in a given time interval. Transportation facilities also generally exhibit congestion; that is, the number of their users in a time unit affects their performance. When the flow approaches the capacity of a given facility (e.g., a road section), interactions among users significantly increase and congestion effects can become important. Congestion on a facility can significantly affect the level of service received by its users; for example, travel time, service delay, and fuel consumption all increase with the level of congestion.

Finally, the performance of the transportation system influences the relative accessibility of different zones of the urban area by determining, for each zone, the generalized cost (disutility) of reaching other zones (active accessibility), or of being reached from other zones (passive accessibility). As has been noted, both these types of accessibilities influence the location of households and economic activities and ultimately the real estate market. For example, in choosing their residence zone, households take account of active accessibility to the workplace and other services (commerce, education, etc.). Similarly, economic activities are located to take into account passive accessibility on behalf of their potential clients; public services should be located to allow for passive accessibility by their users, and so on.

1.1.3 Relationship Between Transportation Systems and Activity Systems

Several feedback cycles can be identified in an urban transportation system. These are cycles of interdependence between the various elements and subsystems, as shown in Figure 1.1. The **innermost cycle**, the one that involves the least number of elements and that usually shows the shortest reaction time to perturbations, is the interaction between facility flows, the performance due to congestion and transportation costs, in particular those connected with road transportation. The trips made by a given mode (e.g., car) choose from among the available paths and use traffic elements of the transportation network (e.g., road sections). Due to congestion, these flows affect the level of service on the different paths and so, in turn, influence user path choices.

There are also **outer cycles**, cycles that influence multiple choice dimensions and that involve changes occurring over longer time periods. These cycles affect the split of trips among the alternative modes and the distribution of these trips among the possible destinations. Finally, there are cycles spanning even longer time spans, in which interactions between activity location choices and travel demand are important. Again, through congestion, travel demand influences accessibility of the different areas of the city and hence the location choices of households and firms.

It is clear from the above that a transportation system is a complex system, that is, a system made up of multiple elements with nonlinear interactions and multiple feedback cycles. Furthermore, the inherent unpredictability of many features of the system, such as the time needed to traverse a road section or the particular choice made by a user, may require the system state to be represented by random variables.

As a first approximation, these random variables are often represented by their expected values.

Transportation systems engineering has traditionally focused on modeling and analysis of the elements and relationships that make up the transportation system, considering the activity system as exogenously given. More specifically, it has typically considered the influence of the activity system on the transportation system (in particular on travel demand), whereas the inverse influence of accessibility on activity location and level has usually been neglected. However, this divide is rapidly vanishing and transportation system analysis increasingly studies the whole activity– transportation system, albeit at different levels of detail than do disciplines such as regional science and spatial economics.

If the problem at hand is longterm planning of the whole urban transportation system, including the construction of new motorways, railway lines, parking facilities, and the like, the analysis has to include the entire multimode transportation system and possibly its relationships with the urban activity system. Indeed, the resulting modifications in the transportation network and service performance characteristics and the time needed to implement the plan are such that all components of the transportation and activity systems will likely be affected.

There are cases, however, in which the problem is more limited. If, for example, the aim is to design the service characteristics of an urban transit system without building new facilities (and without implementing new policies affecting other modes, such as car use restrictions), it is common practice to include in the analysis system only those elements (demand, services, prices, vehicles, etc.) related to public transportation. The rest of the transportation system is included in the external environment interacting with the public transportation system.

1.2 Transportation System Identification

1.2.1 Relevant Spatial Dimensions

The identification of relevant spatial dimensions consists of three phases:

- Definition of the study area
- Subdivision of the area into traffic zones (zoning)
- Identification of the basic network

Study Area

This phase delineates the geographical area that includes the transportation system under analysis and encompasses most of the project effects. First, the analyst must consider the decision-making context and the type of relevant trips: commuting, leisure, and so on. Most trips of interest should have their origin and destination inside the study area. Similarly, the study area should include transportation facilities and services that are likely to be affected by the transportation project. As one example, the study area for a new traffic scheme should include possible alternative roads for rerouting; as another, the study area for a new infrastructure project should include locations where the number of trips starting or ending may change due to variations in accessibility. The limit of the study area is the area boundary. Outside this boundary is the external area, which is only considered through its connections with the analysis system. For instance, the study area might be a whole country if the transportation project is at a national level; alternatively, it may be a specific urban area, or part of an urban area for a traffic management project.

Zoning

In principle, the trips undertaken in a given area may start and end at a large number of points. To model the system, it is necessary to subdivide the study area (and possibly portions of the external area) into a number of discrete geographic units called traffic analysis zones (TAZs). Trips between two different traffic zones are known as interzonal trips, whereas intrazonal trips are those that start and end within the same zone.

In most transportation models, all trips that start or end within a zone are represented as if their terminal points were at a single fictitious node called the zone centroid, located in the zone near the geographic “center of gravity” of the full set of actual trip terminal points that it represents. In this representation, intrazonal trips both start and end at the same centroid location, so their effects on the network cannot be modeled.

Zoning can have different levels of detail, that is, a coarser or finer grain. For example, traffic zones may consist of entire cities or groups of cities in a regional or national model, or of one or a few blocks in urban traffic model.

For a given model, the density of zoning should approximately correspond to the density of the relevant network elements: a denser set of network elements corresponds to a finer zoning and vice versa (Figure 1.2). For example, if the urban system includes public transportation, it is common practice to consider smaller traffic zones than for a system including only individual cars. This allows walking access to transit stops and/or stations to be realistically represented in terms of the distance from the zone centroid.

The external area is usually subdivided into larger traffic zones. External zones represent trips that use the study area’s transportation system but start or end outside of the study area itself. External zones are also represented by zone centroids sometimes called stations.

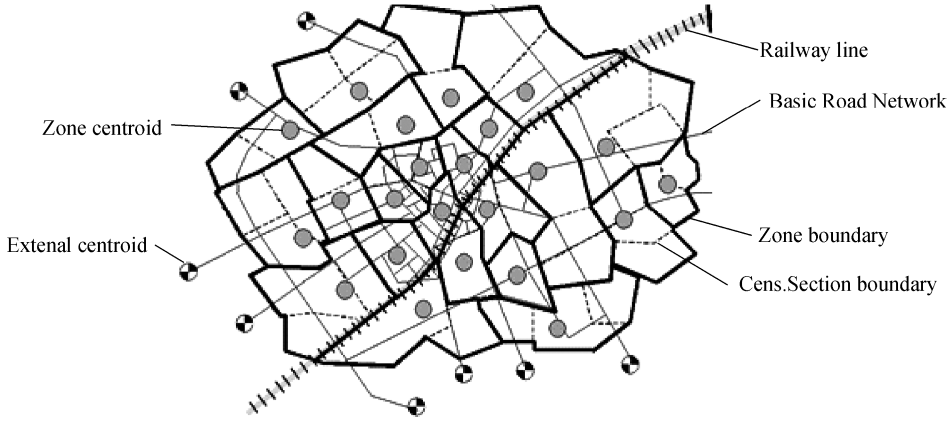


Figure 1.2 Zoning and Basic Network

For a given study area and analysis problem, there may be several possible zoning systems. However, some general guidelines are usually followed.

- Physical geographic separators (e.g., rivers, railway lines, etc.) are conventionally used as zone boundaries because they prevent "diffuse" connections between adjacent areas and therefore usually imply different access conditions to transportation facilities and services.

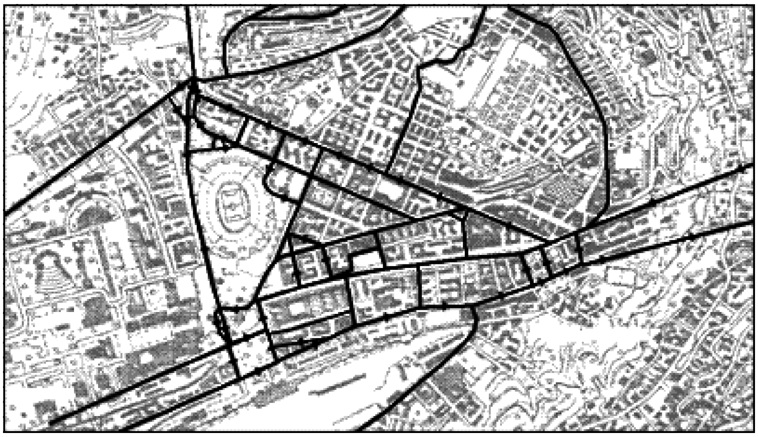


Figure 1.3 Basic road Network for a Portion of Urban Area

- Traffic zones are often defined as aggregations of official administrative areas (e.g., census geographic units, municipalities, or provinces). This allows each zone to be associated with the statistical data (population, employment, etc.) usually available for such areas.

- A different level of zoning detail may be adopted for different parts of the study area depending on the precision needed. For example, smaller zones may be used in the vicinity of a specific facility (e.g., a new road, railway, etc.) for which traffic flows and impacts must be predicted more precisely.

- A traffic zone should group connected portions of the study area that are relatively homogeneous with respect both to their land use (e.g., residential or commercial uses in urban areas; industrial or rural uses in outlying areas) and to their accessibility to transportation facilities and services.

Basic Network

The set of physical elements represented for a given application is called the basic network. For example, in urban road systems, the road sections and their main traffic regulations such as one-way, no turn, and the like are indicated (Figure 1.3). For scheduled service systems, the infrastructure over which the service is operated (road sections, railways, etc.) will be indicated, together with the main stops or stations, the lines operating along the physical sections, and so on.

The facilities and services included in the network might relate to one or to several transportation modes. The former is referred to as a single mode system and the latter as a multimodal system.

Relevant facilities and services are identified based on their role in connecting the traffic zones in the study area and the external zones. This implies a close interdependence between the identification of the basic network and zone systems. Facilities and services may also be included according to their relationship to the transportation alternatives under consideration.

Because the flows on network elements resulting from intrazonal trips are not modeled, very fine zoning with a coarse basic network will probably cause overestimation of the traffic flows on the included network elements. Conversely, a very detailed basic network with coarse zoning may lead to underestimation of some traffic flows.

Identification of the relevant elements is obviously easier when all the services and facilities play a role in connecting traffic zones, as may be the case, for example, for a national **airways** network. In the case of road networks, only a subset of roads is relevant in connecting the different zones. In urban areas, for example, local roads are usually excluded from the basic network of the whole area, although they may be included in the basic networks of spatially limited subsystems (a neighborhood or part of it). Similarly, when dealing with a whole region, most of the roads within each city will not be included in the basic network.

1.2.2 Relevant Temporal Dimensions

A transportation system operates and evolves over time, with the characteristics of both travel demand and supply varying at different time scales. For example, the number of trips undertaken in an urban area and the frequency of transit services vary by time of day, by day of the week, and so on. Although space has always been recognized as a fundamental dimension of transportation systems, the time dimension has often been overlooked. However, determinations of the relevant analysis time intervals as well as assumptions about system variability within those intervals are crucial modeling decisions.

Design and evaluation of transportation projects typically involve two distinct time scales. Design (e.g., determining the required number of road lanes, the settings of a traffic signal at an intersection or the service frequency of a transit line) usually requires information on short maximum-load periods such as the peak hour. This information is obtained from a transportation model by analyzing conditions in a particular reference or model period. On the other hand, economic or financial evaluations usually require information about a project's performance over a time span comparable to its technical life. The analysis period is the entire time duration relevant to the study of a given system.

Depending on the application, the analysis period may include one or more model periods. For major infrastructure projects, for example, the analysis period may span several years or even decades, but the system is typically modeled for only a limited number of reference periods (e.g., one average day per year); the results obtained for the model periods are then expanded to the whole analysis period. By contrast, applications such as traffic signal setting, for example, may only require the modeling of a single reference period (e.g., the A.M. peak period on an average weekday).

If both demand and supply remained approximately constant over the whole analysis period, then any shorter interval could be adopted as a reference period, and the results obtained from modeling the reference period could validly be extrapolated to the whole analysis period. However, because transportation system characteristics change over time, a selected reference period will only be representative of a portion of the analysis period. Thus, the latter is typically subdivided into several model periods, corresponding to different representative situations. Figure 1.4 shows the variation of urban travel demand by trip purpose within an average weekday. In this case, inasmuch as the hypothesis of constancy within the day would clearly be unrealistic, the day would typically be subdivided into shorter model periods (e.g., morning peak, off peak, evening peak).

One approach is to assume that all relevant transportation characteristics are constant on average during the reference period, and independent of the particular instant at which they are modeled: this is the assumption of within-period stationarity. Traditional mathematical models of transportation systems assume that demand and supply remain constant over a period of time long enough to allow the system to reach a stationary or steady-state condition.

The other approach explicitly models the variations in demand and supply within the reference period; this is the assumption of within-period dynamics. It should be noted that, in practice, within-period dynamic models typically assume that some elements of the system (e.g., activity-system variables or global travel demand) remain constant within the model period.

In general, three kinds of time variations of system characteristics are important.

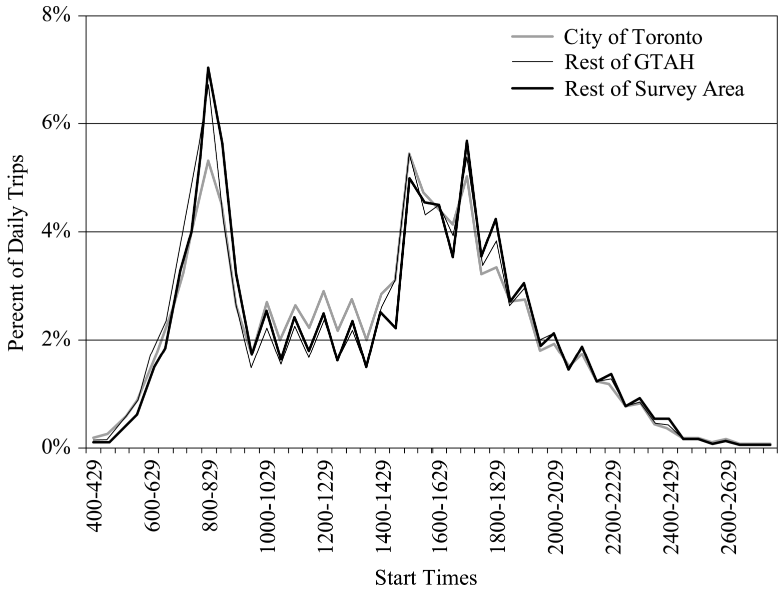
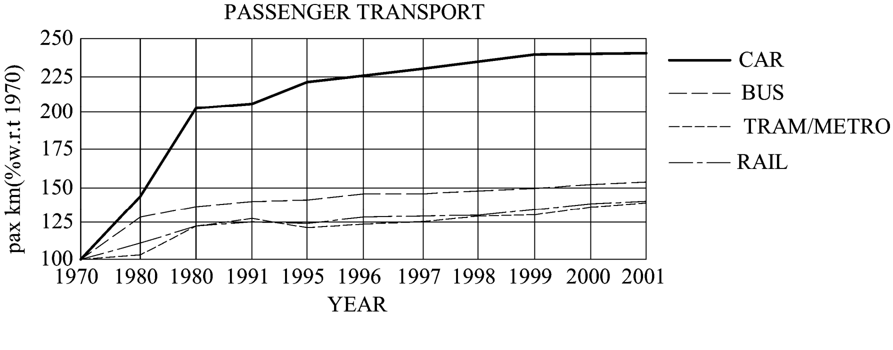


Figure 1.4 Breakdown of Urban Travel Demand by Time of Day and Purpose

(a) **Long-term** variations or trends at the global level and/or systematic variations that can be identified by averaging over multiple reference periods. For example, if reference intervals are single days, a trend consists of variations in the total level and/or in the structure of the average annual demand, observed over several years. In this case, the daily demand is averaged over 365 elementary periods. **Long-period** variations are often the result of structural changes in the socioeconomic variables underlying travel demand, or in transp ortation supply. For example, variations in the level of economic activity, production technologies, household income, individual vehicle ownership, sociodemographic population characteristics, lifestyles, urban migration, and the stock of transportation facilities and services have significantly modified the level and structure of passenger and freight travel demand over the years (Figure 1.5).



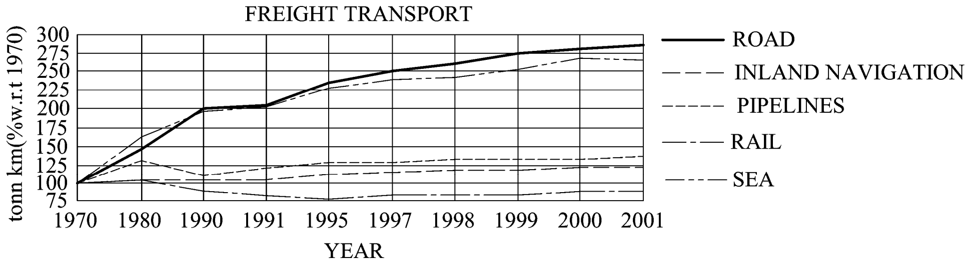


Figure 1.5 Average Long-Term Trends in European Passenger and Freight Demand

(b) Cyclical (seasonal) variations occurring within the analysis period and involving several reference periods. These variations repeat themselves cyclically and can be observed by averaging over a number of cycles. This is the case, for example, with variations in daily demand on different days of the week, or with variations at different times within a typical day. For instance, the fluctuations of urban travel demand by time of day, shown in Figure 1.4, repeat cyclically over successive workdays. In an analysis period, several cyclic variations with different cycle lengths may occur and overlap with long-term variations. For example, demand and supply change over an analysis period of several years (long-term variation), but they also vary cyclically over the different months of the year, the days of the week and the hours of each day.

(c) Between-period variations are variations in demand and supply over reference periods with otherwise identical characteristics, after accounting for the trend and cyclic variations. This is the case with demand variations during morning peak hours of different days with similar characteristics. These fluctuations can be considered random because they cannot be associated with specific events. Travel demand results from the choices made by a large number of decision makers; its actual value in a period therefore depends both on the unpredictable behavioral elements connected with these choices, and on the influence of choices made in previous periods. Similarly, the actual values of some key supply parameters, such as road capacities or travel times, may vary due to unpredictable events, such as an accident. Variations in demand and supply between successive reference periods, for example, the same hours within typical days, are called between-period (or period-to-period) dynamics.

As already mentioned, in reality the three types of dynamics overlap and their identification depends to a great extent on the perspective adopted. In addition, the length of the reference period depends on the modeling approach followed. Some models can endogenously represent the variations in relevant parameters within a typical day, which in this case may be taken as the model period. Other models may require the analyst to explicitly specify different exogenous input variables in order to represent variations over different reference periods of the day; in this case, single hours may be the best model periods. Moreover, different applications usually require different assumptions on the relevant temporal dimensions.

Consider, for example, a freight system project for which no significant congestion is expected. This project might require an analysis period several years long. Furthermore, it might be appropriate to consider longterm variations of the system over a number of years, and to account for seasonal variations by considering one or a few typical months as model periods, while ignoring cyclic variations within each month.

For a project with a shortterm horizon, such as the traffic plan of an urban area, the longterm trend of daily demand (say over several years) can be ignored. The analysis period might consist of one or more typical days (e.g., average week and weekend days).

Cyclic variations could be modeled as hourly variations within the typical day. Model periods may encompass the morning and evening peak and off-peak hours, with traffic conditions during each period assumed to be stationary. Alternatively, the analyst may consider a different perspective, by which the analysis period is an entire week, cyclic variations are relative to both days of the week and hours of the day, and reference periods encompass full days. In this case, the models would explicitly represent the distribution of demand and supply performances over subintervals of each day, following a **within-period** dynamic approach (Figure 1.6).

1.2.3 Relevant Components of Travel Demand

Passengers and goods moving in a given area demand the transportation services supplied by the system. Travel demand clearly plays a central role in the analysis and modeling of transportation systems because most transportation projects attempt to satisfy this demand (although some projects, such as travel-demand management policies, attempt instead to modify some of its characteristics). In turn, traveler choices can significantly affect the performance of transportation supply elements through congestion.

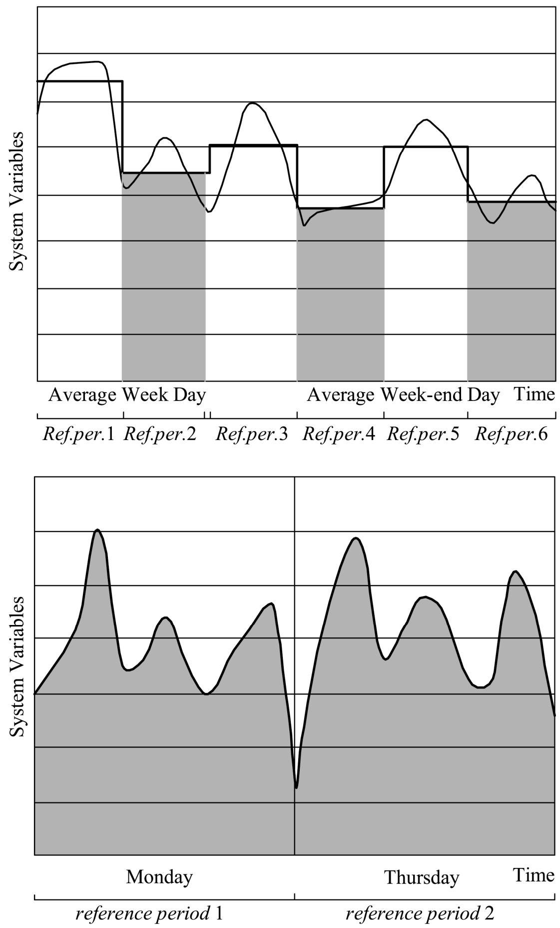


Figure 1.6 Alternative Reference Periods

Travel does not generally provide utility in itself, but is rather an auxiliary activity necessary for other activities carried out in different locations. Travelers make work, school, and shopping-related trips. Goods are shipped from production sites to markets. Travel demand is therefore a derived demand, the result of the interactions between the activity system and the transportation services and facilities, as well as of the habits underlying travel behavior in a given area.

A travel-demand flow can formally be defined as the number of users with given characteristics consuming particular services offered by a transportation system in a given time period. It is clear that travel demand flows result from the aggregation of individual trips made in the study area during the reference period. A trip is defined as the act of moving from one place (origin) to another (destination) using one or more modes of transportation, in order to carry out one or more activities. A sequence of trips, following each other in such a way that the destination of one trip coincides with the origin of the next, is referred to as a journey or trip chain. With passenger travel, trip chains usually start and end at home; for example, a home-work-shopping-home chain consists of three distinct trips. For freight, individual movements of goods from one place to another are usually referred to as shipments or consignments. The sequence of manipulations (e.g., packaging) and storage activities applied to shipments is often referred to as the logistic or supply chain.

Transportation system users, and the trips they undertake, can be characterized in a variety of ways in addition to the temporal characterization described in the previous section. In the following chapters, h stands for the reference period, describing the average weekday, the morning or evening peak hours, the winter or summer seasons, and so on. Some of these ways are described here.

The spatial characterization of trips is made by grouping them by place (zone or centroid) of origin and destination, and demand flows can be arranged in tables, called origin-destination matrices (OD matrices), whose rows and columns correspond to the different origin and destination zones, respectively (Figure 1.7). Matrix entry gives the number of trips made in the reference period from origin zone o to destination zone d (the OD flow). Some aggregations of the OD matrix elements are also useful. The sum of the elements of row o:

$$d_{o*} = \sum_d d_{od} \tag{1.1}$$

accumulates the total number of trips leaving zone o in the reference period and is known as the flow produced or generated by zone o. The sum of the elements of column d accumulates the number of trips arriving in zone d in the reference period:

$$d_{*d} = \sum_o d_{od} \tag{1.2}$$

and is known as the flow attracted by zone d . The total number of trips made in the study area in the reference interval is indicated by d ..:

$$d_{**} = \sum_o \sum_d d_{od} \tag{1.3}$$

Trips can be characterized by whether their endpoints are located within or outside of the study area. For internal (II) trips, the origin and the destination are both within the study area. For exchange (IE or EI) trips, the origin is within the study area and the destination outside, or vice versa. Finally, crossing (EE) trips have both their origin and their destination external to the study area, but traverse the study area, that is, use the transportation system under study. Figure 1.7 is a schematic representation of the three types of trips and their position in the OD matrix.

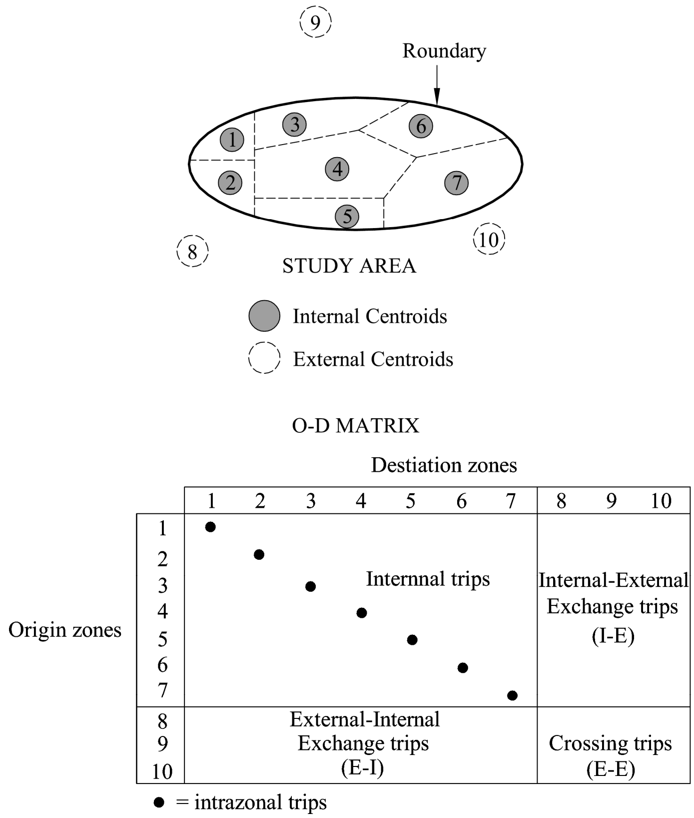


Figure 1.7 Trip Types and Their Identification in the Origin–Destination Matrix

Travel demand can also be classified in terms of user and trip characteristics. In the case of person trips, user characteristics of interest usually relate to the trip-maker’s socioeconomic attributes, such as income level or possession of a driver’s license. Groups of users who are homogeneous with respect to a particular set of socioeconomic characteristics are referred to as market segments. In a study of different pricing policies, for example, market segments might be defined according to personal or household income. In the case of goods movements, the user characteristics of interest typically relate to attributes of the shipping firm, such as sector of economic activity, firm size, type of plant, production cycle, and so on. In the following chapters, market segments are indicated by i.

Characteristics of individual trips are also of interest. Person trips are often described in terms of the general activities carried out at the origin and destination ends. The pair of activities defines the trip purpose: home-based work trips, work-based shopping trips, and so on. A whole sequence of purposes (activities) can be associated with a trip chain. The trip purpose is indicated by s.

Other trip characteristics of interest in a particular analysis may include desired arrival or departure times, and mode, among others, for person trips; and consignment size, type of goods (time sensitivity, value, etc.) and mode for freight trips.