

Fundamentals of Transportation

Fundamentals of Transportation

Fundamentals of Transportation

- /About/
- /Introduction/
- /Economics/
- /Geography and Networks/

/Planning/

- /Trip Generation/
- /Destination Choice/
- /Mode Choice/
- /Route Choice/
- /Evaluation/

/Operations/

- /Queueing/
- /Traffic Flow/
- /Queueing and Traffic Flow/
- /Shockwaves/
- /Traffic Signals/

/Design/

- /Sight Distance/
- /Grade/
- /Earthwork/
- /Horizontal Curves/
- /Vertical Curves/

Other Topics

- /Pricing/
- /Conclusions/
- /Analogs/
- /Decision Making/

Fundamentals of Transportation/ About

This book is aimed at undergraduate civil engineering students, though the material may provide a useful review for practitioners and graduate students in transportation. Typically, this would be for an **Introduction to Transportation** course, which might be taken by most students in their sophomore or junior year. Often this is the first engineering course students take, which requires a switch in thinking from simply solving given problems to formulating the problem mathematically before solving it, i.e. from straight-forward calculation often found in undergraduate Calculus to vaguer word problems more reflective of the real world.

How an idea becomes a road

The plot of this textbook can be thought of as "How an idea becomes a road". The book begins with the generation of ideas. This is followed by the analysis of ideas, first determining the origin and destination of a transportation facility (usually a road), then the required width of the facility to accommodate demand, and finally the design of the road in terms of curvature. As such the book is divided into three main parts: planning, operations, and design, which correspond to the three main sets of practitioners within the transportation engineering community: transportation planners, traffic engineers, and highway engineers. Other topics, such as pavement design, and bridge design, are beyond the scope of this work. Similarly transit operations and railway engineering are also large topics beyond the scope of this book.

Each page is roughly the notes from one fifty-minute lecture.

Authors

Authors of this book include David Levinson ^[1], Henry Liu ^[2], William Garrison ^[3], Adam Danczyk, Michael Corbett

References

- [1] <http://nexus.umn.edu>
 - [2] <http://www.ce.umn.edu/~liu/>
 - [3] [http://en.wikipedia.org/wiki/William_Garrison_\(geographer\)](http://en.wikipedia.org/wiki/William_Garrison_(geographer))
-

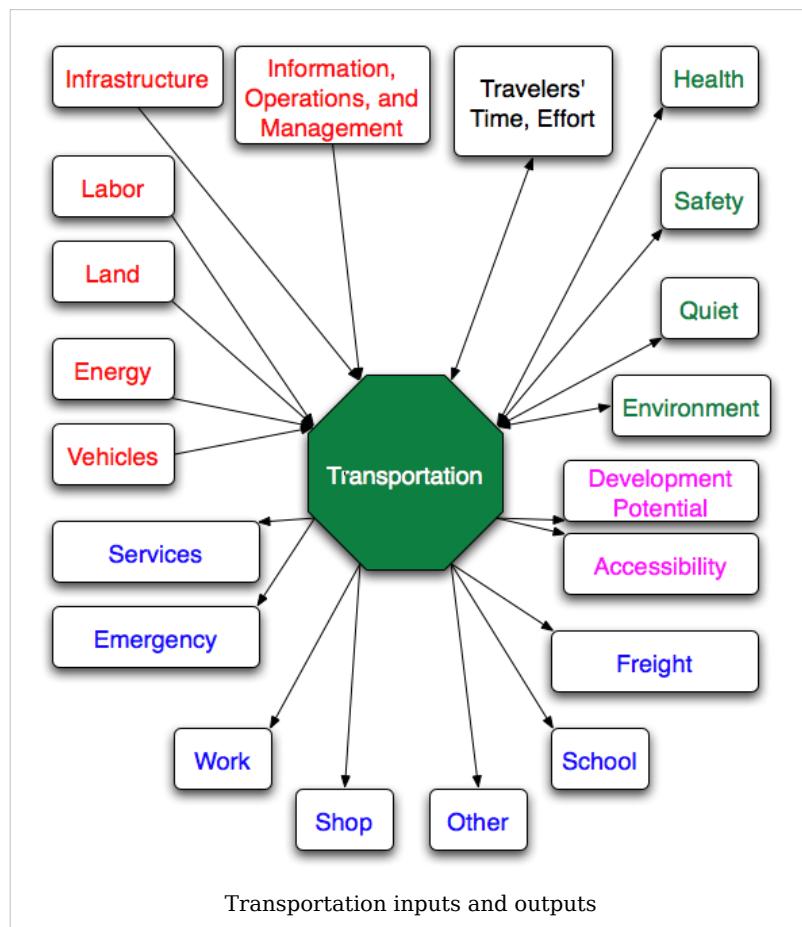
Fundamentals of Transportation/ Introduction

Transportation moves people and goods from one place to another using a variety of vehicles across different infrastructure systems. It does this using not only technology (namely vehicles, energy, and infrastructure), but also people's time and effort; producing not only the desired outputs of passenger trips and freight shipments, but also adverse outcomes such as air pollution, noise, congestion, crashes, injuries, and fatalities.

Figure 1 illustrates the inputs, outputs, and outcomes of transportation. In the upper left are traditional inputs (infrastructure (including pavements, bridges, etc.), labor required to produce transportation, land consumed by infrastructure, energy inputs, and vehicles). Infrastructure is the traditional preserve of civil engineering, while vehicles are anchored in mechanical engineering. Energy, to the extent it is powering existing vehicles is a mechanical engineering question, but the design of systems to reduce or minimize energy consumption require thinking beyond traditional disciplinary boundaries.

On the top of the figure are Information, Operations, and Management, and Travelers' Time and Effort. Transportation systems serve people, and are created by people, both the system owners and operators, who run, manage, and maintain the system and travelers who use it. Travelers' time depends both on freeflow time, which is a product of the infrastructure design and on delay due to congestion, which is an interaction of system capacity and its use. On the upper right side of the figure are the adverse outcomes of transportation, in particular its negative externalities:

- by polluting, systems consume health and increase morbidity and mortality;
- by being dangerous, they consume safety and produce injuries and fatalities;
- by being loud they consume quiet and produce noise (decreasing quality of life and property values); and
- by emitting carbon and other pollutants, they harm the environment.



All of these factors are increasingly being recognized as costs of transportation, but the most notable are the environmental effects, particularly with concerns about global climate change. The bottom of the figure shows the outputs of transportation. Transportation is central to economic activity and to people's lives, it enables them to engage in work, attend school, shop for food and other goods, and participate in all of the activities that comprise human existence. More transportation, by increasing accessibility to more destinations, enables people to better meet their personal objectives, but entails higher costs both individually and socially. While the "transportation problem" is often posed in terms of congestion, that delay is but one cost of a system that has many costs and even more benefits. Further, by changing accessibility, transportation gives shape to the development of land.

Modalism and Intermodalism

Transportation is often divided into infrastructure modes: e.g. highway, rail, water, pipeline and air. These can be further divided. Highways include different vehicle types: cars, buses, trucks, motorcycles, bicycles, and pedestrians. Transportation can be further separated into freight and passenger, and urban and inter-city. Passenger transportation is divided in public (or mass) transit (bus, rail, commercial air) and private transportation (car, taxi, general aviation).

These modes of course intersect and interconnect. At-grade crossings of railroads and highways, inter-modal transfer facilities (ports, airports, terminals, stations).

Different combinations of modes are often used on the same trip. I may walk to my car, drive to a parking lot, walk to a shuttle bus, ride the shuttle bus to a stop near my building, and walk into the building where I take an elevator.

Transportation is usually considered to be between buildings (or from one address to another), although many of the same concepts apply within buildings. The operations of an elevator and bus have a lot in common, as do a forklift in a warehouse and a crane at a port.

Motivation

Transportation engineering is usually taken by undergraduate Civil Engineering students. Not all aim to become transportation professionals, though some do. Loosely, students in this course may consider themselves in one of two categories: Students who intend to specialize in transportation (or are considering it), and students who don't. The remainder of civil engineering often divides into two groups: "Wet" and "Dry". Wets include those studying water resources, hydrology, and environmental engineering, Drys are those involved in structures and geotechnical engineering.

Transportation students

Transportation students have an obvious motivation in the course above and beyond the fact that it is required for graduation. Transportation Engineering is a pre-requisite to further study of Highway Design, Traffic Engineering, Transportation Policy and Planning, and Transportation Materials. It is our hope, that by the end of the semester, many of you will consider yourselves Transportation Students. However not all will.

"Wet Students"

I am studying Environmental Engineering or Water Resources, why should I care about Transportation Engineering?

Transportation systems have major environmental impacts (air, land, water), both in their construction and utilization. By understanding how transportation systems are designed and operate, those impacts can be measured, managed, and mitigated.

"Dry Students"

I am studying Structures or Geomechanics, why should I care about Transportation Engineering?

Transportation systems are huge structures of themselves, with very specialized needs and constraints. Only by understanding the systems can the structures (bridges, footings, pavements) be properly designed. Vehicle traffic is the dynamic structural load on these structures.

Citizens and Taxpayers

Everyone participates in society and uses transportation systems. Almost everyone complains about transportation systems. In developed countries you seldom here similar levels of complaints about water quality or bridges falling down. Why do transportation systems engender such complaints, why do they fail on a daily basis? Are transportation engineers just incompetent? Or is something more fundamental going on?

By understanding the systems as citizens, you can work toward their improvement. Or at least you can entertain your friends at parties

Goal

It is often said that the goal of Transportation Engineering is "The Safe and Efficient Movement of People and Goods."

But that goal (safe and efficient movement of people and goods) doesn't answer:

Who, What, When, Where, How, Why?

Overview

This wikibook is broken into 3 major units

- Transportation Planning: Forecasting, determining needs and standards.
- Traffic Engineering (Operations): Queueing, Traffic Flow Highway Capacity and Level of Service (LOS)
- Highway Engineering (Design): Vehicle Performance/Human Factors, Geometric Design

Thought Questions

- What constraints keeps us from achieving the goal of transportation systems?
- What is the "Transportation Problem"?

Sample Problem

- Identify a transportation problem (local, regional, national, or global) and consider solutions. Research the efficacy of various solutions. Write a one-page memo documenting the problem and solutions, documenting your references.

Abbreviations

- LOS - Level of Service
- ITE - Institute of Transportation Engineers
- TRB - Transportation Research Board
- TLA - Three letter abbreviation

Key Terms

- Hierarchy of Roads
- Functional Classification
- Modes
- Vehicles
- Freight, Passenger
- Urban, Intercity
- Public, Private

Transportation Economics/ Introduction

Transportation systems are subject to constraints and face questions of resource allocation. The topics of supply and demand, as well as equilibrium and disequilibrium, arise and give shape to the use and capability of the system.

Demand Curve

How much would people pay for a final grade of an A in a transportation engineering class?

- How many people would pay \$5000 for an A?
- How many people would pay \$500 for an A?
- How many people would pay \$50 for an A?
- How many people would pay \$5 for an A?

If we draw out these numbers, with the price on the Y-axis, and the number of people willing to pay on the X-axis, we trace out a demand curve. Unless you run into an exceptionally ethical (or hypocritical) group, the lower the price, the more people are willing to pay for an "A". We can of course replace an "A" with any other good or service, such as the price of gasoline and get a similar though not identical curve.

Demand and Budgets in Transportation

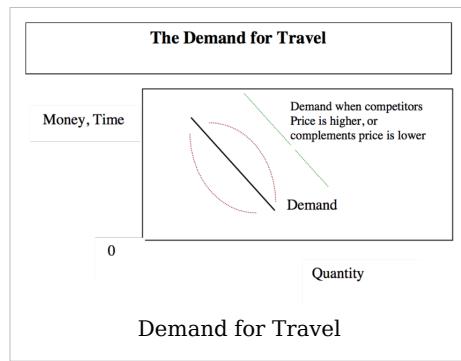
We often say "travel is a derived demand". There would be no travel but for the activities being undertaken at the trip ends. Travel is seldom consumed for its own sake, the occasional "Sunday Drive" or walk in the park excepted. On the other hand, there seems to be some innate need for people to get out of the house, a 20-30 minute separation between the home and workplace is common, and 60 - 90 minutes of travel per day total is common, even for nonworkers. We do know that the more expensive something is, the less of it that will be consumed. E.g. if gas prices were doubled there will be less travel overall. Similarly, the longer it takes to get from A to B, the less likely it is that people will go from A to B.

In short, we are dealing with a downward sloping demand curve, where the curve itself depends not only on the characteristics of the good in question, but also on its complements or substitutes.

The Shape of Demand

What we need to estimate is the shape of demand (is it linear or curved, convex or concave, what function best describes it), the sensitivity of demand for a particular thing (a mode, an origin destination pair, a link, a time of day) to price and time (elasticity) in the short run and the long run.

- Are the choices continuous (the number of miles driven) or discrete (car vs. bus)?
- Are we treating demand as an absolute or a probability?



- Does the probability apply to individuals (disaggregate) or the population as a whole (aggregate)?
- What is the trade-off between money and time?
- What are the effects on demand for a thing as a function of the time and money costs of competitive or complementary choices (cross elasticity).

Supply Curve

How much would a person need to pay you to write an A-quality 20 page term paper for a given transportation class?

- How many would write it for \$100,000?
- How many would write it for \$10,000?
- How many would write it for \$1,000?
- How many would write it for \$100?
- How many would write it for \$10?

If we draw out these numbers for all the potential entrepreneurial people available, we trace out a supply curve. The lower the price, the fewer people are willing to supply the paper-writing service.

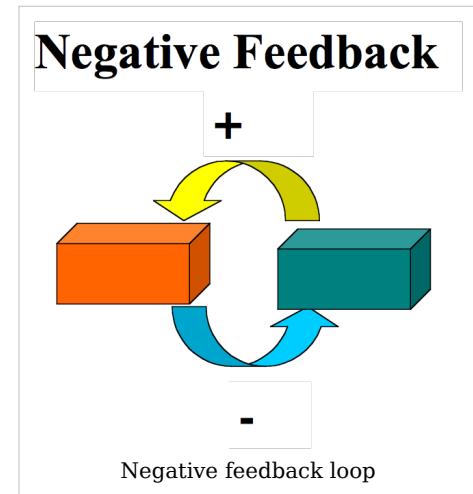
Equilibrium in a Negative Feedback System

Supply and Demand comprise the economists view of transportation systems. They are equilibrium systems. What does that mean?

It means the system is subject to a negative feedback process:

An increase in *A* begets a decrease in *B*. An increase *B* begets an increase in *A*.

Example: *A*: Traffic Congestion and *B*: Traffic Demand
... more congestion limits demand, but more demand creates more congestion.



Supply and Demand Equilibrium

As with earning grades and cheating, transportation is not free, it costs both time and money. These costs are represented by a supply curve, which rises with the amount of travel demanded. As described above, demand (e.g. the number of vehicles which want to use the facility) depends on the price, the lower the price, the higher the demand. These two curves intersect at an equilibrium point. In the example figure, they intersect at a toll of \$0.50 per km, and flow of 3000 vehicles per hour. Time is usually converted to money (using a Value of Time), to simplify the analysis.

Costs may be *variable* and include users' time, out-of-pocket costs (paid on a per trip or per distance basis) like tolls, gasolines, and fares, or *fixed* like insurance or buying an automobile, which are only borne once in a while and are largely independent of the cost of an individual trip.

Disequilibrium

However, many elements of the transportation system do not necessarily generate an equilibrium. Take the case where an increase in A begets an increase in B. An increase in B begets an increase in A. An example where A is an increase in Traffic Demand generates more Gas Tax Revenue (B) and more Gas Tax Revenue generates more Road Building, which in turn increases traffic demand. (This example assumes the gas tax generates more demand from the resultant road building than costs in sensitivity of demand to the price, i.e. the investment is worthwhile). This is dubbed a positive feedback system, and in some contexts a "Virtuous Circle", where the "virtue" is a value judgment that depends on your perspective.

Similarly, one might have a "Vicious Circle" where a decrease in A begets a decrease in B and a decrease in B begets a decrease in A. A classic example of this is where (A) is Transit Service and (B) is Transit Demand. Again "vicious" is a value judgment. Less service results in fewer transit riders, fewer transit riders cannot make as great a claim on transportation resources, leading to more service cutbacks.

These systems of course interact: more road building may attract transit riders to cars, while those additional drivers pay gas taxes and generate more roads.

One might ask whether positive feedback systems converge or diverge. The answer is "it depends on the system", and in particular where or when in the system you observe. There might be some point where no matter how many additional roads you built, there would be no more traffic demand, as everyone already consumes as much travel as they want to. We have yet to reach that point for roads, but on the other hand, we have lots of goods. If you live in most parts of the United States, the price of water at your house probably does not affect how much you drink, and a lower price for tap water would not increase your rate of ingestion. You might use substitutes if their prices were lower (or tap water were costlier), e.g. bottled water. Price might affect other behaviors such as lawn watering and car washing though.

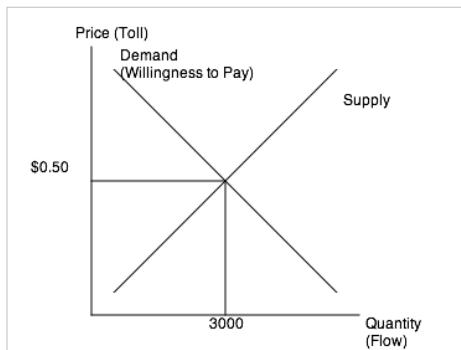
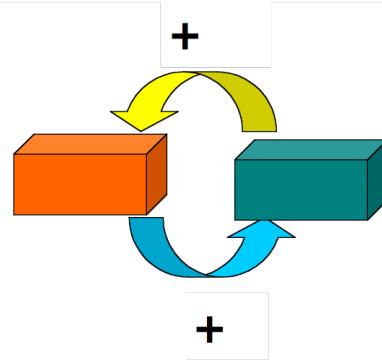


Illustration of equilibrium between supply and demand

Positive Feedback (A Virtuous circle)



Positive feedback loop (virtuous circle)

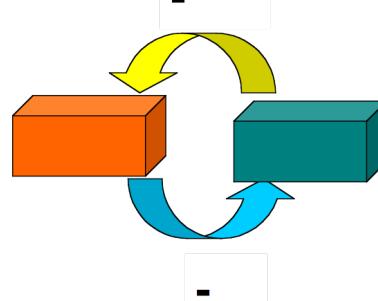
Provision

Transportation services are provided by both the public and private sector.

- Roads are generally publicly owned in the United States, though the same is not true of highways in other countries. Furthermore, public ownership has not always been the norm, many countries had a long history of privately owned turnpikes, in the United States private roads were known through the early 1900s.
- Railroads are generally private.
- Carriers (Airlines, Bus Companies, Truckers, Train Operators) are often private firms
- Formerly private urban transit operators have been taken over by local government from the 1950s in a process called municipalization. With the rise of the automobile, transit systems were steadily losing passengers and money.

The situation is complicated by the idea of contracting or franchising. Often private firms operate "public transit" routes, either under a contract, for a fixed price, or an agreement where the private firm collects the revenue on the route (a franchise agreement). Franchises may be subsidized if the route is a money-loser, or may require bidding if the route is profitable. Private provision of public transport is common in the United Kingdom.

Positive Feedback (A Vicious Circle)



Positive feedback loop (vicious circle)

Thought questions

1. Should the government subsidize public transportation? Why or why not?
2. Should the government operate public transportation systems?
3. Is building roads a good idea even if it results in more travel demand?

Sample Problem

Problem (Solution)



London Routemaster Bus

Key Terms

- Supply
- Demand
- Negative Feedback System
- Equilibrium
- Disequilibrium
- Public Sector
- Private Sector

Fundamentals of Transportation/ Geography and Networks

Transportation systems have specific structure. Roads have length, width, and depth. The characteristics of roads depends on their purpose.

Roads

A road is a path connecting two points. The English word 'road' comes from the same root as the word 'ride' -the Middle English 'rood' and Old English 'rad' -meaning the act of riding. Thus a road refers foremost to the right of way between an origin and destination. In an urban context, the word street is often used rather than road, which dates to the Latin word 'strata', meaning pavement (the additional layer or stratum that might be on top of a path).

Modern roads are generally paved, and unpaved routes are considered trails. The pavement of roads began early in history. Approximately 2600 BCE, the Egyptians constructed a paved road out of sandstone and limestone slabs to assist with the movement of stones on rollers between the quarry and the site of construction of the pyramids. The Romans and others used brick or stone pavers to provide a more level, and smoother surface, especially in urban areas, which allows faster travel, especially of wheeled vehicles. The innovations of Thomas Telford and John McAdam reinvented roads in the early nineteenth century, by using less expensive smaller and broken stones, or aggregate, to maintain a smooth ride and allow for drainage. Later in the nineteenth century, application of tar (asphalt) further smoothed the ride. In 1824, asphalt blocks were used on the Champs-Elysees in Paris. In 1872, the first asphalt street (Fifth Avenue) was paved in New York (due to Edward de Smedt), but it wasn't until bicycles became popular in the late nineteenth century that the "Good Roads Movement" took off. Bicycle travel, more so than travel by other vehicles at the time, was sensitive to rough roads. Demands for higher quality roads really took off with the widespread adoption of the automobile in the United States in the early twentieth century.

The first good roads in the twentieth century were constructed of Portland cement concrete (PCC). The material is stiffer than asphalt (or asphalt concrete) and provides a smoother ride. Concrete lasts slightly longer than asphalt between major repairs, and can carry a heavier load, but is more expensive to build and repair. While urban streets had been paved with concrete in the US as early as 1889, the first rural concrete road was in Wayne County, Michigan, near to Detroit in 1909, and the first concrete highway in 1913 in Pine

Bluff, Arkansas. By the next year over 2300 miles of concrete pavement had been laid nationally. However over the remainder of the twentieth century, the vast majority of roadways were paved with asphalt. In general only the most important roads, carrying the heaviest loads, would be built with concrete.

Roads are generally classified into a hierarchy. At the top of the hierarchy are freeways, which serve entirely a function of moving vehicles between other roads. Freeways are grade-separated and limited access, have high speeds and carry heavy flows. Below freeways are arterials. These may not be grade-separated, and while access is still generally limited, it is not limited to the same extent as freeways, particularly on older roads. These serve both a movement and an access function. Next are collector/distributor roads. These serve more of an access function, allowing vehicles to access the network from origins and destinations, as well as connecting with smaller, local roads, that have only an access function, and are not intended for the movement of vehicles with neither a local origin nor destination. Local roads are designed to be low speed and carry relatively little traffic.

The class of the road determines which level of government administers it. The highest roads will generally be owned, operated, or at least regulated (if privately owned) by the higher level of government involved in road operations; in the United States, these roads are operated by the individual states. As one moves down the hierarchy of roads, the level of government is generally more and more local (counties may control collector/distributor roads, towns may control local streets). In some countries freeways and other roads near the top of the hierarchy are privately owned and regulated as utilities, these are generally operated as toll roads. Even publicly owned freeways are operated as toll roads under a toll authority in other countries, and some US states. Local roads are often owned by adjoining property owners and neighborhood associations.

The design of roads is specified in a number of design manual, including the AASHTO Policy on the Geometric Design of Streets and Highways (or Green Book). Relevant concerns include the alignment of the road, its horizontal and vertical curvature, its super-elevation or banking around curves, its thickness and pavement material, its cross-slope, and its width.

Freeways

A motorway or freeway (sometimes called an expressway or thruway) is a multi-lane divided road that is designed to be high-speed free flowing, access-controlled, built to high standards, with no traffic lights on the mainline. Some motorways or freeways are financed with tolls, and so may have tollbooths, either across the entrance ramp or across the mainline. However in the United States and Great Britain, most are financed with gas or other tax revenue.

Though of course there were major road networks during the Roman Empire and before, the history of motorways and freeways dates at least as early as 1907, when the first limited access automobile highway, the Bronx River Parkway began construction in Westchester County, New York (opening in 1908). In this same period, William Vanderbilt constructed the Long Island Parkway as a toll road in Queens County, New York. The Long Island Parkway was built for racing and speeds of 60 miles per hour (96 km/hr) were accommodated. Users however had to pay a then expensive \$2.00 toll (later reduced) to recover the construction costs of \$2 million. These parkways were paved when most roads

were not. In 1919 General John Pershing assigned Dwight Eisenhower to discover how quickly troops could be moved from Fort Meade between Baltimore and Washington to the Presidio in San Francisco by road. The answer was 62 days, for an average speed of 3.5 miles per hour (5.6 km/hr). While using segments of the Lincoln Highway, most of that road was still unpaved. In response, in 1922 Pershing drafted a plan for an 8,000 mile (13,000 km) interstate system which was ignored at the time.

The US Highway System was a set of paved and consistently numbered highways sponsored by the states, with limited federal support. First built in 1924, they succeeded some previous major highways such as the Dixie Highway, Lincoln Highway and Jefferson Highway that were multi-state and were constructed with the aid of private support. These roads however were not in general access-controlled, and soon became congested as development along the side of the road degraded highway speeds.

In parallel with the US Highway system, limited access parkways were developed in the 1920s and 1930s in several US cities. Robert Moses built a number of these parkways in and around New York City. A number of these parkways were grade separated, though they were intentionally designed with low bridges to discourage trucks and buses from using them. German Chancellor Adolf Hitler appointed a German engineer Fritz Todt Inspector General for German Roads. He managed the construction of the German Autobahns, the first limited access high-speed road network in the world. In 1935, the first section from Frankfurt am Main to Darmstadt opened, the total system today has a length of 11,400 km. The Federal-Aid Highway Act of 1938 called on the Bureau of Public Roads to study the feasibility of a toll-financed superhighway system (three east-west and three north-south routes). Their report Toll Roads and Free Roads declared such a system would not be self-supporting, advocating instead a 43,500 km (27,000 mile) free system of interregional highways, the effect of this report was to set back the interstate program nearly twenty years in the US.

The German autobahn system proved its utility during World War II, as the German army could shift relatively quickly back and forth between two fronts. Its value in military operations was not lost on the American Generals, including Dwight Eisenhower.

On October 1, 1940, a new toll highway using the old, unutilized South Pennsylvania Railroad right-of-way and tunnels opened. It was the first of a new generation of limited access highways, generally called superhighways or freeways that transformed the American landscape. This was considered the first freeway in the US, as it, unlike the earlier parkways, was a multi-lane route as well as being limited access. The Arroyo Seco Parkway, now the Pasadena Freeway, opened December 30, 1940. Unlike the Pennsylvania Turnpike, the Arroyo Seco parkway had no toll barriers.

A new National Interregional Highway Committee was appointed in 1941, and reported in 1944 in favor of a 33,900 mile system. The system was designated in the Federal Aid Highway Act of 1933, and the routes began to be selected by 1947, yet no funding was provided at the time. The 1952 highway act only authorized a token amount for construction, increased to \$175 million annually in 1956 and 1957.

The US Interstate Highway System was established in 1956 following a decade and half of discussion. Much of the network had been proposed in the 1940s, but it took time to authorize funding. In the end, a system supported by gas taxes (rather than tolls), paid for 90% by the federal government with a 10% local contribution, on a pay-as-you-go" system, was established. The Federal Aid Highway Act of 1956 had authorized the expenditure of

\$27.5 billion over 13 years for the construction of a 41,000 mile interstate highway system. As early as 1958 the cost estimate for completing the system came in at \$39.9 billion and the end date slipped into the 1980s. By 1991, the final cost estimate was \$128.9 billion. While the freeways were seen as positives in most parts of the US, in urban areas opposition grew quickly into a series of freeway revolts. As soon as 1959, (three years after the Interstate act), the San Francisco Board of Supervisors removed seven of ten freeways from the city's master plan, leaving the Golden Gate bridge unconnected to the freeway system. In New York, Jane Jacobs led a successful freeway revolt against the Lower Manhattan Expressway sponsored by business interests and Robert Moses among others. In Baltimore, I-70, I-83, and I-95 all remain unconnected thanks to highway revolts led by now Senator Barbara Mikulski. In Washington, I-95 was rerouted onto the Capital Beltway. The pattern repeated itself elsewhere, and many urban freeways were removed from Master Plans.

In 1936, the Trunk Roads Act ensured that Great Britain's Minister of Transport controlled about 30 major roads, of 7,100 km (4,500 miles) in length. The first Motorway in Britain, the Preston by-pass, now part of the M-6, opened in 1958. In 1959, the first stretch of the M1 opened. Today there are about 10,500 km (6300 miles) of trunk roads and motorways in England.

Australia has 790 km of motorways, though a much larger network of roads. However the motorway network is not truly national in scope (in contrast with Germany, the United States, Britain, and France), rather it is a series of local networks in and around metropolitan areas, with many intercity connection being on undivided and non-grade separated highways. Outside the Anglo-Saxon world, tolls were more widely used. In Japan, when the Meishin Expressway opened in 1963, the roads in Japan were in far worse shape than Europe or North American prior to this. Today there are over 6,100 km of expressways (3,800 miles), many of which are private toll roads. France has about 10,300 km of expressways (6,200 miles) of motorways, many of which are toll roads. The French motorway system developed through a series of franchise agreements with private operators, many of which were later nationalized. Beginning in the late 1980s with the wind-down of the US interstate system (regarded as complete in 1990), as well as intercity motorway programs in other countries, new sources of financing needed to be developed. New (generally suburban) toll roads were developed in several metropolitan areas.

An exception to the dearth of urban freeways is the case of the Big Dig in Boston, which relocates the Central Artery from an elevated highway to a subterranean one, largely on the same right-of-way, while keeping the elevated highway operating. This project is estimated to be completed for some \$14 billion; which is half the estimate of the original complete US Interstate Highway System.

As mature systems in the developed countries, improvements in today's freeways are not so much widening segments or constructing new facilities, but better managing the roadspace that exists. That improved management, takes a variety of forms. For instance, Japan has advanced its highways with application of Intelligent Transportation Systems, in particular traveler information systems, both in and out of vehicles, as well as traffic control systems. The US and Great Britain also have traffic management centers in most major cities that assess traffic conditions on motorways, deploy emergency vehicles, and control systems like ramp meters and variable message signs. These systems are beneficial, but cannot be seen as revolutionizing freeway travel. Speculation about future automated highway systems has

taken place almost as long as highways have been around. The *Futurama* exhibit at the New York 1939 World's Fair posited a system for 1960. Yet this technology has been twenty years away for over sixty years, and difficulties remain.

Layers of Networks

The road is itself part of a layer of subsystems of which the pavement surface is only one part. We can think of a hierarchy of systems.

- Places
- Trip Ends
- End to End Trip
- Driver/Passenger
- Service (Vehicle & Schedule)
- Signs and Signals
- Markings
- Pavement Surface
- Structure (Earth & Pavement and Bridges)
- Alignment (Vertical and Horizontal)
- Right-Of-Way
- Space

At the base is *space*. On space, a specific *right-of-way* is designated, which is property where the road goes. Originally right-of-way simply meant legal permission for travelers to cross someone's property. Prior to the construction of roads, this might simply be a well-worn dirt path.

On top of the right-of-way is the *alignment*, the specific path a transportation facility takes within the right-of-way. The path has both vertical and horizontal elements, as the road rises or falls with the topography and turns as needed.

Structures are built on the alignment. These include the roadbed as well as bridges or tunnels that carry the road.

Pavement surface is the gravel or asphalt or concrete surface that vehicles actually ride upon and is the top layer of the structure. That surface may have *markings* to help guide drivers to stay to the right (or left), delineate lanes, regulate which vehicles can use which lanes (bicycles-only, high occupancy vehicles, buses, trucks) and provide additional information. In addition to marking, signs and signals to the side or above the road provide additional regulatory and navigation information.

Services use roads. Buses may provide scheduled services between points with stops along the way. Coaches provide scheduled point-to-point without stops. Taxis handle irregular passenger trips.

Drivers and *passengers* use services or drive their own vehicle (producing their own transportation services) to create an *end-to-end trip*, between an origin and destination. Each origin and destination comprises a *trip end* and those trip ends are only important because of the *places* at the ends and the activity that can be engaged in. As transportation is a derived demand, if not for those activities, essentially no passenger travel would be undertaken.

With modern information technologies, we may need to consider additional systems, such as Global Positioning Systems (GPS), differential GPS, beacons, transponders, and so on

that may aide the steering or navigation processes. Cameras, in-pavement detectors, cell phones, and other systems monitor the use of the road and may be important in providing feedback for real-time control of signals or vehicles.

Each layer has rules of behavior:

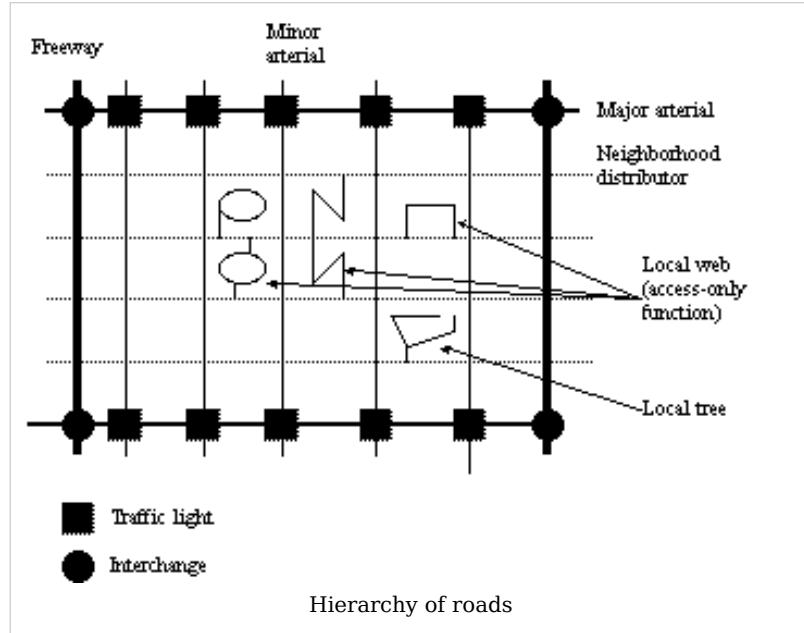
- some rules are physical and never violated, others are physical but probabilistic
- some are legal rules or social norms which are occasionally violated

Hierarchy of Roads

Even within each layer of the system of systems described above, there is differentiation.

Transportation facilities have two distinct functions: through movement and land access. This differentiation:

- permits the aggregation of traffic to achieve economies of scale in construction and operation (high speeds);
- reduces the number of conflicts;
- helps maintain the desired quiet character of residential neighborhoods by keeping through traffic away from homes;
- contains less redundancy, and so may be less costly to build.



| Functional Classification | Types of Connections | Relation to Abutting Property | Minnesota Examples |
|---|--|--|--------------------------------------|
| Limited Access (highway) | Through traffic movement between cities and across cities | Limited or controlled access highways with ramps and/or curb cut controls. | I-94, Mn280 |
| Linking (arterial:principal and minor) | Traffic movement between limited access and local streets. | Direct access to abutting property. | University Avenue, Washington Avenue |
| Local (collector and distributor roads) | Traffic movement in and between residential areas | Direct access to abutting property. | Pillsbury Drive, 17th Avenue |

Model Elements

Transportation forecasting, to be discussed in more depth in subsequent modules, abstracts the real world into a simplified representation.

Recall the hierarchy of roads. What can be simplified? It is typical for a regional forecasting model to eliminate local streets and replace them with a centroid (a point representing a traffic analysis zone). Centroids are the source and sink of all transportation demand on the network. Centroid connectors are artificial or dummy links connecting the centroid to the "real" network. An illustration of traffic analysis zones can be found at this external link for Fulton County, Georgia, here: traffic zone map, 3MB ^[1]. Keep in mind that *Models are abstractions*.

Network

- Zone Centroid - special node whose number identifies a zone, located by an "x" "y" coordinate representing longitude and latitude (sometimes "x" and "y" are identified using planar coordinate systems).
- Node (vertices) - intersection of links , located by x and y coordinates
- Links (arcs) - short road segments indexed by from and to nodes (including centroid connectors), attributes include lanes, capacity per lane, allowable modes
- Turns - indexed by at, from, and to nodes
- Routes, (paths) - indexed by a series of nodes from origin to destination. (e.g. a bus route)
- Modes - car, bus, HOV, truck, bike, walk etc.

Matrices

Scalar

A scalar is a single value that applies model-wide; e.g. the price of gas or total trips.

| | Total Trips |
|----------|-------------|
| Variable | T |

Vectors

Vectors are values that apply to particular zones in the model system, such as trips produced or trips attracted or number of households. They are arrayed separately when treating a zone as an origin or as a destination so that they can be combined into full matrices.

- vector (origin) - a column of numbers indexed by traffic zones, describing attributes at the origin of the trip (e.g. the number of households in a zone)

| | Trips Produced at Origin Zone |
|---------------|-------------------------------|
| Origin Zone 1 | T_{i_1} |
| Origin Zone 2 | T_{i_2} |
| Origin Zone 3 | T_{i_3} |

- vector (destination) - a row of numbers indexed by traffic zones, describing attributes at the destination

| | Destination Zone 1 | Destination Zone 2 | Destination Zone 3 |
|--|----------------------------|----------------------------|----------------------------|
| Trips Attracted to Destination Zone | T _{j₁} | T _{j₂} | T _{j₃} |
| | | | |

Full Matrices

A full or interaction matrix is a table of numbers, describing attributes of the origin-destination pair

| | Destination Zone 1 | Destination Zone 2 | Destination Zone 3 |
|----------------------|---------------------------|---------------------------|---------------------------|
| Origin Zone 1 | T ₁₁ | T ₁₂ | T ₁₃ |
| Origin Zone 2 | T ₂₁ | T ₂₂ | T ₂₃ |
| Origin Zone 3 | T ₃₁ | T ₃₂ | T ₃₃ |

Thought Questions

- Identify the rules associated with each layer?
- Why aren't all roads the same?
- How might we abstract the real transportation system when representing it in a model for analysis?
- Why is abstraction useful?

Variables

- T - Total Trips
- T_{i_k} - Trips Produced from Origin Zone k
- T_{j_k} - Trips Attracted to Destination Zone k
- T_{ij} - Trips Going Between Origin Zone i and Destination Zone j

Key Terms

- Zone Centroid
- Node
- Links
- Turns
- Routes
- Modes
- Matrices
- Right-of-way
- Alignment
- Structures
- Pavement Surface
- Markings
- Signs and Signals
- Services
- Driver
- Passenger
- End to End Trip

- Trip Ends
- Places

External Exercises

Use the ADAM software at the STREET website [2] and examine the network structure. Familiarize yourself with the software, and edit the network, adding at least two nodes and four one-way links (two two-way links), and deleting nodes and links. What are the consequences of such network adjustments? Are some adjustments better than others?

References

[1] http://wms.co.fulton.ga.us/apps/doc_archive/get.php/69253.pdf

[2] <http://street.umn.edu>

Fundamentals of Transportation/Trip Generation

Trip Generation is the first step in the conventional four-step transportation forecasting process (followed by Destination Choice, Mode Choice, and Route Choice), widely used for forecasting travel demands. It predicts the number of trips originating in or destined for a particular traffic analysis zone.

Every trip has two ends, and we need to know where both of them are. The first part is determining how many trips originate in a zone and the second part is how many trips are destined for a zone. Because land use can be divided into two broad category (residential and non-residential)<footnote>There are two types of people in the world, those that divide the world into two kinds of people and those that don't. Some people say there are three types of people in the world, those who can count, and those who can't.</footnote> we have models that are household based and non-household based (e.g. a function of number of jobs or retail activity).

For the residential side of things, trip generation is thought of as a function of the social and economic attributes of households (households and housing units are very similar measures, but sometimes housing units have no households, and sometimes they contain multiple households, clearly housing units are easier to measure, and those are often used instead for models, it is important to be clear which assumption you are using).

At the level of the traffic analysis zone, the language is that of land uses "producing" or *attracting* trips, where by assumption trips are "produced" by households and "attracted" to non-households. Production and attractions differ from origins and destinations. Trips are produced by households even when they are returning home (that is, when the household is a destination). Again it is important to be clear what assumptions you are using.

Activities

People engage in activities, these activities are the "purpose" of the trip. Major activities are home, work, shop, school, eating out, socializing, recreating, and serving passengers (picking up and dropping off). There are numerous other activities that people engage on a less than daily or even weekly basis, such as going to the doctor, banking, etc. Often less frequent categories are dropped and lumped into the catchall "Other".

Every trip has two ends, an origin and a destination. Trips are categorized by *purposes*, the activity undertaken at a destination location.

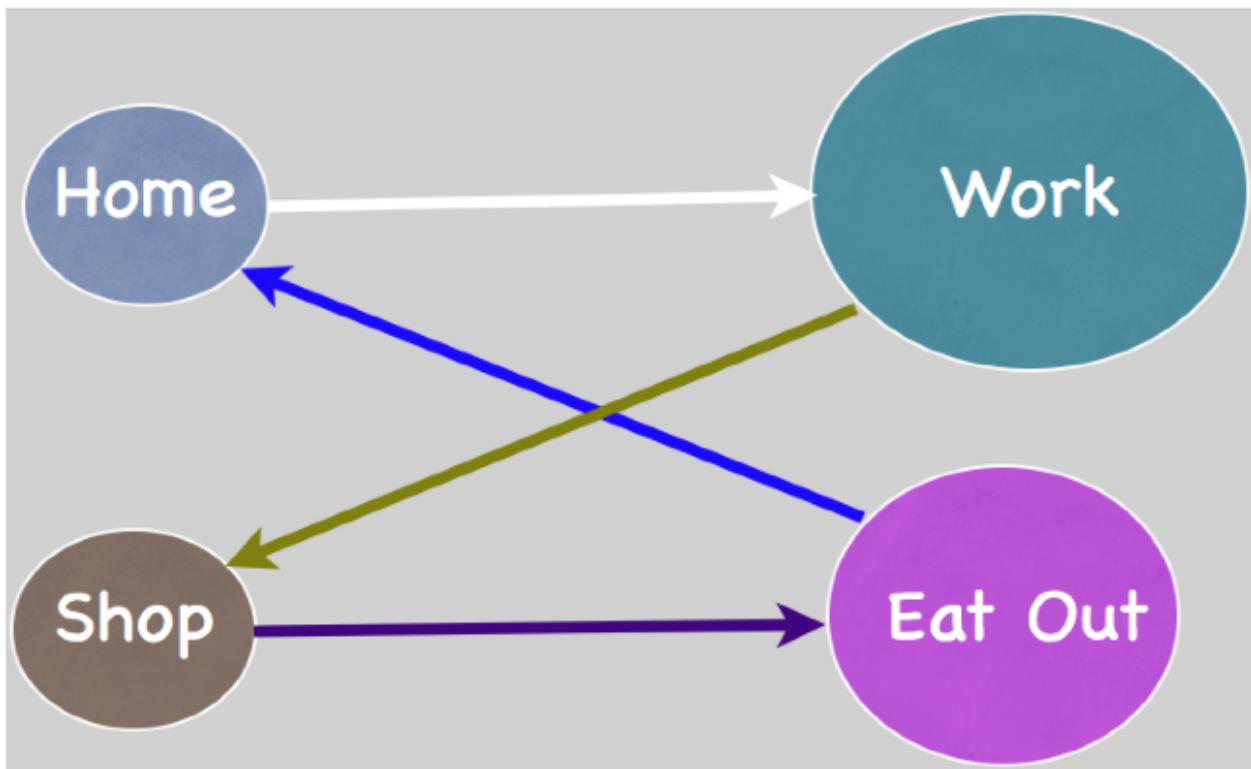
Observed trip making from the Twin Cities (2000-2001) Travel Behavior Inventory by Gender

| Trip Purpose | Males | Females | Total |
|---|-------|---------|-------|
| Work | 4008 | 3691 | 7691 |
| Work related | 1325 | 698 | 2023 |
| Attending school | 495 | 465 | 960 |
| Other school activities | 108 | 134 | 242 |
| Childcare, daycare, after school care | 111 | 115 | 226 |
| Quickstop | 45 | 51 | 96 |
| Shopping | 2972 | 4347 | 7319 |
| Visit friends or relatives | 856 | 1086 | 1942 |
| Personal business | 3174 | 3928 | 7102 |
| Eat meal outside of home | 1465 | 1754 | 3219 |
| Entertainment, recreation, fitness | 1394 | 1399 | 2793 |
| Civic or religious | 307 | 462 | 769 |
| Pick up or drop off passengers | 1612 | 2490 | 4102 |
| With another person at their activities | 64 | 48 | 112 |
| At home activities | 288 | 384 | 672 |

Some observations:

- Men and women behave differently, splitting responsibilities within households, and engaging in different activities,
- Most trips are not work trips, though work trips are important because of their peaked nature (and because they tend to be longer in both distance and travel time),
- The vast majority of trips are not people going to (or from) work.

People engage in activities in sequence, and may chain their trips. In the Figure below, the trip-maker is traveling from home to work to shop to eating out and then returning home.



Specifying Models

How do we predict how many trips will be generated by a zone? The number of trips originating from or destined to a purpose in a zone are described by trip rates (a cross-classification by age or demographics is often used) or equations. First, we need to identify what we think are the relevant variables.

Home-end

The total number of trips leaving or returning to homes in a zone may be described as a function of:

$$T_h = f(\text{housing units}, \text{household size}, \text{age}, \text{income}, \text{accessibility}, \text{vehicle ownership}).$$

Home-End Trips are sometimes functions of:

- Housing Units
- Household Size
- Age
- Income
- Accessibility
- Vehicle Ownership
- Other Home-Based Elements

Work-end

At the work-end of work trips, the number of trips generated might be a function as below:

$$T_w = f(jobs \text{ (square feet of space by type, occupancy rate)})$$

Work-End Trips are sometimes functions of:

- Jobs
- Square Footage of Workspace
- Occupancy Rate
- Other Job-Related Elements

Shop-end

Similarly shopping trips depend on a number of factors:

$$T_s = f(number \text{ of retail workers, type of retail, square foot, location, competition})$$

Shop-End Trips are sometimes functions of:

- Number of Retail Workers
- Type of Retail Available
- Square Footage of Retail Available
- Location
- Competition
- Other Retail-Related Elements

Input Data

A forecasting activity conducted by planners or economists, such as one based on the concept of economic base analysis, provides aggregate measures of population and activity growth. Land use forecasting distributes forecast changes in activities across traffic zones.

Estimating Models

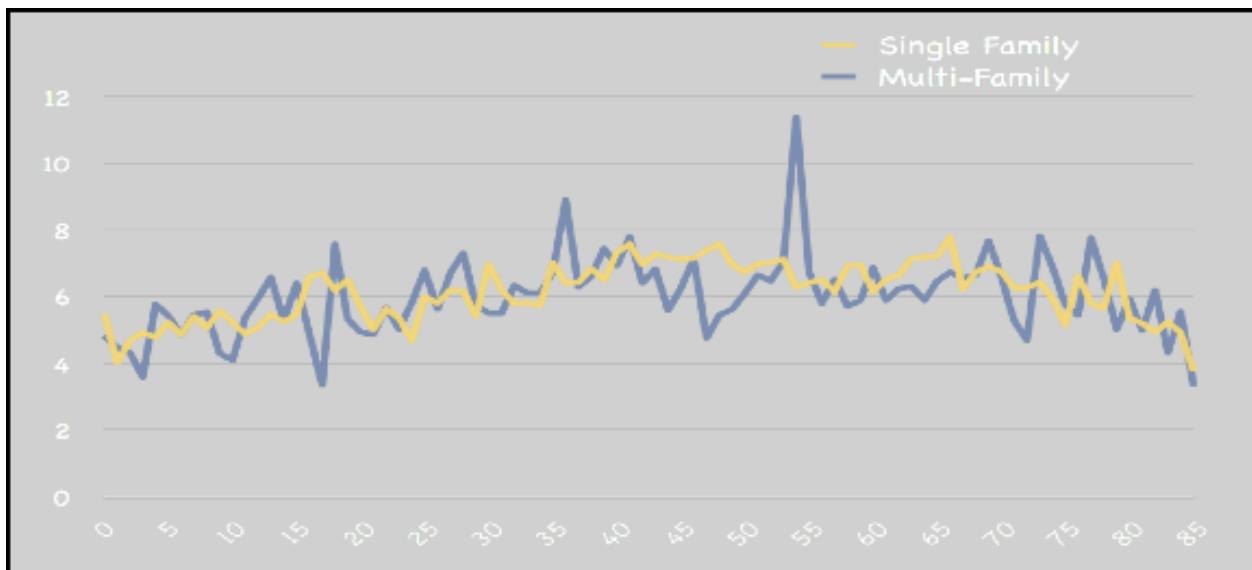
Which is more accurate: the data or the average? The problem with averages (or aggregates) is that every individual's trip-making pattern is different.

Home-end

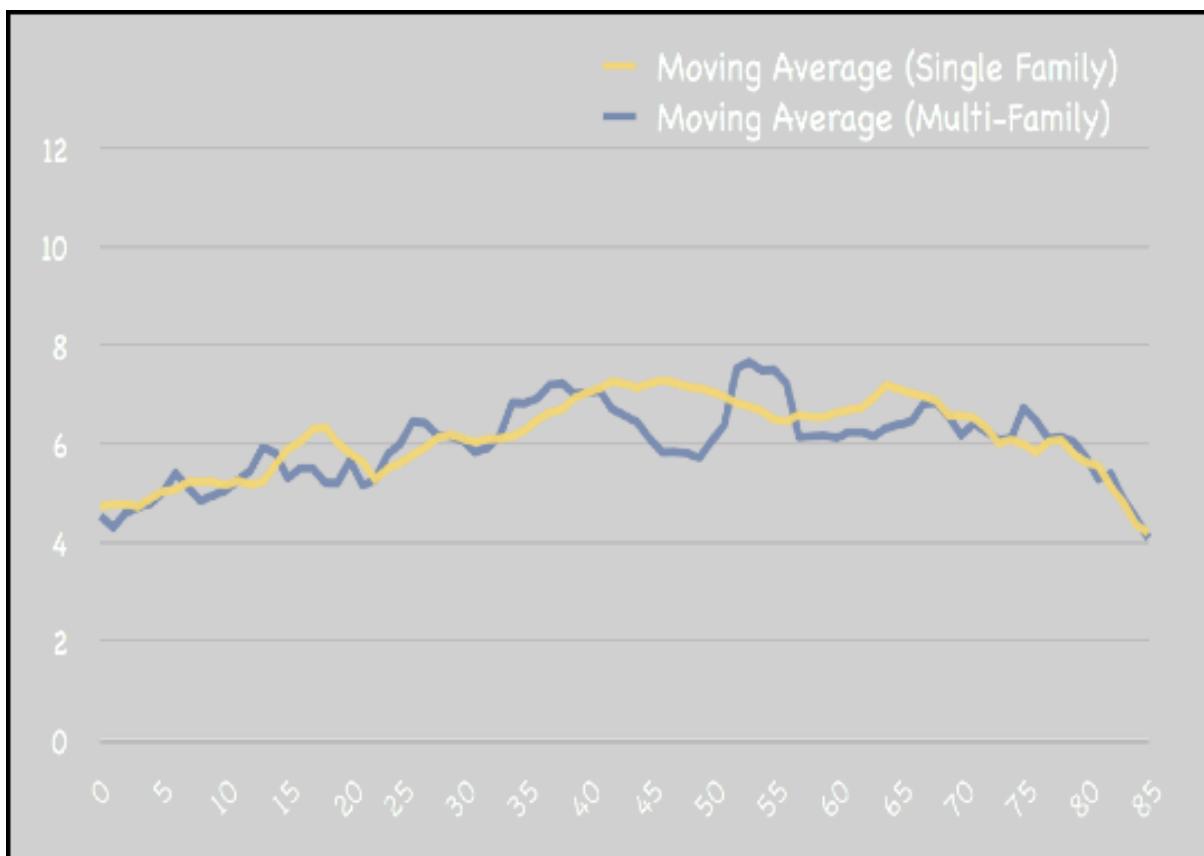
To estimate trip generation at the home end, a cross-classification model can be used, this is basically constructing a table where the rows and columns have different attributes, and each cell in the table shows a predicted number of trips, this is generally derived directly from data.

In the example cross-classification model: The dependent variable is trips per person. The independent variables are dwelling type (single or multiple family), household size (1, 2, 3, 4, or 5+ persons per household), and person age.

The figure below shows a typical example of how trips vary by age in both single-family and multi-family residence types.



The figure below shows a moving average.



Non-home-end

The trip generation rates for both “work” and “other” trip ends can be developed using Ordinary Least Squares (OLS) regression (a statistical technique for fitting curves to minimize the sum of squared errors (the difference between predicted and actual value) relating trips to employment by type and population characteristics).

The variables used in estimating trip rates for the work-end are Employment in Offices (E_{off}), Retail (E_{ret}), and Other (E_{oth})

A typical form of the equation can be expressed as:

$$T_i = a_1 E_{off,i} + a_2 E_{oth,i} + a_3 E_{ret,i}$$

Where:

- T_i - Person trips attracted per worker in the ith zone
- $E_{off,i}$ - office employment in the ith zone
- $E_{oth,i}$ - other employment in the ith zone
- $E_{ret,i}$ - retail employment in the ith zone
- a_1, a_2, a_3 - model coefficients

Normalization

For each trip purpose (e.g. home to work trips), the number of trips originating at home must equal the number of trips destined for work. Two distinct models may give two results. There are several techniques for dealing with this problem. One can either assume one model is correct and adjust the other, or split the difference.

It is necessary to ensure that the total number of trip origins equals the total number of trip destinations, since each trip interchange by definition must have two trip ends.

The rates developed for the home end are assumed to be most accurate,

The basic equation for normalization:

$$T'_j = T_j \frac{\sum_{i=1}^I T_i}{\sum_{j=1}^J T_j}$$

Sample Problems

- Problem (Solution)

Variables

- T_i - Person trips originating in Zone i
- T_j - Person Trips destined for Zone j
- T'_i - Normalized Person trips originating in Zone i
- T'_j - Normalized Person Trips destined for Zone j
- T_h - Person trips generated at home end (typically morning origins, afternoon destinations)
- T_w - Person trips generated at work end (typically afternoon origins, morning destinations)

- T_s - Person trips generated at shop end
- H_i - Number of Households in Zone i
- $E_{off,i}$ - office employment in the ith zone
- $E_{ret,i}$ - retail employment in the ith zone
- $E_{oth,i}$ - other employment in the ith zone
- B_n - model coefficients

Abbreviations

- H2W - Home to work
- W2H - Work to home
- W2O - Work to other
- O2W - Other to work
- H2O - Home to other
- O2H - Other to home
- O2O - Other to other
- HBO - Home based other (includes H2O, O2H)
- HBW - Home based work (H2W, W2H)
- NHB - Non-home based (O2W, W2O, O2O)

External Exercises

Use the ADAM software at the STREET website ^[1] and try Assignment #1 to learn how changes in analysis zone characteristics generate additional trips on the network.

End Notes

[1] <http://street.umn.edu/>

Further Reading

- Trip Generation article on wikipedia (http://en.wikipedia.org/wiki/Trip_generation)

Fundamentals of Transportation/Trip Generation/Problem


Problem:

Planners have estimated the following models for the AM Peak Hour

$$T_i = 1.5 * H_i$$

$$T_j = (1.5 * E_{off,j}) + (1 * E_{oth,j}) + (0.5 * E_{ret,j})$$

Where:

T_i = Person Trips Originating in Zone i

T_j = Person Trips Destined for Zone j

H_i = Number of Households in Zone i

You are also given the following data

Data

| Variable | Dakotopolis | New Fargo |
|-----------|-------------|-----------|
| H | 10000 | 15000 |
| E_{off} | 8000 | 10000 |
| E_{oth} | 3000 | 5000 |
| E_{ret} | 2000 | 1500 |

- A. What are the number of person trips originating in and destined for each city?
- B. Normalize the number of person trips so that the number of person trip origins = the number of person trip destinations. Assume the model for person trip origins is more accurate.
- Solution

Fundamentals of Transportation/Trip Generation/Solution


Problem:

Planners have estimated the following models for the AM Peak Hour

$$T_i = 1.5 * H_i$$

$$T_j = (1.5 * E_{off,j}) + (1 * E_{oth,j}) + (0.5 * E_{ret,j})$$

Where:

T_i = Person Trips Originating in Zone i

T_j = Person Trips Destined for Zone j

H_i = Number of Households in Zone i

You are also given the following data

Data

| Variable | Dakotopolis | New Fargo |
|-----------|-------------|-----------|
| H | 10000 | 15000 |
| E_{off} | 8000 | 10000 |
| E_{oth} | 3000 | 5000 |
| E_{ret} | 2000 | 1500 |

- A. What are the number of person trips originating in and destined for each city?
- B. Normalize the number of person trips so that the number of person trip origins = the number of person trip destinations. Assume the model for person trip origins is more accurate.


Solution:

- A. What are the number of person trips originating in and destined for each city?

Solution to Trip Generation Problem Part A

| | Households (H_i) | Office Employees (E_{off}) | Other Employees (E_{oth}) | Retail Employees (E_{ret}) | Origins $T_j = (1.5 * E_{off,j}) + (1 * E_{oth,j}) + (0.5 * E_{ret,j})$ | Destinations |
|-------------|----------------------|--------------------------------|-------------------------------|--------------------------------|--|--------------|
| Dakotopolis | 10000 | 8000 | 3000 | 2000 | 15000 | 16000 |
| New Fargo | 15000 | 10000 | 5000 | 1500 | 22500 | 20750 |
| Total | 25000 | 18000 | 8000 | 3000 | 37500 | 36750 |

- B. Normalize the number of person trips so that the number of person trip origins = the number of person trip destinations. Assume the model for person trip origins is more

accurate.

$$\text{Use: } T'_j = T_j \frac{\sum_{i=1}^I T_i}{\sum_{j=1}^J T_j} \Rightarrow T_j \frac{37500}{36750} = T_j * 1.0204$$

Solution to Trip Generation Problem Part B

| | Origins (T_i) | Destinations (T_j) | Adjustment Factor | Normalized Destinations (T'_j) | Rounded |
|-------------|-------------------|------------------------|-------------------|------------------------------------|---------|
| Dakotopolis | 15000 | 16000 | 1.0204 | 16326.53 | 16327 |
| New Fargo | 22500 | 20750 | 1.0204 | 21173.47 | 21173 |
| Total | 37500 | 36750 | 1.0204 | 37500 | 37500 |

Fundamentals of Transportation/ Destination Choice

Everything is related to everything else, but near things are more related than distant things. - Waldo Tobler's 'First Law of Geography'

Trip distribution (or **destination choice** or **zonal interchange analysis**), is the second component (after Trip Generation, but before Mode Choice and Route Choice) in the traditional four-step transportation forecasting model. This step matches tripmakers' origins and destinations to develop a "trip table", a matrix that displays the number of trips going from each origin to each destination. Historically, trip distribution has been the least developed component of the transportation planning model.

Table: Illustrative Trip Table

| Origin \ Destination | 1 | 2 | 3 | Z |
|----------------------|----------|----------|----------|----------|
| 1 | T_{11} | T_{12} | T_{13} | T_{1Z} |
| 2 | T_{21} | | | |
| 3 | T_{31} | | | |
| Z | T_{Z1} | | | T_{ZZ} |

Where: T_{ij} = Trips from origin i to destination j . Work trip distribution is the way that travel demand models understand how people take jobs. There are trip distribution models for other (non-work) activities, which follow the same structure.

Fratar Models

The simplest trip distribution models (Fratar or Growth models) simply extrapolate a base year trip table to the future based on growth, $T_{ijy+1} = g * T_{ijy}$

where:

- T_{ijy} - Trips from i to j in year y
- g - growth factor

Fratar Model takes no account of changing spatial accessibility due to increased supply or changes in travel patterns and congestion.

Gravity Model

The gravity model illustrates the macroscopic relationships between places (say homes and workplaces). It has long been posited that the interaction between two locations declines with increasing (distance, time, and cost) between them, but is positively associated with the amount of activity at each location (Isard, 1956). In analogy with physics, Reilly (1929) formulated Reilly's law of retail gravitation, and J. Q. Stewart (1948) formulated definitions of demographic gravitation, force, energy, and potential, now called accessibility (Hansen, 1959). The distance decay factor of $1/distance$ has been updated to a more comprehensive function of generalized cost, which is not necessarily linear - a negative exponential tends to be the preferred form. In analogy with Newton's law of gravity, a gravity model is often used in transportation planning.

The gravity model has been corroborated many times as a basic underlying aggregate relationship (Scott 1988, Cervero 1989, Levinson and Kumar 1995). The rate of decline of the interaction (called alternatively, the impedance or friction factor, or the utility or propensity function) has to be empirically measured, and varies by context.

Limiting the usefulness of the gravity model is its aggregate nature. Though policy also operates at an aggregate level, more accurate analyses will retain the most detailed level of information as long as possible. While the gravity model is very successful in explaining the choice of a large number of individuals, the choice of any given individual varies greatly from the predicted value. As applied in an urban travel demand context, the disutilities are primarily time, distance, and cost, although discrete choice models with the application of more expansive utility expressions are sometimes used, as is stratification by income or auto ownership.

Mathematically, the gravity model often takes the form:

$$T_{ij} = K_i K_j T_i T_j f(C_{ij})$$

$$\sum_j T_{ij} = T_i, \sum_i T_{ij} = T_j$$

$$K_i = \frac{1}{\sum_j K_j T_j f(C_{ij})}, K_j = \frac{1}{\sum_i K_i T_i f(C_{ij})}$$

where

- T_{ij} = Trips between origin i and destination j
- T_i = Trips originating at i
- T_j = Trips destined for j
- C_{ij} = travel cost between i and j
- K_i, K_j = balancing factors solved iteratively.

- f = impedance or distance decay factor

It is doubly constrained so that Trips from i to j equal number of origins and destinations.

Balancing a matrix

1. Assess Data, you have T_i , T_j , C_{ij}

2. Compute $f(C_{ij})$, e.g.

- $f(C_{ij}) = 1/C_{ij}^2$
- $f(C_{ij}) = e^{-\beta C_{ij}}$

3. Iterate to Balance Matrix

(a) Multiply Trips from Zone i (T_i) by Trips to Zone j (T_j) by Impedance in Cell ij ($f(C_{ij})$) for all ij

(b) Sum Row Totals T'_i , Sum Column Totals T'_j

(c) Multiply Rows by $N_i = T_i/T'_i$

(d) Sum Row Totals T'_i , Sum Column Totals T'_j

(e) Compare T_i and T'_i , T_j and T'_j if within tolerance stop, Otherwise goto (f)

(f) Multiply Columns by $N_j = T_j/T'_j$

(g) Sum Row Totals T'_i , Sum Column Totals T'_j

(h) Compare T_i and T'_i , T_j and T'_j if within tolerance stop, Otherwise goto (b)

Issues

Feedback

One of the key drawbacks to the application of many early models was the inability to take account of congested travel time on the road network in determining the probability of making a trip between two locations. Although Wohl noted as early as 1963 research into the feedback mechanism or the “interdependencies among assigned or distributed volume, travel time (or travel ‘resistance’) and route or system capacity”, this work has yet to be widely adopted with rigorous tests of convergence or with a so-called “equilibrium” or “combined” solution (Boyce et al. 1994). Haney (1972) suggests internal assumptions about travel time used to develop demand should be consistent with the output travel times of the route assignment of that demand. While small methodological inconsistencies are necessarily a problem for estimating base year conditions, forecasting becomes even more tenuous without an understanding of the feedback between supply and demand. Initially heuristic methods were developed by Irwin and Von Cube (as quoted in Florian et al. (1975)) and others, and later formal mathematical programming techniques were established by Evans (1976).

Feedback and time budgets

A key point in analyzing feedback is the finding in earlier research by Levinson and Kumar (1994) that commuting times have remained stable over the past thirty years in the Washington Metropolitan Region, despite significant changes in household income, land use pattern, family structure, and labor force participation. Similar results have been found in the Twin Cities by Barnes and Davis (2000).

The stability of travel times and distribution curves over the past three decades gives a good basis for the application of aggregate trip distribution models for relatively long term forecasting. This is not to suggest that there exists a constant travel time budget.

In terms of time budgets:

- 1440 Minutes in a Day
- Time Spent Traveling: ~ 100 minutes + or -
- Time Spent Traveling Home to Work: 20 - 30 minutes + or -

Research has found that auto commuting times have remained largely stable over the past forty years, despite significant changes in transportation networks, congestion, household income, land use pattern, family structure, and labor force participation. The stability of travel times and distribution curves gives a good basis for the application of trip distribution models for relatively long term forecasting.

Examples

Example 1: Solving for impedance



Problem:

You are given the travel times between zones, compute the impedance matrix $f(C_{ij})$, assuming $f(C_{ij}) = 1/C_{ij}^2$.

Travel Time OD Matrix (

| Origin Zone | Destination Zone 1 | Destination Zone 2 |
|-------------|--------------------|--------------------|
| 1 | 2 | 5 |
| 2 | 5 | 2 |

Compute impedances ($f(C_{ij})$)



Solution:

Impedance Matrix (

| Origin Zone | Destination Zone 1 | Destination Zone 2 |
|-------------|------------------------|------------------------|
| 1 | $\frac{1}{2^2} = 0.25$ | $\frac{1}{5^2} = 0.04$ |
| 2 | $\frac{1}{5^2} = 0.04$ | $\frac{1}{2^2} = 0.25$ |

Example 2: Balancing a Matrix Using Gravity Model



Problem:

You are given the travel times between zones, trips originating at each zone (zone 1 = 15, zone 2 = 15) trips destined for each zone (zone 1 = 10, zone 2 = 20) and asked to use the classic gravity model $f(C_{ij}) = 1/C_{ij}^2$

Travel Time OD Matrix (

| Origin Zone | Destination Zone 1 | Destination Zone 2 |
|-------------|--------------------|--------------------|
| 1 | 2 | 5 |
| 2 | 5 | 2 |



Solution:

(a) Compute impedances ($f(C_{ij})$)

Impedance Matrix (

| Origin Zone | Destination Zone 1 | Destination Zone 2 |
|-------------|--------------------|--------------------|
| 1 | 0.25 | 0.04 |
| 2 | 0.04 | 0.25 |

(b) Find the trip table

Balancing Iteration 0 (Set-up)

| Origin Zone | Trips Originating | Destination Zone 1 | Destination Zone 2 |
|----------------|-------------------|--------------------|--------------------|
| Trips Destined | | 10 | 20 |
| 1 | 15 | 0.25 | 0.04 |
| 2 | 15 | 0.04 | 0.25 |

Balancing Iteration 1 (

| Origin Zone | Trips Originating | Destination Zone 1 | Destination Zone 2 | Row Total T'_i | Normalizing Factor $N_i = T_i/T'_i$ |
|----------------|-------------------|--------------------|--------------------|---------------------|--|
| Trips Destined | | 10 | 20 | | |
| 1 | 15 | 37.50 | 12 | 49.50 | 0.303 |
| 2 | 15 | 6 | 75 | 81 | 0.185 |
| Column Total | | 43.50 | 87 | | |

Balancing Iteration 2 (

| Origin Zone | Trips Originating | Destination Zone 1 | Destination Zone 2 | Row Total T'_i | Normalizing Factor $N_i = T_i/T'_i$ |
|--|-------------------|--------------------|--------------------|------------------|--|
| Trips Destined | | 10 | 20 | | |
| 1 | 15 | 11.36 | 3.64 | 15.00 | 1.00 |
| 2 | 15 | 1.11 | 13.89 | 15.00 | 1.00 |
| Column Total | | 12.47 | 17.53 | | |
| Normalizing Factor $N_j = T_j/T'_j$ | | 0.802 | 1.141 | | |

Balancing Iteration 3 (

| Origin Zone | Trips Originating | Destination Zone 1 | Destination Zone 2 | Row Total T'_i | Normalizing Factor $N_i = T_i/T'_i$ |
|--|-------------------|--------------------|--------------------|------------------|--|
| Trips Destined | | 10 | 20 | | |
| 1 | 15 | 9.11 | 4.15 | 13.26 | 1.13 |
| 2 | 15 | 0.89 | 15.85 | 16.74 | 0.90 |
| Column Total | | 10.00 | 20.00 | | |
| Normalizing Factor $N_j = T_j/T'_j$ | | 1.00 | 1.00 | | |

Balancing Iteration 4 (

| Origin Zone | Trips Originating | Destination Zone 1 | Destination Zone 2 | Row Total T'_i | Normalizing Factor $N_i = T_i/T'_i$ |
|--|-------------------|--------------------|--------------------|------------------|--|
| Trips Destined | | 10 | 20 | | |
| 1 | 15 | 10.31 | 4.69 | 15.00 | 1.00 |
| 2 | 15 | 0.80 | 14.20 | 15.00 | 1.00 |
| Column Total | | 11.10 | 18.90 | | |
| Normalizing Factor $N_j = T_j/T'_j$ | | 0.90 | 1.06 | | |

...

Balancing Iteration 16 (

| Origin Zone | Trips Originating | Destination Zone 1 | Destination Zone 2 | Row Total T'_i | Normalizing Factor $N_i = T_i/T'_i$ |
|--|--------------------------|---------------------------|---------------------------|------------------------------------|---|
| Trips Destined | | 10 | 20 | | |
| 1 | 15 | 9.39 | 5.61 | 15.00 | 1.00 |
| 2 | 15 | 0.62 | 14.38 | 15.00 | 1.00 |
| Column Total | | 10.01 | 19.99 | | |
| Normalizing Factor = $N_j = T_j/T'_j$ | | 1.00 | 1.00 | | |

So while the matrix is not strictly balanced, it is very close, well within a 1% threshold, after 16 iterations.

External Exercises

Use the ADAM software at the STREET website ^[1] and try Assignment #2 to learn how changes in link characteristics adjust the distribution of trips throughout a network.

Further Reading

- Background

References

[1] <http://street.umn.edu/>

Fundamentals of Transportation/ Destination Choice/Background

Some additional background on Fundamentals of Transportation/Destination Choice

History

Over the years, modelers have used several different formulations of trip distribution. The first was the Fratar or Growth model (which did not differentiate trips by purpose). This structure extrapolated a base year trip table to the future based on growth, but took no account of changing spatial accessibility due to increased supply or changes in travel patterns and congestion.

The next models developed were the gravity model and the intervening opportunities model. The most widely used formulation is still the gravity model.

While studying traffic in Baltimore, Maryland, Alan Voorhees developed a mathematical formula to predict traffic patterns based on land use. This formula has been instrumental in the design of numerous transportation and public works projects around the world. He wrote "A General Theory of Traffic Movement," (Voorhees, 1956) which applied the gravity model to trip distribution, which translates trips generated in an area to a matrix that identifies the number of trips from each origin to each destination, which can then be loaded onto the network.

Evaluation of several model forms in the 1960s concluded that "the gravity model and intervening opportunity model proved of about equal reliability and utility in simulating the 1948 and 1955 trip distribution for Washington, D.C." (Heanue and Pyers 1966). The Fratar model was shown to have weakness in areas experiencing land use changes. As comparisons between the models showed that either could be calibrated equally well to match observed conditions, because of computational ease, gravity models became more widely spread than intervening opportunities models. Some theoretical problems with the intervening opportunities model were discussed by Whitaker and West (1968) concerning its inability to account for all trips generated in a zone which makes it more difficult to calibrate, although techniques for dealing with the limitations have been developed by Ruiter (1967).

With the development of logit and other discrete choice techniques, new, demographically disaggregate approaches to travel demand were attempted. By including variables other than travel time in determining the probability of making a trip, it is expected to have a better prediction of travel behavior. The logit model and gravity model have been shown by Wilson (1967) to be of essentially the same form as used in statistical mechanics, the entropy maximization model. The application of these models differ in concept in that the gravity model uses impedance by travel time, perhaps stratified by socioeconomic variables, in determining the probability of trip making, while a discrete choice approach brings those variables inside the utility or impedance function. Discrete choice models require more information to estimate and more computational time.

Ben-Akiva and Lerman (1985) have developed combination destination choice and mode choice models using a logit formulation for work and non-work trips. Because of computational intensity, these formulations tended to aggregate traffic zones into larger

districts or rings in estimation. In current application, some models, including for instance the transportation planning model used in Portland, Oregon use a logit formulation for destination choice. Allen (1984) used utilities from a logit based mode choice model in determining composite impedance for trip distribution. However, that approach, using mode choice log-sums implies that destination choice depends on the same variables as mode choice. Levinson and Kumar (1995) employ mode choice probabilities as a weighting factor and develops a specific impedance function or "f-curve" for each mode for work and non-work trip purposes.

Mathematics

At this point in the transportation planning process, the information for zonal interchange analysis is organized in an origin-destination table. On the left is listed trips produced in each zone. Along the top are listed the zones, and for each zone we list its attraction. The table is $n \times n$, where $n =$ the number of zones.

Each cell in our table is to contain the number of trips from zone i to zone j . We do not have these within cell numbers yet, although we have the row and column totals. With data organized this way, our task is to fill in the cells for tables headed $t=1$ through say $t=n$.

Actually, from home interview travel survey data and attraction analysis we have the cell information for $t = 1$. The data are a sample, so we generalize the sample to the universe. The techniques used for zonal interchange analysis explore the empirical rule that fits the $t = 1$ data. That rule is then used to generate cell data for $t = 2, t = 3, t = 4$, etc., to $t = n$.

The first technique developed to model zonal interchange involves a model such as this:

$$T_{ij} = T_i \frac{T_j f(C_{ij}) K_{ij}}{\sum_{j=1}^n T_j f(C_{ij}) K_{ij}}$$

where:

- T_{ij} : trips from i to j .
- T_i : trips from i , as per our generation analysis
- T_j : trips attracted to j , as per our generation analysis
- $f(C_{ij})$: travel cost friction factor, say = C_{ij}^b
- K_{ij} : Calibration parameter

Zone i generates T_i trips; how many will go to zone j ? That depends on the attractiveness of j compared to the attractiveness of all places; attractiveness is tempered by the distance a zone is from zone i . We compute the fraction comparing j to all places and multiply T_i by it.

The rule is often of a gravity form:

$$T_{ij} = a \frac{P_i P_j}{C_{ij}^b}$$

where:

- P_i, P_j : populations of i and j
- a, b : parameters

But in the zonal interchange mode, we use numbers related to trip origins (T_i) and trip destinations (T_j) rather than populations.

There are lots of model forms because we may use weights and special calibration parameters, e.g., one could write say:

$$T_{ij} = a \frac{T_i^c T_j^d}{C_{ij}^b}$$

or $T_{ij} = \frac{c T_i d T_j}{C_{ij}^b}$

where:

- a, b, c, d are parameters
- C_{ij} : travel cost (e.g. distance, money, time)
- T_j : inbound trips, destinations
- T_i : outbound trips, origin

Entropy Analysis

Wilson (1970) gives us another way to think about zonal interchange problem. This section treats Wilson's methodology to give a grasp of central ideas. To start, consider some trips where we have seven people in origin zones commuting to seven jobs in destination zones. One configuration of such trips will be:

Table: Configuration of Trips

| zone | 1 | 2 | 3 |
|------|---|---|---|
| 1 | 2 | 1 | 1 |
| 2 | 0 | 2 | 1 |

$$w(T_{ij}) = \frac{7!}{2!1!1!0!2!1!} = 1260$$

where $0! = 1$ That configuration can appear in 1,260 ways. We have calculated the number of ways that configuration of trips might have occurred, and to explain the calculation, let's recall those coin tossing experiments talked about so much in elementary statistics. The number of ways a two-sided coin can come up is 2^n , where n is the number of times we toss the coin. If we toss the coin once, it can come up heads or tails, $2^1 = 2$. If we toss it twice, it can come up HH, HT, TH, or TT, 4 ways, and $2^2 = 4$. To ask the specific question about, say, four coins coming up all heads, we calculate $4!/4!0! = 1$. Two heads and two tails would be $4!/2!2! = 6$. We are solving the equation:

$$w = \frac{n!}{\prod_{i=1}^n n_i!}$$

An important point is that as n gets larger, our distribution gets more and more peaked, and it is more and more reasonable to think of a most likely state.

However, the notion of most likely state comes not from this thinking; it comes from statistical mechanics, a field well known to Wilson and not so well known to transportation planners. The result from statistical mechanics is that a descending series is most likely. Think about the way the energy from lights in the classroom is affecting the air in the classroom. If the effect resulted in an ascending series, many of the atoms and molecules would be affected a lot and a few would be affected a little. The descending series would have a lot affected not at all or not much and only a few affected very much. We could take a given level of energy and compute excitation levels in ascending and descending series. Using the formula above, we would compute the ways particular series could occur, and we would conclude that descending series dominate.

That's more or less Boltzmann's Law,

$$p_j = p_0 e^{\beta e_j}$$

That is, the particles at any particular excitation level, j , will be a negative exponential function of the particles in the ground state, p_0 , the excitation level, e_j , and a parameter β , which is a function of the (average) energy available to the particles in the system. The two paragraphs above have to do with ensemble methods of calculation developed by Gibbs, a topic well beyond the reach of these notes.

Returning to our O-D matrix, note that we have not used as much information as we would have from an O and D survey and from our earlier work on trip generation. For the same travel pattern in the O-D matrix used before, we would have row and column totals, i.e.:

Table: Illustrative O-D Matrix with row and column totals

| | zone | 1 | 2 | 3 |
|------|---------------------|---|---|---|
| zone | $T_i \setminus T_j$ | 2 | 3 | 2 |
| 1 | 4 | 2 | 1 | 1 |
| 2 | 3 | 0 | 2 | 1 |

Consider the way the four folks might travel, $4!/2!1!1! = 12$; consider three folks, $3!/0!2!1! = 3$. All travel can be combined in $12*3 = 36$ ways. The possible configuration of trips is, thus, seen to be much constrained by the column and row totals.

We put this point together with the earlier work with our matrix and the notion of most likely state to say that we want to

$$\max w(T_{ij}) = \frac{T!}{\prod_{ij} T_{ij}!}$$

subject to

$$\sum_j T_{ij} = T_i; \sum_i T_{ij} = T_j$$

$$\text{where: } T = \sum_j \sum_i T_{ij} = \sum_i T_i = \sum_j T_j$$

and this is the problem that we have solved above.

Wilson adds another consideration; he constrains the system to the amount of energy available (i.e., money), and we have the additional constraint,

$$\sum_i \sum_j T_{ij} C_{ij} = C$$

where C is the quantity of resources available and C_{ij} is the travel cost from i to j .

The discussion thus far contains the central ideas in Wilson's work, but we are not yet to the place where the reader will recognize the model as it is formulated by Wilson.

First, writing the function to be maximized using Lagrangian multipliers, we have:

$$\frac{T!}{\prod_{ij} T_{ij}!} + \sum_i \lambda_i \left(T_i - \sum_j T_{ij} \right) + \sum_j \lambda_j \left(T_j - \sum_i T_{ij} \right) + \beta \left(C - \sum_i \sum_j T_{ij} C_{ij} \right)$$

where λ_i, λ_j , and β are the Lagrange multipliers, β having an energy sense.

Second, it is convenient to maximize the natural log (\ln) rather than $w(T_{ij})$, for then we may use Stirling's approximation.

$$\ln N! \approx N \ln N - N$$

$$\text{so } \frac{\partial \ln N!}{\partial N} \approx \ln N$$

Third, evaluating the maximum, we have

$$\frac{\partial T}{\partial T_{ij}} = -\ln T_{ij} - \lambda_i - \lambda_j - \beta C_{ij} = 0$$

with solution

$$\ln T_{ij} = -\lambda_i - \lambda_j - \beta C_{ij}$$

$$T_{ij} = e^{-\lambda_i - \lambda_j - \beta C_{ij}}$$

Finally, substituting this value of T_{ij} back into our constraint equations, we have:

$$\sum_j e^{-\lambda_i - \lambda_j - \beta C_{ij}} = 0 \quad \sum_i e^{-\lambda_i - \lambda_j - \beta C_{ij}} = 0$$

and, taking the constant multiples outside of the summation sign

$$e^{-\lambda_j} = \frac{T_i}{\sum_j e^{-\lambda_j - \beta C_{ij}}}; \quad e^{-\lambda_i} = \frac{T_j}{\sum_i e^{-\lambda_i - \beta C_{ij}}}$$

$$\text{let } \frac{e^{-\lambda_j}}{T_i} = A_i; \quad \frac{e^{-\lambda_i}}{T_j} = B_j$$

$$\text{we have } T_{ij} = A_i B_j T_i T_j e^{-\beta C_{ij}}$$

which says that the most probable distribution of trips has a gravity model form, T_{ij} is proportional to trip origins and destinations. A_i , B_j , and β ensure constraints are met.

Turning now to computation, we have a large problem. First, we do not know the value of C , which earlier on we said had to do with the money available, it was a cost constraint. Consequently, we have to set β to different values and then find the best set of values for A_i and B_j . We know what β means - the greater the value of β , the less the cost of average distance traveled. (Compare β in Boltzmann's Law noted earlier.) Second, the values of α_i and β_j depend on each other. So for each value of β , we must use an iterative solution. There are computer programs to do this.

Wilson's method has been applied to the Lowry model.

References

- Allen, B. 1984 Trip Distribution Using Composite Impedance Transportation Research Record 944 pp. 118-127
- Barnes, G. and Davis, G. 2000. Understanding Urban Travel Demand: Problems, Solutions, and the Role of Forecasting, University of Minnesota Center for Transportation Studies: Transportation and Regional Growth Study
- Ben-Akiva M. and Lerman S. 1985 Discrete Choice Analysis, MIT Press, Cambridge MA
- Boyce, D., Lupa, M. and Zhang, Y.F. 1994 Introducing "Feedback" into the Four-Step Travel Forecasting Procedure vs. the Equilibrium Solution of a Combined Model presented at 73rd Annual Meeting of Transportation Research Board
- Evans, Suzanne P. 1976 . Derivation and Analysis of Some Models for Combining Trip Distribution and Assignment. Transportation Research, Vol. 10, PP 37-57 1976
- Florian M., Nguyen S., and Ferland J. 1975 On the Combined Distribution-Assignment of Traffic", Transportation Science, Vol. 9, pp. 43-53, 1975

- Haney, D. 1972 Consistency in Transportation Demand and Evaluation Models, Highway Research Record 392, pp. 13-25 1972
- Hansen, W. G. 1959. How accessibility shapes land use. Journal of the American Institute of Planners, 25(2), 73-76.
- Heanue, Kevin E. and Pyers, Clyde E. 1966. A Comparative Evaluation of Trip Distribution Procedures,
- Levinson, D. and A. Kumar 1994 The Rational Locator: Why Travel Times Have Remained Stable, Journal of the American Planning Association, 60:3 319-332
- Levinson, D. and Kumar A. 1995. A Multi-modal Trip Distribution Model. Transportation Research Record #1466: 124-131.
- Portland MPO Report to Federal Transit Administration on Transit Modeling
- Reilly, W.J. 1929 "Methods for the Study of Retail Relationships" University of Texas Bulletin No 2944, Nov. 1929.
- Reilly, W.J., 1931 The Law of Retail Gravitation, New York.
- Ruiter, E. 1967 Improvements in Understanding, Calibrating, and Applying the Opportunity Model Highway Research Record No. 165 pp. 1-21
- Stewart, J.Q. 1948 "Demographic Gravitation: Evidence and Application" Sociometry Vol. XI Feb.-May 1948 pp 31-58.
- Stewart, J.Q., 1947. Empirical Mathematical Rules Concerning the Distribution and Equilibrium of Population, Geographical Review, Vol 37, 461-486.
- Stewart, J.Q., 1950. Potential of Population and its Relationship to Marketing. In: Theory in Marketing , R. Cox and W. Alderson (Eds) (Richard D. Irwin, Inc., Homewood, Illinois).
- Stewart, J.Q., 1950. The Development of Social Physics, American Journal of Physics, Vol 18, 239-253
- Voorhees, Alan M., 1956, "A General Theory of Traffic Movement," 1955 Proceedings, Institute of Traffic Engineers, New Haven, Connecticut.
- Whitaker, R. and K. West 1968 The Intervening Opportunities Model: A Theoretical Consideration Highway Research Record 250 pp.1-7
- Wilson, A.G. A Statistical Theory of Spatial Distribution Models Transportation Research, Volume 1, pp. 253-269 1967
- Wohl, M. 1963 Demand, Cost, Price and Capacity Relationships Applied to Travel Forecasting. Highway Research Record 38:40-54
- Zipf, G.K., 1946. The Hypothesis on the Intercity Movement of Persons. American Sociological Review, vol. 11, Oct
- Zipf, G.K., 1949. Human Behaviour and the Principle of Least Effort. Massachusetts

Fundamentals of Transportation/ Mode Choice

Mode choice analysis is the third step in the conventional four-step transportation forecasting model, following Trip Generation and Destination Choice but before Route Choice. While trip distribution's zonal interchange analysis yields a set of origin destination tables which tells where the trips will be made, mode choice analysis allows the modeler to determine what mode of transport will be used.

The early transportation planning model developed by the Chicago Area Transportation Study (CATS) focused on transit, it wanted to know how much travel would continue by transit. The CATS divided transit trips into two classes: trips to the CBD (mainly by subway/elevated transit, express buses, and commuter trains) and other (mainly on the local bus system). For the latter, increases in auto ownership and use were trade off against bus use; trend data were used. CBD travel was analyzed using historic mode choice data together with projections of CBD land uses. Somewhat similar techniques were used in many studies. Two decades after CATS, for example, the London study followed essentially the same procedure, but first dividing trips into those made in inner part of the city and those in the outer part. This procedure was followed because it was thought that income (resulting in the purchase and use of automobiles) drove mode choice.

Diversion Curve techniques

The CATS had diversion curve techniques available and used them for some tasks. At first, the CATS studied the diversion of auto traffic from streets and arterial to proposed expressways. Diversion curves were also used as bypasses were built around cities to establish what percentage of the traffic would use the bypass. The mode choice version of diversion curve analysis proceeds this way: one forms a ratio, say:

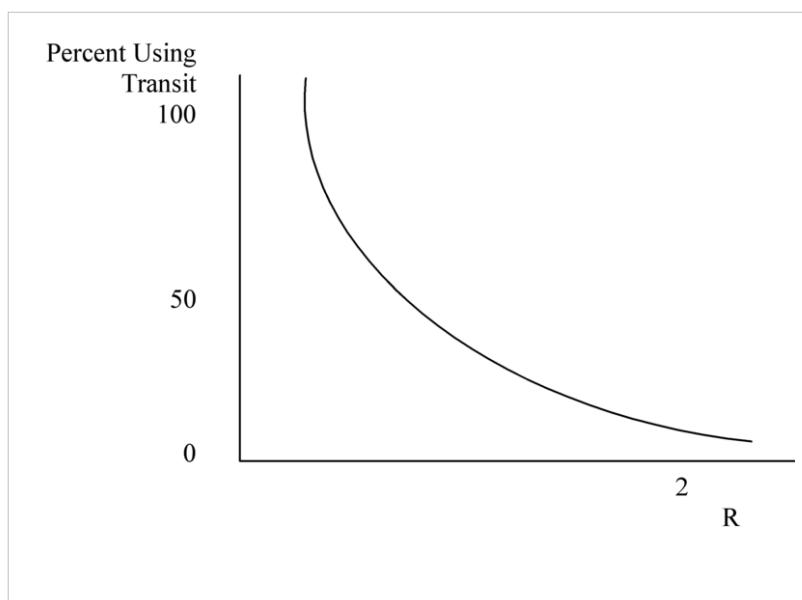
$$\frac{c_{transit}}{c_{auto}} = R$$

where:

c_m = travel time by mode m and

R is empirical data in the form:

Given the R that we have calculated, the graph tells us the percent of users in the market that will choose transit. A variation on the technique is to use costs rather than time in the diversion ratio. The decision to use a time or cost ratio turns on the problem at hand. Transit agencies developed diversion curves for different kinds of situations, so variables like income and population density entered implicitly.



Diversion curves are based on empirical observations, and their improvement has resulted from better (more and more pointed) data. Curves are available for many markets. It is not difficult to obtain data and array results. Expansion of transit has motivated data development by operators and planners. Yacov Zahavi's UMOT studies contain many examples of diversion curves.

In a sense, diversion curve analysis is expert system analysis. Planners could "eyeball" neighborhoods and estimate transit ridership by routes and time of day. Instead, diversion is observed empirically and charts can be drawn.

Disaggregate Travel Demand models

Travel demand theory was introduced in the appendix on traffic generation. The core of the field is the set of models developed following work by Stan Warner in 1962 (*Strategic Choice of Mode in Urban Travel: A Study of Binary Choice*). Using data from the CATS, Warner investigated classification techniques using models from biology and psychology. Building from Warner and other early investigators, disaggregate demand models emerged. Analysis is disaggregate in that individuals are the basic units of observation, yet aggregate because models yield a single set of parameters describing the choice behavior of the population. Behavior enters because the theory made use of consumer behavior concepts from economics and parts of choice behavior concepts from psychology. Researchers at the University of California, Berkeley (especially Daniel McFadden, who won a Nobel Prize in Economics for his efforts) and the Massachusetts Institute of Technology (Moshe Ben-Akiva) (and in MIT associated consulting firms, especially Cambridge Systematics) developed what has become known as choice models, direct demand models (DDM), Random Utility Models (RUM) or, in its most used form, the multinomial logit model (MNL).

Choice models have attracted a lot of attention and work; the Proceedings of the International Association for Travel Behavior Research chronicles the evolution of the models. The models are treated in modern transportation planning and transportation engineering textbooks.

One reason for rapid model development was a felt need. Systems were being proposed (especially transit systems) where no empirical experience of the type used in diversion curves was available. Choice models permit comparison of more than two alternatives and the importance of attributes of alternatives. There was the general desire for an analysis technique that depended less on aggregate analysis and with a greater behavioral content. And, there was attraction too, because choice models have logical and behavioral roots extended back to the 1920s as well as roots in Kelvin Lancaster's consumer behavior theory, in utility theory, and in modern statistical methods.

Psychological roots

Early psychology work involved the typical experiment: Here are two objects with weights, w_1 and w_2 , which is heavier? The finding from such an experiment would be that the greater the difference in weight, the greater the probability of choosing correctly. Graphs similar to the one on the right result.

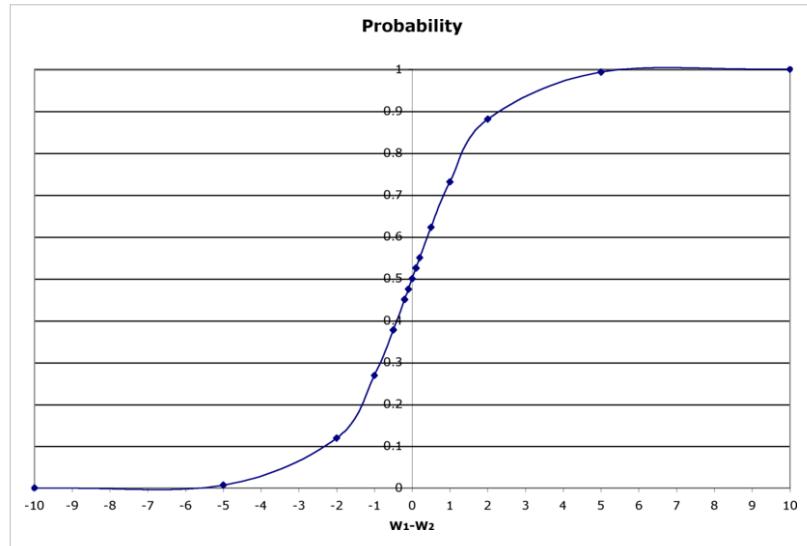
Louis Leon Thurstone proposed (in the 1920s) that perceived weight,

$$w = v + e,$$

where v is the true weight and e is random with

$$E(e) = 0.$$

The assumption that e is normally and identically distributed (NID) yields the binary probit model.



Econometric formulation

Economists deal with utility rather than physical weights, and say that

$$\text{observed utility} = \text{mean utility} + \text{random term}.$$

The characteristics of the object, x , must be considered, so we have

$$u(x) = v(x) + e(x).$$

If we follow Thurston's assumption, we again have a probit model.

An alternative is to assume that the error terms are independently and identically distributed with a Weibull, Gumbel Type I, or double exponential distribution (They are much the same, and differ slightly in their tails (thicker) from the normal distribution). This yields the multinomial logit model (MNL). Daniel McFadden argued that the Weibull had desirable properties compared to other distributions that might be used. Among other things, the error terms are normally and identically distributed. The logit model is simply a log ratio of the probability of choosing a mode to the probability of not choosing a mode.

$$\log \left(\frac{P_i}{1 - P_i} \right) = v(x_i)$$

Observe the mathematical similarity between the logit model and the S-curves we estimated earlier, although here share increases with utility rather than time. With a choice model we are explaining the share of travelers using a mode (or the probability that an individual traveler uses a mode multiplied by the number of travelers).

The comparison with S-curves is suggestive that modes (or technologies) get adopted as their utility increases, which happens over time for several reasons. First, because the utility itself is a function of network effects, the more users, the more valuable the service, higher the utility associated with joining the network. Second, because utility increases as user costs drop, which happens when fixed costs can be spread over more users (another network effect). Third, technological advances, which occur over time and as the number of users increases, drive down relative cost.

An illustration of a utility expression is given:

$$\log \left(\frac{P_A}{1 - P_A} \right) = \beta_0 + \beta_1 (c_A - c_T) + \beta_2 (t_A - t_T) + \beta_3 I + \beta_4 N = v_A$$

where

P_i = Probability of choosing mode i.

P_A = Probability of taking auto

c_A, c_T = cost of auto, transit

t_A, t_T = travel time of auto, transit

I = income

N = Number of travelers

With algebra, the model can be translated to its most widely used form:

$$\begin{aligned} \frac{P_A}{1 - P_A} &= e^{v_A} \\ P_A &= e^{v_A} - P_A e^{v_A} \\ P_A (1 + e^{v_A}) &= e^{v_A} \\ P_A &= \frac{e^{v_A}}{1 + e^{v_A}} \end{aligned}$$

It is fair to make two conflicting statements about the estimation and use of this model:

1. It's a "house of cards", and
2. Used by a technically competent and thoughtful analyst, it's useful.

The "house of cards" problem largely arises from the utility theory basis of the model specification. Broadly, utility theory assumes that (1) users and suppliers have perfect information about the market; (2) they have deterministic functions (faced with the same options, they will always make the same choices); and (3) switching between alternatives is costless. These assumptions don't fit very well with what is known about behavior. Furthermore, the aggregation of utility across the population is impossible since there is no universal utility scale.

Suppose an option has a net utility u_{jk} (option k , person j). We can imagine that having a systematic part v_{jk} that is a function of the characteristics of an object and person j , plus a random part e_{jk} , which represents tastes, observational errors, and a bunch of other things (it gets murky here). (An object such as a vehicle does not have utility, it is characteristics

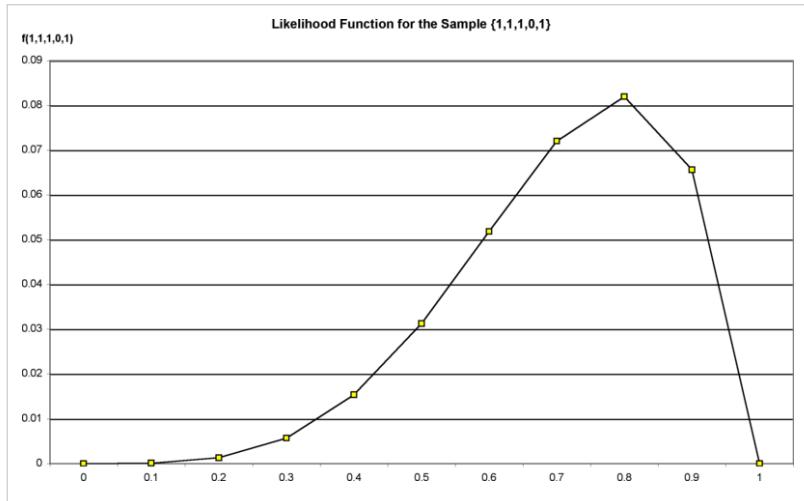
of a vehicle that have utility.) The introduction of e lets us do some aggregation. As noted above, we think of observable utility as being a function:

$$v_A = \beta_0 + \beta_1(c_A - c_T) + \beta_2(t_A - t_T) + \beta_3 I + \beta_4 N$$

where each variable represents a characteristic of the auto trip. The value β_0 is termed an alternative specific constant. Most modelers say it represents characteristics left out of the equation (e.g., the political correctness of a mode, if I take transit I feel morally righteous, so β_0 may be negative for the automobile), but it includes whatever is needed to make error terms NID.

Econometric estimation

Turning now to some technical matters, how do we estimate $v(x)$? Utility ($v(x)$) isn't observable. All we can observe are choices (say, measured as 0 or 1), and we want to talk about probabilities of choices that range from 0 to 1. (If we do a regression on 0s and 1s we might measure for j a probability of 1.4 or -0.2 of taking an auto.) Further, the distribution of the error terms wouldn't have appropriate statistical characteristics.



The MNL approach is to make a maximum likelihood estimate of this functional form. The likelihood function is:

$$L^* = \prod_{n=1}^N f(y_n | x_n, \theta)$$

we solve for the estimated parameters

$$\hat{\theta}$$

that max L*. This happens when:

$$\frac{\partial L}{\partial \hat{\theta}_N} = 0$$

The log-likelihood is easier to work with, as the products turn to sums:

$$\ln L^* = \sum_{n=1}^N \ln f(y_n | x_n, \theta)$$

Consider an example adopted from John Bitzan's Transportation Economics Notes. Let X be a binary variable that is *gamma* and 0 with probability (1- *gamma*). Then $f(0) = (1- \gamma)$ and $f(1) = \gamma$. Suppose that we have 5 observations of X , giving the sample $\{1,1,1,0,1\}$. To find the maximum likelihood estimator of *gamma* examine various values of *gamma*, and for these values determine the probability of drawing the sample $\{1,1,1,0,1\}$. If *gamma* takes the value 0, the probability of drawing our sample is 0. If *gamma* is 0.1, then the probability of getting our sample is: $f(1,1,1,0,1) = f(1)f(1)f(1)f(0)f(1) = 0.1 * 0.1 * 0.1 * 0.9 * 0.1 = 0.00009$. We can compute the probability of obtaining our sample over

a range of *gamma* - this is our likelihood function. The likelihood function for n independent observations in a logit model is

$$L^* = \prod_{n=1}^N P_i^{Y_i} (1 - P_i)^{1-Y_i}$$

where: $Y_i = 1$ or 0 (choosing e.g. auto or not-auto) and P_i = the probability of observing $Y_i=1$

The log likelihood is thus:

$$\ell = \ln L^* = \sum_{i=1}^n [Y_i \ln P_i + (1 - Y_i) \ln (1 - P_i)]$$

In the binomial (two alternative) logit model,

$$P_{auto} = \frac{e^{v(x_{auto})}}{1 + e^{v(x_{auto})}}, \text{ so}$$

$$\ell = \ln L^* = \sum_{i=1}^n \left[Y_i v(x_{auto}) - \ln \left(1 + e^{v(x_{auto})} \right) \right]$$

The log-likelihood function is maximized setting the partial derivatives to zero:

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^n (Y_i - \hat{P}_i) = 0$$

The above gives the essence of modern MNL choice modeling.

Independence of Irrelevant Alternatives (IIA)

Independence of Irrelevant Alternatives is a property of Logit, but not all Discrete Choice models. In brief, the implication of IIA is that if you add a mode, it will draw from present modes in proportion to their existing shares. (And similarly, if you remove a mode, its users will switch to other modes in proportion to their previous share). To see why this property may cause problems, consider the following example: Imagine we have seven modes in our logit mode choice model (drive alone, carpool 2 passenger, carpool 3+ passenger, walk to transit, auto driver to transit (park and ride), auto passenger to transit (kiss and ride), and walk or bike). If we eliminated Kiss and Ride, a disproportionate number may use Park and Ride or carpool.

Consider another example. Imagine there is a mode choice between driving and taking a red bus, and currently each has 50% share. If we introduce another mode, let's call it a blue bus with identical attributes to the red bus, the logit mode choice model would give each mode 33.3% of the market, or in other words, buses will collectively have 66.7% market share. Logically, if the mode is truly identical, it would not attract any additional passengers (though one can imagine scenarios where adding capacity would increase bus mode share, particularly if the bus was capacity constrained).

There are several strategies that help with the IIA problem. Nesting of choices allows us to reduce this problem. However, there is an issue of the proper Nesting structure. Other alternatives include more complex models (e.g. Mixed Logit) which are more difficult to estimate.

Consumers' Surplus

Topics not touched on include the “red bus, blue bus” problem; the use of nested models (e.g., estimate choice between auto and transit, and then estimate choice between rail and bus transit); how consumers’ surplus measurements may be obtained; and model estimation, goodness of fit, etc. For these topics see a textbook such as Ortuzar and Willumsen (2001).

Returning to roots

The discussion above is based on the economist's utility formulation. At the time MNL modeling was developed there was some attention to psychologist's choice work (e.g., Luce's choice axioms discussed in his Individual Choice Behavior, 1959). It has an analytic side in computational process modeling. Emphasis is on how people think when they make choices or solve problems (see Newell and Simon 1972). Put another way, in contrast to utility theory, it stresses not the choice but the way the choice was made. It provides a conceptual framework for travel choices and agendas of activities involving considerations of long and short term memory, effectors, and other aspects of thought and decision processes. It takes the form of rules dealing with the way information is searched and acted on. Although there is a lot of attention to behavioral analysis in transportation work, the best of modern psychological ideas are only beginning to enter the field. (e.g. Golledge, Kwan and Garling 1984; Garling, Kwan, and Golledge 1994).

Examples

Example 1: Mode Choice Model



Problem:

You are given this mode choice model

$$U_{ijm} = -0.412(C_c/w) - 0.0201*C_{ivt} - 0.0531*C_{ovt} - 0.89*D_1 - 1.78*D_3 - 2.15*D_4$$

Where:

- C_c/w = cost of mode (cents) / wage rate (in cents per minute)
- C_{ivt} = travel time in-vehicle (min)
- C_{ovt} = travel time out-of-vehicle (min)
- D = mode specific dummies: (dummies take the value of 1 or 0)
 - D_1 = driving,
 - D_2 = transit with walk access, [base mode]
 - D_3 = transit with auto access,
 - D_4 = carpool

With these inputs:

| | Driving | Walk Connect Transit | Auto Connect Transit | Carpool |
|----------------------------------|---------|-------------------------|-------------------------|---------|
| t = travel time in-vehicle (min) | 10 | 30 | 15 | 12 |

| | | | | |
|---|----|-----|-----|-----|
| t_0 = travel time out-of-vehicle (min) | 0 | 15 | 10 | 3 |
| D_1 = driving, | 1 | 0 | 0 | 0 |
| D_2 = transit with walk access, [base mode] | 0 | 1 | 0 | 0 |
| D_3 = transit with auto access, | 0 | 0 | 1 | 0 |
| D_4 = carpool | 0 | 0 | 0 | 1 |
| COST | 25 | 100 | 100 | 150 |
| WAGE | 60 | 60 | 60 | 60 |

What are the resultant mode shares?



Solution:

| Outputs | 1: Driving | 2: Walk-connect transit | 3: Auto-connect transit | 4: Carpool | Sum |
|-----------|------------|-------------------------|-------------------------|------------|------|
| Utilities | -1.26 | -2.09 | -3.30 | -3.58 | |
| EXP(V) | 0.28 | 0.12 | 0.04 | 0.03 | |
| P(V) | 59.96% | 26.31% | 7.82% | 5.90% | 100% |

Interpretation

Value of Time:

$$.0411/2.24 = \$0.0183/min = \$1.10/hour$$

(in 1967 \$, when the wage rate was about \$2.85/hour)

implication, if you can improve the travel time (by more buses, less bottlenecks, e.g.) for less than \$1.10/hour/person, then it is socially worthwhile.

Example 2: Mode Choice Model Interpretation



Case 1

| " | Bus | Car | Parameter |
|-------|--------|--------|-----------|
| T_w | 10 min | 5 min | -0.147 |
| T_t | 40 min | 20 min | -0.0411 |
| C | \$2 | \$1 | -2.24 |

Car always wins (independent of parameters as long as all are < 0)



Case 1

| | Bus | Car | Parameter |
|---------|------------|------------|------------------|
| Tw | 5 min | 5 min | -0.147 |
| Tt | 40 min | 20 min | -0.0411 |
| C | \$2 | \$4 | -2.24 |
| Results | -6.86 | -10.51 | |

Under observed parameters, bus always wins, but not necessarily under all parameters.

It is important to note that individuals differ in parameters. We could introduce socio-economic and other observable characteristics as well as a stochastic error term to make the problem more realistic.

Sample Problem

- Problem (Solution)

Variables

- U_{ijm} - Utility of traveling from i to j by mode m
- D_n = mode specific dummies: (dummies take the value of 1 or 0)
- P_m = Probability of mode m
- C_c/w = cost of mode (cents) / wage rate (in cents per minute)
- C_{ivt} = travel time in-vehicle (min)
- C_{ovt} = travel time out-of-vehicle (min)
- D_n = mode specific dummies: (dummies take the value of 1 or 0)

Abbreviations

- WCT - walk connected transit
- ADT - auto connect transit (drive alone/park and ride)
- APT - auto connect transit (auto passenger/kiss and ride)
- AU1 - auto driver (no passenger)
- AU2 - auto 2 occupants
- AU3+ - auto 3+ occupants
- WK/BK - walk/bike
- IIA - Independence of Irrelevant Alternatives

Key Terms

- Mode choice
- Logit
- Probability
- Independence of Irrelevant Alternatives (IIA)
- Dummy Variable (takes value of 1 or 0)

References

- Garling, Tommy Mei Po Kwan, and Reginald G. Golledge. Household Activity Scheduling, Transportation Research, 22B, pp. 333-353. 1994.
- Golledge. Reginald G., Mei Po Kwan, and Tommy Garling, "Computational Process Modeling of Household Travel Decisions," Papers in Regional Science, 73, pp. 99-118. 1984.
- Lancaster, K.J., A new approach to consumer theory. Journal of Political Economy, 1966. 74(2): p. 132-157.
- Luce, Duncan R. (1959). Individual choice behavior, a theoretical analysis. New York, Wiley.
- Newell, A. and Simon, H. A. (1972). Human Problem Solving. Englewood Cliffs, NJ: Prentice Hall.
- Ortuzar, Juan de Dios and L. G. Willumsen's Modelling Transport. 3rd Edition. Wiley and Sons. 2001,
- Thurstone, L.L. (1927). A law of comparative judgement. Psychological Review, 34, 278-286.
- Warner, Stan 1962 Strategic Choice of Mode in Urban Travel: A Study of Binary Choice

Fundamentals of Transportation/ Mode Choice/Problem



Problem:

You are given the following mode choice model.

$$U_{ijm} = -1C_{ijm} + 5D_T$$

Where:

- C_{ijm} = travel cost between i and j by mode m
- D_T = dummy variable (alternative specific constant) for transit and Travel Times

Auto Travel Times

| Origin\Destination | Dakotopolis | New Fargo |
|--------------------|-------------|-----------|
| Dakotopolis | 5 | 7 |
| New Fargo | 7 | 5 |

Transit Travel Times

| Origin\Destination | Dakotopolis | New Fargo |
|--------------------|-------------|-----------|
| Dakotopolis | 10 | 15 |
| New Fargo | 15 | 8 |

- A. Using a logit model, determine the probability of a traveler driving.
- B. Using the results from the previous problem (#2), how many car trips will there be?
- Solution

Fundamentals of Transportation/ Mode Choice/Solution


Problem:

You are given the following mode choice model.

$$U_{ijm} = -1C_{ijm} + 5D_T$$

Where:

- C_{ijm} = travel cost between i and j by mode m
- D_T = dummy variable (alternative specific constant) for transit

Auto Travel Times

| Origin\Destination | Dakotopolis | New Fargo |
|--------------------|-------------|-----------|
| Dakotopolis | 5 | 7 |
| New Fargo | 7 | 5 |

Transit Travel Times

| Origin\Destination | Dakotopolis | New Fargo |
|--------------------|-------------|-----------|
| Dakotopolis | 10 | 15 |
| New Fargo | 15 | 8 |


Solution:
Part A

- A. Using a logit model, determine the probability of a traveler driving.

Solution Steps

1. Compute Utility for Each Mode for Each Cell
2. Compute Exponentiated Utilities for Each Cell
3. Sum Exponentiated Utilities
4. Compute Probability for Each Mode for Each Cell
5. Multiply Probability in Each Cell by Number of Trips in Each Cell

Auto Utility:

| Origin\Destination | Dakotopolis | New Fargo |
|---------------------------|--------------------|------------------|
| Dakotopolis | -5 | -7 |
| New Fargo | -7 | -5 |

Transit Utility:

| Origin\Destination | Dakotopolis | New Fargo |
|---------------------------|--------------------|------------------|
| Dakotopolis | -5 | -10 |
| New Fargo | -10 | -3 |

| Origin\Destination | Dakotopolis | New Fargo |
|---------------------------|--------------------|------------------|
| Dakotopolis | 0.0067 | 0.0009 |
| New Fargo | 0.0009 | 0.0067 |

| Origin\Destination | Dakotopolis | New Fargo |
|---------------------------|--------------------|------------------|
| Dakotopolis | 0.0067 | 0.0000454 |
| New Fargo | 0.0000454 | 0.0565 |

Sum:

| Origin\Destination | Dakotopolis | New Fargo |
|---------------------------|--------------------|------------------|
| Dakotopolis | 0.0134 | 0.0009454 |
| New Fargo | 0.0009454 | 0.0498 |

P(Auto) =

| Origin\Destination | Dakotopolis | New Fargo |
|---------------------------|--------------------|------------------|
| Dakotopolis | 0.5 | 0.953 |
| New Fargo | 0.953 | 0.12 |

P(Transit) =

| Origin\Destination | Dakotopolis | New Fargo |
|---------------------------|--------------------|------------------|
| Dakotopolis | 0.5 | 0.047 |
| New Fargo | 0.047 | 0.88 |

Part B

B. Using the results from the previous problem (#2), how many car trips will there be?

Recall

Total Trips

| Origin\Destination | Dakotopolis | New Fargo |
|--------------------|-------------|-----------|
| Dakotopolis | 9395 | 5606 |
| New Fargo | 6385 | 15665 |

Total Trips by Auto =

| Origin\Destination | Dakotopolis | New Fargo |
|--------------------|-------------|-----------|
| Dakotopolis | 4697 | 5339 |
| New Fargo | 6511 | 1867 |

Fundamentals of Transportation/ Route Choice

Route assignment, route choice, or traffic assignment concerns the selection of routes (alternative called paths) between origins and destinations in transportation networks. It is the fourth step in the conventional transportation forecasting model, following Trip Generation, Destination Choice, and Mode Choice. The zonal interchange analysis of trip distribution provides origin-destination trip tables. Mode choice analysis tells which travelers will use which mode. To determine facility needs and costs and benefits, we need to know the number of travelers on each route and link of the network (a route is simply a chain of links between an origin and destination). We need to undertake traffic (or trip) assignment. Suppose there is a network of highways and transit systems and a proposed addition. We first want to know the present pattern of travel times and flows and then what would happen if the addition were made.

Link Performance Function

The cost that a driver imposes on others is called the marginal cost. However, when making decisions, a driver only faces his own cost (the average cost) and ignores any costs imposed on others (the marginal cost).

- $AverageCost = C_T/Q$
- $MarginalCost = \delta C_T / \delta Q$

Can Flow Exceed Capacity?

On a link, the capacity is thought of as “outflow.” Demand is inflow.

If inflow > outflow for a period of time, there is queueing (and delay).

For Example, for a 1 hour period, if 2100 cars arrive and 2000 depart, 100 are still there. The link performance function tries to represent that phenomenon in a simple way.

Wardrop's Principles of Equilibrium

User Equilibrium

Each user acts to minimize his/her own cost, subject to every other user doing the same. Travel times are equal on all used routes and lower than on any unused route.

System optimal

Each user acts to minimize the total travel time on the system.

Price of Anarchy

The reason we have congestion is that people are selfish. The cost of that selfishness (when people behave according to their own interest rather than society's) is the *price of anarchy*.

The ratio of system-wide travel time under User Equilibrium and System Optimal conditions.

$$\text{Price of Anarchy} = UE/SO > 1$$

For a two-link network with linear link performance functions (latency functions), Price of Anarchy is $< 4/3$.

Is this too much? Should something be done, or is 33% waste acceptable? [The loss may be larger/smaller in other cases, under different assumptions, etc.]

Conservation of Flow

An important factor in road assignment is the conservation of flow. This means that the number of vehicles entering the intersection (link segment) equals the number of vehicles exiting the intersection for a given period of time (except for sources and sinks).

Similarly, the number of vehicles entering the back of the link equals the number exiting the front (over a long period of time).

Auto assignment

Long-standing techniques

The above examples are adequate for a problem of two links, however real networks are much more complicated. The problem of estimating how many users are on each route is long standing. Planners started looking hard at it as freeways and expressways began to be developed. The freeway offered a superior level of service over the local street system and diverted traffic from the local system. At first, diversion was the technique. Ratios of travel time were used, tempered by considerations of costs, comfort, and level of service.

The Chicago Area Transportation Study (CATS) researchers developed diversion curves for freeways versus local streets. There was much work in California also, for California had early experiences with freeway planning. In addition to work of a diversion sort, the CATS attacked some technical problems that arise when one works with complex networks. One result was the Moore algorithm for finding shortest paths on networks.

The issue the diversion approach didn't handle was the feedback from the quantity of traffic on links and routes. If a lot of vehicles try to use a facility, the facility becomes congested and travel time increases. Absent some way to consider feedback, early planning studies (actually, most in the period 1960-1975) ignored feedback. They used the Moore algorithm to determine shortest paths and assigned all traffic to shortest paths. That's called all or nothing assignment because either all of the traffic from i to j moves along a route or it does not.

The all-or-nothing or shortest path assignment is not trivial from a technical-computational view. Each traffic zone is connected to $n - 1$ zones, so there are numerous paths to be considered. In addition, we are ultimately interested in traffic on links. A link may be a part of several paths, and traffic along paths has to be summed link by link.

An argument can be made favoring the all-or-nothing approach. It goes this way: The planning study is to support investments so that a good level of service is available on all links. Using the travel times associated with the planned level of service, calculations indicate how traffic will flow once improvements are in place. Knowing the quantities of traffic on links, the capacity to be supplied to meet the desired level of service can be calculated.

Heuristic procedures

To take account of the affect of traffic loading on travel times and traffic equilibria, several heuristic calculation procedures were developed. One heuristic proceeds incrementally. The traffic to be assigned is divided into parts (usually 4). Assign the first part of the traffic. Compute new travel times and assign the next part of the traffic. The last step is repeated until all the traffic is assigned. The CATS used a variation on this; it assigned row by row in the O-D table.

The heuristic included in the FHWA collection of computer programs proceeds another way.

- Step 0: Start by loading all traffic using an all or nothing procedure.
- Step 1: Compute the resulting travel times and reassign traffic.
- Step 2: Now, begin to reassigned using weights. Compute the weighted travel times in the previous two loadings and use those for the next assignment. The latest iteration gets a weight of 0.25 and the previous gets a weight of 0.75.
- Step 3. Continue.

These procedures seem to work "pretty well," but they are not exact.

Frank-Wolfe algorithm

Dafermos (1968) applied the Frank-Wolfe algorithm ^[1] (1956, Florian 1976), which can be used to deal with the traffic equilibrium problem. Suppose we are considering a highway network. For each link there is a function stating the relationship between resistance and volume of traffic. The Bureau of Public Roads (BPR) developed a link (arc) congestion (or volume-delay, or link performance) function, which we will term $S_a(v_a)$

$$S_a(v_a) = t_a \left(1 + 0.15 \left(\frac{v_a}{c_a} \right)^4 \right)$$

t_a = free flow travel time on link a per unit of time

v_a = volume of traffic on link a per unit of time (somewhat more accurately: flow attempting to use link a)

c_a = capacity of link a per unit of time

$S_a(v_a)$ is the average travel time for a vehicle on link a . There are other congestion functions. The CATS has long used a function different from that used by the BPR, but there seems to be little difference between results when the CATS and BPR functions are compared.

Equilibrium Assignment

To assign traffic to paths and links we have to have rules, and there are the well-known Wardrop equilibrium ^[2] (1952) conditions. The essence of these is that travelers will strive to find the shortest (least resistance) path from origin to destination, and network equilibrium occurs when no traveler can decrease travel effort by shifting to a new path. These are termed user optimal conditions, for no user will gain from changing travel paths once the system is in equilibrium.

The user optimum equilibrium can be found by solving the following nonlinear programming problem

$$\min \sum_a \int_0^{v_a} S_a(x) dx$$

subject to:

$$v_a = \sum_i \sum_j \sum_r \alpha_{ij}^{ar} x_{ij}^r$$

$$\sum_r x_{ij}^r = T_{ij}$$

$$v_a \geq 0, x_{ij}^r \geq 0$$

where x_{ij}^r is the number of vehicles on path r from origin i to destination j . So constraint (2) says that all travel must take place - $i = 1 \dots n; j = 1 \dots n$

$\alpha_{ij}^{ar} = 1$ if link a is on path r from i to j ; zero otherwise.

So constraint (1) sums traffic on each link. There is a constraint for each link on the network. Constraint (3) assures no negative traffic.

Transit assignment

There are also methods that have been developed to assign passengers to transit vehicles. Some of these are discussed in Spiess (1993) Transit Equilibrium Assignment based on Optimal Strategies [3]

Integrating travel choices

The urban transportation planning model evolved as a set of steps to be followed and models evolved for use in each step. Sometimes there were steps within steps, as was the case for the first statement of the Lowry model. In some cases, it has been noted that steps can be integrated. More generally, the steps abstract from decisions that may be made simultaneously and it would be desirable to better replicate that in the analysis.

Disaggregate demand models were first developed to treat the mode choice problem. That problem assumes that one has decided to take a trip, where that trip will go, and at what time the trip will be made. They have been used to treat the implied broader context. Typically, a nested model will be developed, say, starting with the probability of a trip being made, then examining the choice among places, and then mode choice. The time of travel is a bit harder to treat.

Wilson's doubly constrained entropy model has been the point of departure for efforts at the aggregate level. That model contains the constraint

$$t_{ij}c_{ij} = C$$

where the c_{ij} are the link travel costs, t_{ij} refers to traffic on a link, and C is a resource constraint to be sized when fitting the model with data. Instead of using that form of the constraint, the monotonically increasing resistance function used in traffic assignment can be used. The result determines zone-to-zone movements and assigns traffic to networks, and that makes much sense from the way one would imagine the system works. Zone-to-zone traffic depends on the resistance occasioned by congestion.

Alternatively, the link resistance function may be included in the objective function (and the total cost function eliminated from the constraints).

A generalized disaggregate choice approach has evolved as has a generalized aggregate approach. The large question is that of the relations between them. When we use a macro model, we would like to know the disaggregate behavior it represents. If we are doing a micro analysis, we would like to know the aggregate implications of the analysis.

Wilson derives a gravity-like model with weighted parameters that say something about the attractiveness of origins and destinations. Without too much math we can write probability of choice statements based on attractiveness, and these take a form similar to some varieties of disaggregate demand models.

Integrating travel demand with route assignment

It has long been recognized that travel demand is influenced by network supply. The example of a new bridge opening where none was before inducing additional traffic has been noted for centuries. Much research has gone into developing methods for allowing the forecasting system to directly account for this phenomenon. Evans (1974) published a doctoral dissertation on a mathematically rigorous combination of the gravity distribution model with the equilibrium assignment model. The earliest citation of this integration is the

work of Irwin and Von Cube, as related by Florian et al. (1975), who comment on the work of Evans:

"The work of Evans resembles somewhat the algorithms developed by Irwin and Von Cube ("Capacity Restraint in Multi-Travel Mode Assignment Programs" H.R.B. Bulletin 347 (1962)) for a transportation study of Toronto, Canada. Their work allows for feedback between congested assignment and trip distribution, although they apply sequential procedures. Starting from an initial solution of the distribution problem, the interzonal trips are assigned to the initial shortest routes. For successive iterations, new shortest routes are computed, and their lengths are used as access times for input the distribution model. The new interzonal flows are then assigned in some proportion to the routes already found. The procedure is stopped when the interzonal times for successive iteration are quasi-equal."

Florian et al. proposed a somewhat different method for solving the combined distribution assignment, applying directly the Frank-Wolfe algorithm. Boyce et al. (1988) summarize the research on Network Equilibrium Problems, including the assignment with elastic demand.

Discussion

A three link problem can not be solved graphically, and most transportation network problems involve a large numbers of nodes and links. Eash et al, for instance, studied the road net on DuPage County where there were about 30,000 one-way links and 9,500 nodes. Because problems are large, an algorithm is needed to solve the assignment problem, and the Frank-Wolfe algorithm (modified a bit since first published) is used. Start with an all or nothing assignment and then follow the rule developed by Frank-Wolfe to iterate toward the minimum value of the objective function. The algorithm applies successive feasible solutions to achieve convergence to the optimal solution. It uses an efficient search procedure to move the calculation rapidly toward the optimal solution. Travel times correspond to the dual variables in this programming problem.

It is interesting that the Frank-Wolfe algorithm was available in 1956. Its application was developed in 1968 and it took almost another two decades before the first equilibrium assignment algorithm was embedded in commonly used transportation planning software (Emme and Emme/2, developed by Florian and others in Montreal). We would not want to draw any general conclusion from the slow application observation, mainly because we can find counter examples about the pace and pattern of technique development. For example, the simplex method for the solution of linear programming problems was worked out and widely applied prior to the development of much of programming theory.

The problem statement and algorithm have general applications across civil engineering -- hydraulics, structures, and construction. (See Hendrickson and Janson 1984).

Examples

Example 1



Problem:

Solve for the flows on Links 1 and 2 in the Simple Network of two parallel links just shown if the link performance function on link 1:

$$C_1 = 5 + 2 * Q_1$$

and the function on link 2:

$$C_2 = 10 + Q_2$$

where total flow between the origin and destination is 1000 trips.



Solution:

Time (Cost) is equal on all used routes so $C_1 = C_2$

And we have Conservation of flow so, $Q_1 + Q_2 = Q_o = Q_d = 1000$

$$5 + 2 * (1000 - Q_2) = 10 + Q_2$$

$$1995 = 3Q_2$$

$$Q_2 = 665; Q_1 = 335$$

Example 2



Problem:

An example from Eash, Janson, and Boyce (1979) will illustrate the solution to the nonlinear program problem. There are two links from node 1 to node 2, and there is a resistance function for each link (see Figure 1). Areas under the curves in Figure 2 correspond to the integration from 0 to a in equation 1, they sum to 220,674. Note that the function for link b is plotted in the reverse direction.

$$S_a = 15 \left(1 + 0.15 \left(\frac{v_a}{1000} \right)^4 \right)$$

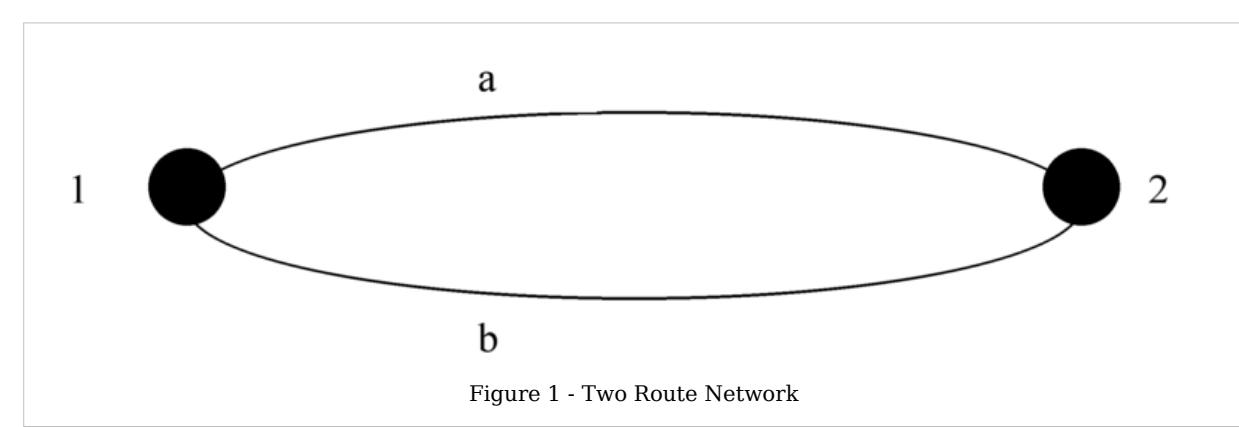
$$S_b = 20 \left(1 + 0.15 \left(\frac{v_b}{3000} \right)^4 \right)$$

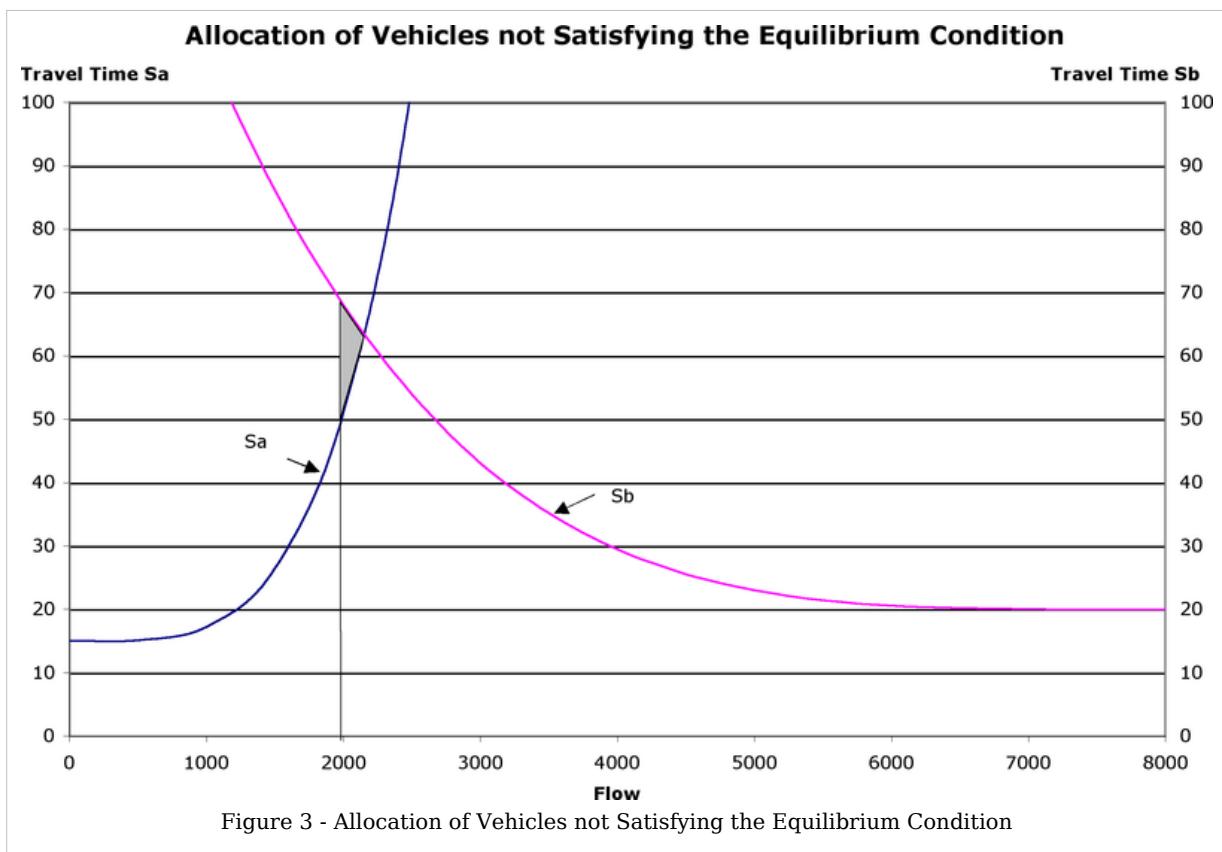
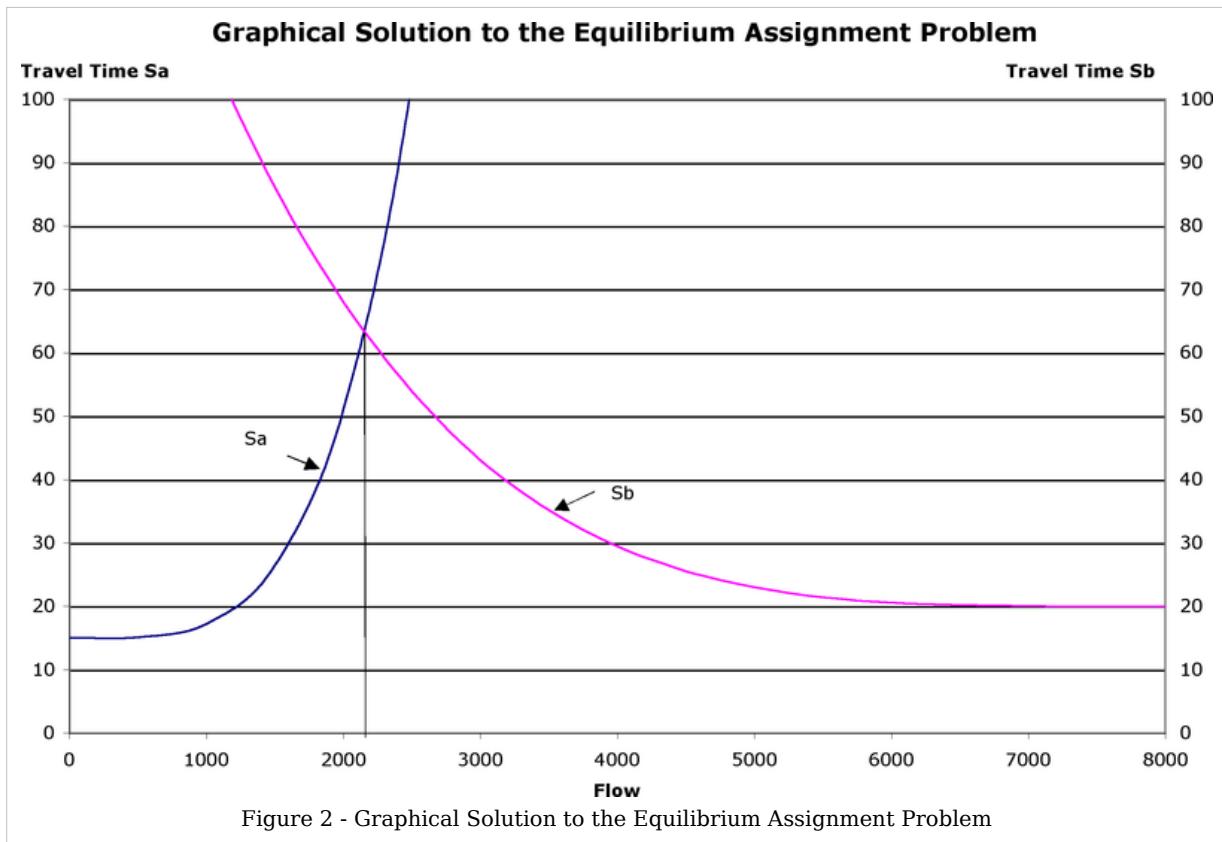
$$v_a + v_b = 8000$$

Show graphically the equilibrium result.



Solution:





At equilibrium there are 2,152 vehicles on link a and 5,847 on link b . Travel time is the same on each route: about 63.

Figure 3 illustrates an allocation of vehicles that is not consistent with the equilibrium solution. The curves are unchanged, but with the new allocation of vehicles to routes the shaded area has to be included in the solution, so the Figure 3 solution is larger than the solution in Figure 2 by the area of the shaded area.

Thought Questions

- How can we get drivers to consider their marginal cost?
- Alternatively: How can we get drivers to behave in a “System Optimal” way?

Sample Problem

- Problem (Solution)

Variables

- C_T - total cost
- C_k - travel cost on link k
- Q_k - flow (volume) on link k

Abbreviations

- VDF - Volume Delay Function
- LPF - Link Performance Function
- BPR - Bureau of Public Roads
- UE - User Equilibrium
- SO - System Optimal
- DTA - Dynamic Traffic Assignment
- DUE - Deterministic User Equilibrium
- SUE - Stochastic User Equilibrium
- AC - Average Cost
- MC - Marginal Cost

Key Terms

- Route assignment, route choice, auto assignment
- Volume-delay function, link performance function
- User equilibrium
- System optimal
- Conservation of flow
- Average cost
- Marginal cost

External Exercises

Use the ADAM software at the STREET website ^[4] and try Assignment #3 to learn how changes in network characteristics impact route choice.

References

- Dafermos, Stella. C. and F.T. Sparrow The Traffic Assignment Problem for a General Network." J. of Res. of the National Bureau of Standards, 73B, pp. 91-118. 1969.
- Florian, Michael ed., Traffic Equilibrium Methods, Springer-Verlag, 1976.
- Wardsop, J. C. Some Theoretical Aspects of Road Traffic Research," Proceedings, Institution of Civil Engineers Part 2, 9, pp. 325-378. 1952
- Eash, Ronald, Bruce N. Janson, and David Boyce Equilibrium Trip Assignment: Advantages and Implications for Practice, Transportation Research Record 728, pp. 1-8, 1979.
- Evans, Suzanne P. . "Derivation and Analysis of Some Models for Combining Trip Distribution and Assignment." Transportation Research, Vol 10, pp 37-57 1976
- Hendrickson, C.T. and B.N. Janson, "A Common Network Flow Formulation to Several Civil Engineering Problems" Civil Engineering Systems 1(4), pp. 195-203, 1984

References

- [1] http://en.wikipedia.org/wiki/Frank-Wolfe_algorithm
- [2] <http://en.wikipedia.org/wiki/Wardrop>
- [3] <http://emme2.spiess.ch/congtras/congtras.html>
- [4] <http://street.umn.edu/>

Fundamentals of Transportation/ Route Choice/Problem


Problem:

Given a flow of six (6) units from origin "o" to destination "r". Flow on each route ab is designated with Q_{ab} in the Time Function. Apply Wardrop's Network Equilibrium Principle (Users Equalize Travel Times on all used routes)

- A.** What is the flow and travel time on each link? (complete the table below) for Network A

Link Attributes

| Link | Link Performance Function | Flow | Time |
|------|----------------------------|------|------|
| o-p | $C_{op} = 5 * Q_{op}$ | | |
| p-r | $C_{pr} = 25 + Q_{pr}$ | | |
| o-q | $C_{oq} = 20 + 2 * Q_{oq}$ | | |
| q-r | $C_{qr} = 5 * Q_{qr}$ | | |

- B.** What is the system optimal assignment?

- C.** What is the Price of Anarchy?

- Solution

Fundamentals of Transportation/ Route Choice/Solution


Problem:

Given a flow of six (6) units from origin "o" to destination "r". Flow on each route ab is designated with Q_{ab} in the Time Function. Apply Wardrop's Network Equilibrium Principle (Users Equalize Travel Times on all used routes)

- A.** What is the flow and travel time on each link? (complete the table below) for Network A

Link Attributes

| Link | Link Performance Function | Flow | Time |
|------|----------------------------|------|------|
| o-p | $C_{op} = 5 * Q_{op}$ | | |
| p-r | $C_{pr} = 25 + Q_{pr}$ | | |
| o-q | $C_{oq} = 20 + 2 * Q_{oq}$ | | |
| q-r | $C_{qr} = 5 * Q_{qr}$ | | |

- B.** What is the system optimal assignment?

- C.** What is the Price of Anarchy?

**Solution:****Part A**

What is the flow and travel time on each link? Complete the table below for Network A:

Link Attributes

| Link | Link Performance Function | Flow | Time |
|------|----------------------------|------|------|
| o-p | $C_{op} = 5 * Q_{op}$ | | |
| p-r | $C_{pr} = 25 + Q_{pr}$ | | |
| o-q | $C_{oq} = 20 + 2 * Q_{oq}$ | | |
| q-r | $C_{qr} = 5 * Q_{qr}$ | | |

These four links are really 2 links O-P-R and O-Q-R, because by conservation of flow $Q_{op} = Q_{pr}$ and $Q_{oq} = Q_{qr}$.

Link Attributes

| Link | Link Performance Function | Flow | Time |
|-------|------------------------------|------|------|
| o-p-r | $C_{opr} = 25 + 5 * Q_{opr}$ | | |
| o-q-r | $C_{oqr} = 20 + 7 * Q_{oqr}$ | | |

By Wardrop's Equilibrium Principle, The travel time (cost) on each used route must be equal. So $C_{opr} = C_{oqr}$. OR $25 + 6 * Q_{opr} = 20 + 7 * Q_{oqr}$ $5 + 6 * Q_{opr} = 7 * Q_{oqr}$
 $Q_{oqr} = 5/7 + 6/7 Q_{opr}$

By the conservation of flow principle $Q_{oqr} + Q_{opr} = 6$ $Q_{opr} = 6 - Q_{oqr}$ By substitution
 $Q_{oqr} = 5/7 + 6/7(6 - Q_{oqr}) = 41/7 - 6/7 Q_{oqr}$ $13/7 Q_{oqr} = 41/7$ $Q_{oqr} = 41/13 = 3.15$
 $Q_{opr} = 2.84$ Check $42.01 = 25 + 6(2.84)$? $20 + 7(3.15) = 42.05$ Check (within rounding error)

Link Attributes

| Link | Link Performance Function | Flow | Time |
|-------|------------------------------|------|-------|
| o-p-r | $C_{opr} = 25 + 5 * Q_{opr}$ | 2.84 | 42.01 |
| o-q-r | $C_{oqr} = 20 + 7 * Q_{oqr}$ | 3.15 | 42.01 |

or expanding back to the original table:

Link Attributes

| Link | Link Performance Function | Flow | Time |
|------|----------------------------|------|-------|
| o-p | $C_{op} = 5 * Q_{op}$ | 2.84 | 14.2 |
| p-r | $C_{pr} = 25 + Q_{pr}$ | 2.84 | 27.84 |
| o-q | $C_{oq} = 20 + 2 * Q_{oq}$ | 3.15 | 26.3 |
| q-r | $C_{qr} = 5 * Q_{qr}$ | 3.15 | 15.75 |

User Equilibrium: Total Delay = $42.01 * 6 = 252.06$

Part B

What is the system optimal assignment?

What is the system optimal assignment for the previous example

Conservation of Flow:

$$Q_{oqr} + Q_{opr} = 6$$

$$Total Delay = Q_{opr}(25 + 6 * Q_{opr}) + Q_{oqr}(20 + 7 * Q_{oqr})$$

$$25Q_{opr} + 6Q_{opr}^2 + (6 - Q_{opr})(20 + 7(6 - Q_{opr}))$$

$$25Q_{opr} + 6Q_{opr}^2 + (6 - Q_{opr})(62 - 7Q_{opr}))$$

$$25Q_{opr} + 6Q_{opr}^2 + 372 - 62Q_{opr} - 42Q_{opr} + 7Q_{opr}^2$$

$$13Q_{opr}^2 - 79Q_{opr} + 372$$

Analytic Solution requires minimizing total delay

$$\delta C / \delta Q = 26Q_{opr} - 79 = 0$$

$$Q_{opr} = 79/26 = 3.04$$

$$Q_{oqr} = 6 - Q_{opr} = 2.96$$

And we can compute the SO travel times on each path

$$C_{opr,SO} = 25 + 6 * 3.04 = 43.24$$

$$C_{oqr,SO} = 20 + 7 * 2.96 = 40.72$$

Note that unlike the UE solution, $C_{opr,SO} > C_{oqr,SO}$

$$\text{Total Delay} = 3.04(25 + 6 * 3.04) + 2.96(20 + 7 * 2.96) = 131.45 + 120.53 = 251.98$$

Note: one could also use software such as a "Solver" algorithm to find this solution.

Part C

What is the Price of Anarchy?

User Equilibrium: Total Delay = 252.06 System Optimal: Total Delay = 251.98

$$\text{Price of Anarchy} = 252.06 / 251.98 = 1.0003 < 4/3$$

Fundamentals of Transportation/ Evaluation

A benefit-cost analysis (BCA)^[1] is often required in determining whether a project should be approved and is useful for comparing similar projects. It determines the stream of quantifiable economic benefits and costs that are associated with a project or policy. If the benefits exceed the costs, the project is worth doing; if the benefits fall short of the costs, the project is not. Benefit-cost analysis is appropriate where the technology is known and well understood or a minor change from existing technologies is being performed. BCA is not appropriate when the technology is new and untried because the effects of the technology cannot be easily measured or predicted. However, just because something is new in one place does not necessarily make it new, so benefit-cost analysis would be appropriate, e.g., for a light-rail or commuter rail line in a city without rail, or for any road project, but would not be appropriate (at the time of this writing) for something truly radical like teleportation.

The identification of the costs, and more particularly the benefits, is the chief component of the “art” of Benefit-Cost Analysis. This component of the analysis is different for every project. Furthermore, care should be taken to avoid double counting; especially counting cost savings in both the cost and the benefit columns. However, a number of benefits and costs should be included at a minimum. In transportation these costs should be separated for users, transportation agencies, and the public at large. Consumer benefits are measured by consumer’s surplus. It is important to recognize that the demand curve is downward sloping, so there a project may produce benefits both to existing users in terms of a reduction in cost and to new users by making travel worthwhile where previously it was too expensive.

Agency benefits come from profits. But since most agencies are non-profit, they receive no direct profits. Agency construction, operating, maintenance, or demolition costs may be reduced (or increased) by a new project; these cost savings (or increases) can either be considered in the cost column, or the benefit column, but not both.

Society is impacted by transportation project by an increase or reduction of negative and positive externalities. Negative externalities, or social costs, include air and noise pollution and accidents. Accidents can be considered either a social cost or a private cost, or divided into two parts, but cannot be considered in total in both columns.

If there are network externalities (i.e. the benefits to consumers are themselves a function of the level of demand), then consumers’ surplus for each different demand level should be computed. Of course this is easier said than done. In practice, positive network externalities are ignored in Benefit Cost Analysis.

Background

Early Beginnings

When Benjamin Franklin was confronted with difficult decisions, he often recorded the pros and cons on two separate columns and attempted to assign weights to them. While not mathematically precise, this “moral or prudential algebra”, as he put it, allowed for careful consideration of each “cost” and “benefit” as well as the determination of a course of action that provided the greatest benefit. While Franklin was certainly a proponent of this technique, he was certainly not the first. Western European governments, in particular, had been employing similar methods for the construction of waterway and shipyard improvements.

Ekelund and Hebert (1999) credit the French as pioneers in the development of benefit-cost analyses for government projects. The first formal benefit-cost analysis in France occurred in 1708. Abbe de Saint-Pierre attempted to measure and compare the incremental benefit of road improvements (utility gained through reduced transport costs and increased trade), with the additional construction and maintenance costs. Over the next century, French economists and engineers applied their analysis efforts to canals (Ekelund and Hebert, 1999). During this time, The Ecole Polytechnique had established itself as France’s premier educational institution, and in 1837 sought to create a new course in “social arithmetic”: “...the execution of public works will in many cases tend to be handled by a system of concessions and private enterprise. Therefore our engineers must henceforth be able to evaluate the utility or inconvenience, whether local or general, or each enterprise; consequently they must have true and precise knowledge of the elements of such investments.” (Ekelund and Hebert, 1999, p. 47). The school also wanted to ensure their students were aware of the effects of currencies, loans, insurance, amortization and how they affected the probable benefits and costs to enterprises.

In the 1840s French engineer and economist Jules Dupuit (1844, tr. 1952) published an article “On Measurement of the Utility of Public Works”, where he posited that benefits to society from public projects were not the revenues taken in by the government (Aruna, 1980). Rather the benefits were the difference between the public’s willingness to pay and the actual payments the public made (which he theorized would be smaller). This “relative utility” concept was what Alfred Marshall would later rename with the more familiar term, “consumer surplus” (Ekelund and Hebert, 1999).

Vilfredo Pareto (1906) developed what became known as Pareto improvement and Pareto efficiency (optimal) criteria. Simply put, a policy is a Pareto improvement if it provides a benefit to at least one person without making anyone else worse off (Boardman, 1996). A policy is Pareto efficient (optimal) if no one else can be made better off without making someone else worse off. British economists Kaldor and Hicks (Hicks, 1941; Kaldor, 1939) expanded on this idea, stating that a project should proceed if the losers could be compensated in some way. It is important to note that the Kaldor-Hicks criteria states it is sufficient if the winners could potentially compensate the project losers. It does not require that they be compensated.

Benefit-cost Analysis in the United States

Much of the early development of benefit-cost analysis in the United States is rooted in water related infrastructure projects. The US Flood Control Act of 1936 was the first instance of a systematic effort to incorporate benefit-cost analysis to public decision-making. The act stated that the federal government should engage in flood control activities if “the benefits to whomsoever they may accrue [be] in excess of the estimated costs,” but did not provide guidance on how to define benefits and costs (Aruna, 1980, Persky, 2001). Early Tennessee Valley Authority (TVA) projects also employed basic forms of benefit-cost analysis (US Army Corp of Engineers, 1999). Due to the lack of clarity in measuring benefits and costs, many of the various public agencies developed a wide variety of criteria. Not long after, attempts were made to set uniform standards.

The U.S. Army Corp of Engineers “Green Book” was created in 1950 to align practice with theory. Government economists used the Kaldor-Hicks criteria as their theoretical foundation for the restructuring of economic analysis. This report was amended and expanded in 1958 under the title of “The Proposed Practices for Economic Analysis of River Basin Projects” (Persky, 2001).

The Bureau of the Budget adopted similar criteria with 1952’s Circular A-47 - “Reports and Budget Estimates Relating to Federal Programs and Projects for Conservation, Development, or Use of Water and Related Land Resources”.

Modern Benefit-cost Analysis

During the 1960s and 1970s the more modern forms of benefit-cost analysis were developed. Most analyses required evaluation of:

1. The present value of the benefits and costs of the proposed project at the time they occurred
2. The present value of the benefits and costs of alternatives occurring at various points in time (opportunity costs)
3. Determination of risky outcomes (sensitivity analysis)
4. The value of benefits and costs to people with different incomes (distribution effects/equity issues) (Layard and Glaister, 1994)

The Planning Programming Budgeting System (PBBS) - 1965

The Planning Programming Budgeting System (PBBS) developed by the Johnson administration in 1965 was created as a means of identifying and sorting priorities. This grew out of a system Robert McNamara created for the Department of Defense a few years earlier (Gramlich, 1981). The PBBS featured five main elements:

1. A careful specification of basic program objectives in each major area of governmental activity.
2. An attempt to analyze the outputs of each governmental program.
3. An attempt to measure the costs of the program, not for one year but over the next several years (“several” was not explicitly defined)
4. An attempt to compare alternative activities.
5. An attempt to establish common analytic techniques throughout the government.

Office of Management and Budget (OMB) - 1977

Throughout the next few decades, the federal government continued to demand improved benefit-cost analysis with the aim of encouraging transparency and accountability. Approximately 12 years after the adoption of the PPBS system, the Bureau of the Budget was renamed the Office of Management and Budget (OMB). The OMB formally adopted a system that attempts to incorporate benefit-cost logic into budgetary decisions. This came from the Zero-Based Budgeting system set up by Jimmy Carter when he was governor of Georgia (Gramlich, 1981).

Recent Developments

Executive Order 12292, issued by President Reagan in 1981, required a regulatory impact analysis (RIA) for every major governmental regulatory initiative over \$100 million. The RIA is basically a benefit-cost analysis that identifies how various groups are affected by the policy and attempts to address issues of equity (Boardman, 1996).

According to Robert Dorfman, (Dorfman, 1997) most modern-day benefit-cost analyses suffer from several deficiencies. The first is their attempt "to measure the social value of all the consequences of a governmental policy or undertaking by a sum of dollars and cents". Specifically, Dorfman mentions the inherent difficulty in assigning monetary values to human life, the worth of endangered species, clean air, and noise pollution. The second shortcoming is that many benefit-cost analyses exclude information most useful to decision makers: the distribution of benefits and costs among various segments of the population. Government officials need this sort of information and are often forced to rely on other sources that provide it, namely, self-seeking interest groups. Finally, benefit-cost reports are often written as though the estimates are precise, and the readers are not informed of the range and/or likelihood of error present.

The Clinton Administration sought proposals to address this problem in revising Federal benefit-cost analyses. The proposal required numerical estimates of benefits and costs to be made in the most appropriate unit of measurement, and "specify the ranges of predictions and shall explain the margins of error involved in the quantification methods and in the estimates used" (Dorfman, 1997). Executive Order 12898 formally established the concept of Environmental Justice with regards to the development of new laws and policies, stating they must consider the "fair treatment for people of all races, cultures, and incomes." The order requires each federal agency to identify and address "disproportionately high and adverse human health or environmental effects of its programs, policies and activities on minority and low-income populations."

Probabilistic Benefit-Cost Analysis

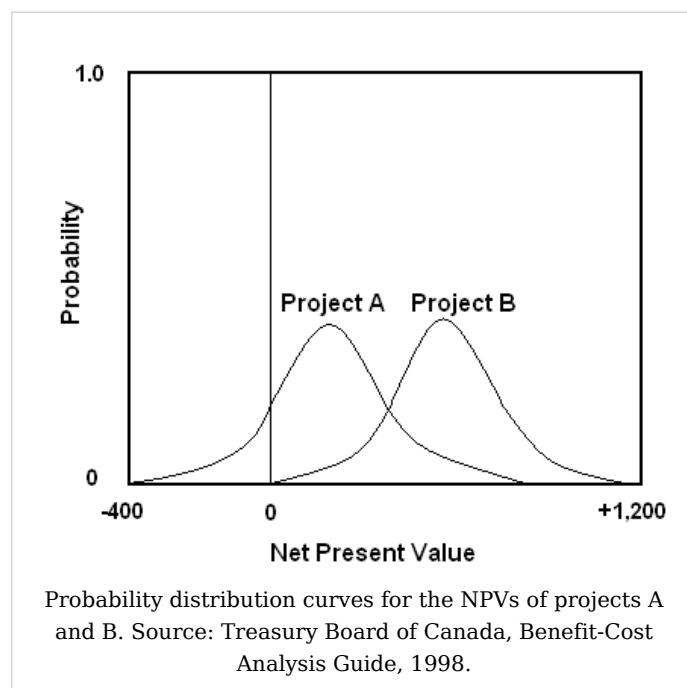
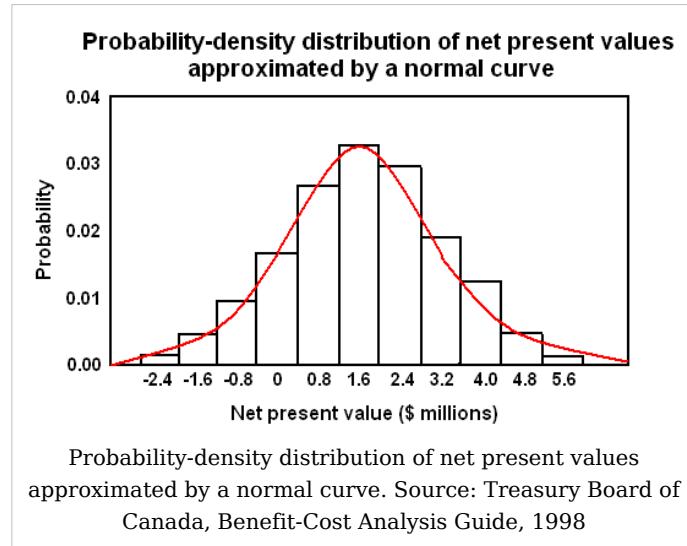
In recent years there has been a push for the integration of sensitivity analyses of possible outcomes of public investment projects with open discussions of the merits of assumptions used. This "risk analysis" process has been suggested by Flyvbjerg (2003) in the spirit of encouraging more transparency and public involvement in decision-making.

The Treasury Board of Canada's Benefit-Cost Analysis Guide recognizes that implementation of a project has a probable range of benefits and costs. It posits that the "effective sensitivity" of an outcome to a particular variable is determined by four factors:

- the responsiveness of the Net Present Value (NPV) to changes in the variable;
- the magnitude of the variable's range of plausible values;
- the volatility of the value of the variable (that is, the probability that the value of the variable will move within that range of plausible values); and
- the degree to which the range or volatility of the values of the variable can be controlled.

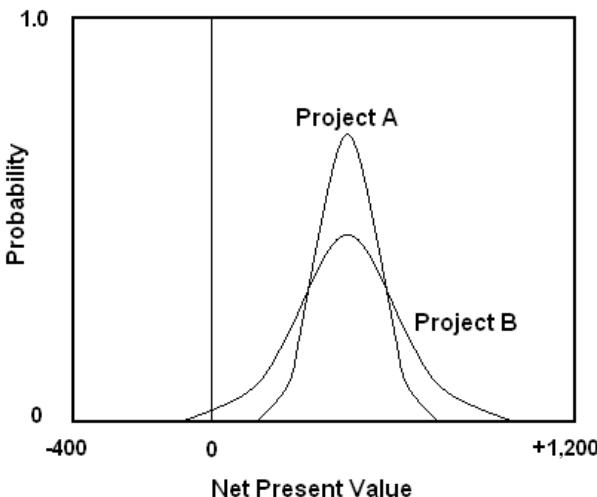
It is helpful to think of the range of probable outcomes in a graphical sense, as depicted in Figure 1 (probability versus NPV).

Once these probability curves are generated, a comparison of different alternatives can also be performed by plotting each one on the same set of ordinates. Consider for example, a comparison between alternative A and B (Figure 2).



In Figure 2, the probability that any specified positive outcome will be exceeded is always higher for project B than it is for project A. The decision maker should, therefore, always prefer project B over project A. In other cases, an alternative may have a much broader or narrower range of NPVs compared to other alternatives (Figure 3).

Some decision-makers might be attracted by the possibility of a higher return (despite the possibility of greater loss) and therefore might choose project B. Risk-averse decision-makers will be attracted by the possibility of lower loss and will therefore be inclined to choose project A.



Probability distribution curves for the NPVs of projects A and B, where Project A has a narrower range of possible NPVs. Source: Treasury Board of Canada, Benefit-Cost Analysis Guide, 1998

Discount rate

Both the costs and benefits flowing from an investment are spread over time. While some costs are one-time and borne up front, other benefits or operating costs may be paid at some point in the future, and still others received as a stream of payments collected over a long period of time. Because of inflation, risk, and uncertainty, a dollar received now is worth more than a dollar received at some time in the future. Similarly, a dollar spent today is more onerous than a dollar spent tomorrow. This reflects the concept of time preference that we observe when people pay bills later rather than sooner. The existence of real interest rates reflects this time preference. The appropriate discount rate depends on what other opportunities are available for the capital. If simply putting the money in a government insured bank account earned 10% per year, then at a minimum, no investment earning less than 10% would be worthwhile. In general, projects are undertaken with those with the highest rate of return first, and then so on until the cost of raising capital exceeds the benefit from using that capital. Applying this efficiency argument, no project should be undertaken on cost-benefit grounds if another feasible project is sitting there with a higher rate of return.

Three alternative bases for the setting the government's test discount rate have been proposed:

1. The social rate of time preference recognizes that a dollar's consumption today will be more valued than a dollar's consumption at some future time for, in the latter case, the dollar will be subtracted from a higher income level. The amount of this difference per dollar over a year gives the annual rate. By this method, a project should not be undertaken unless its rate of return exceeds the social rate of time preference.
2. The opportunity cost of capital basis uses the rate of return of private sector investment, a government project should not be undertaken if it earns less than a private sector investment. This is generally higher than social time preference.

3. The cost of funds basis uses the cost of government borrowing, which for various reasons related to government insurance and its ability to print money to back bonds, may not equal exactly the opportunity cost of capital.

Typical estimates of social time preference rates are around 2 to 4 percent while estimates of the social opportunity costs are around 7 to 10 percent.

Generally, for Benefit-Cost studies an acceptable rate of return (the government's test rate) will already have been established. An alternative is to compute the analysis over a range of interest rates, to see to what extent the analysis is sensitive to variations in this factor. In the absence of knowing what this rate is, we can compute the rate of return (internal rate of return) for which the project breaks even, where the net present value is zero. Projects with high internal rates of return are preferred to those with low rates.

Determine a present value

The basic math underlying the idea of determining a present value is explained using a simple compound interest rate problem as the starting point. Suppose the sum of \$100 is invested at 7 percent for 2 years. At the end of the first year the initial \$100 will have earned \$7 interest and the augmented sum (\$107) will earn a further 7 percent (or \$7.49) in the second year. Thus at the end of 2 years the \$100 invested now will be worth \$114.49.

The discounting problem is simply the converse of this compound interest problem. Thus, \$114.49 receivable in 2 years time, and discounted by 7 per cent, has a present value of \$100.

Present values can be calculated by the following equation:

$$(1) P = \frac{F}{(1+i)^n}$$

where:

- F = future money sum
- P = present value
- i = discount rate per time period (i.e. years) in decimal form (e.g. 0.07)
- n = number of time periods before the sum is received (or cost paid, e.g. 2 years)

Illustrating our example with equations we have:

$$P = \frac{F}{(1+i)^n} = \frac{114.49}{(1+0.07)^2} = 100.00$$

The present value, in year 0, of a stream of equal annual payments of \$A starting year 1, is given by the reciprocal of the equivalent annual cost. That is, by:

$$(2) P = A \left[\frac{(1+i)^n - 1}{i(1+i)^n} \right]$$

where:

- A = Annual Payment

For example: 12 annual payments of \$500, starting in year 1, have a present value at the middle of year 0 when discounted at 7% of: \$3971

$$P = A \left[\frac{(1+i)^n - 1}{i(1+i)^n} \right] = 500 \left[\frac{(1+0.07)^{12} - 1}{0.07(1+0.07)^{12}} \right] = 3971$$

The present value, in year 0, of m annual payments of \$A, starting in year $n + 1$, can be calculated by combining discount factors for a payment in year n and the factor for the present value of m annual payments. For example: 12 annual mid-year payments of \$250 in years 5 to 16 have a present value in year 4 of \$1986 when discounted at 7%. Therefore in year 0, 4 years earlier, they have a present value of \$1515.

$$P_{Y=4} = A \left[\frac{(1+i)^n - 1}{i(1+i)^n} \right] = 250 \left[\frac{(1+0.07)^{12} - 1}{0.07(1+0.07)^{12}} \right] = 1986$$

$$P_{Y=0} = \frac{F}{(1+i)^n} = \frac{P_{Y=4}}{(1+i)^n} = \frac{1986}{(1+0.07)^4} = 1515$$

Evaluation criterion

Three equivalent conditions can tell us if a project is worthwhile

1. The discounted present value of the benefits exceeds the discounted present value of the costs
2. The present value of the net benefit must be positive.
3. The ratio of the present value of the benefits to the present value of the costs must be greater than one.

However, that is not the entire story. More than one project may have a positive net benefit. From the set of mutually exclusive projects, the one selected should have the highest net present value. We might note that if there are insufficient funds to carry out all mutually exclusive projects with a positive net present value, then the discount used in computing present values does not reflect the true cost of capital. Rather it is too low.

There are problems with using the internal rate of return or the benefit/cost ratio methods for project selection, though they provide useful information. The ratio of benefits to costs depends on how particular items (for instance, cost savings) are ascribed to either the benefit or cost column. While this does not affect net present value, it will change the ratio of benefits to costs (though it cannot move a project from a ratio of greater than one to less than one).

Examples

Example 1: Benefit Cost Application



Problem:

This problem, adapted from Watkins (1996)^[2], illustrates how a Benefit Cost Analysis might be applied to a project such as a highway widening. The improvement of the highway saves travel time and increases safety (by bringing the road to modern standards). But there will almost certainly be more total traffic than was carried by the old highway. This example excludes external costs and benefits, though their addition is a straightforward extension. The data for the "No Expansion" can be collected from off-the-shelf sources. However the "Expansion" column's data requires the use of forecasting and modeling.

Table 1: Data

| | No Expansion | Expansion |
|--------------------------------|--------------|-----------|
| Travel Time Savings (\$/yr) | \$100,000 | \$150,000 |
| Safety Improvements (\$/yr) | \$50,000 | \$75,000 |
| Annual Maintenance (\$/yr) | \$20,000 | \$20,000 |
| Land Acquisition (\$/yr) | \$0 | \$50,000 |
| Construction Costs (\$/yr) | \$100,000 | \$150,000 |
| Annual Operation Costs (\$/yr) | \$50,000 | \$50,000 |
| Total Annual Benefits (\$/yr) | \$300,000 | \$450,000 |
| Total Annual Costs (\$/yr) | \$370,000 | \$470,000 |
| Net Annual Benefit (\$/yr) | -\$70,000 | -\$20,000 |

| | | | |
|--------------------|------------|--------|--------|
| Peak | | | |
| Passenger Trips | (per hour) | 18,000 | 24,000 |
| Trip Time | (minutes) | 50 | 30 |
| Off-peak | | | |
| Passenger Trips | (per hour) | 9,000 | 10,000 |
| Trip Time | (minutes) | 35 | 25 |
| Traffic Fatalities | (per year) | 2 | 1 |

Note: the operating cost for a vehicle is unaffected by the project, and is \$4.

Table 2: Model Parameters

| | | |
|------------------------|-------------|-------------|
| Peak Value of Time | (\$/minute) | \$0.15 |
| Off-Peak Value of Time | (\$/minute) | \$0.10 |
| Value of Life | (\$/life) | \$3,000,000 |

What is the benefit-cost relationship?



Solution:

Benefits

A 50 minute trip at \$0.15/minute is \$7.50, while a 30 minute trip is only \$4.50. So for existing users, the expansion saves \$3.00/trip. Similarly in the off-peak, the cost of the trip drops from \$3.50 to \$2.50, saving \$1.00/trip.

Consumers' surplus increases both for the trips which would have been taken without the project and for the trips which are stimulated by the project (so-called "induced demand"), as illustrated above in Figure 1. Our analysis is divided into Old and New Trips, the benefits are given in Table 3.

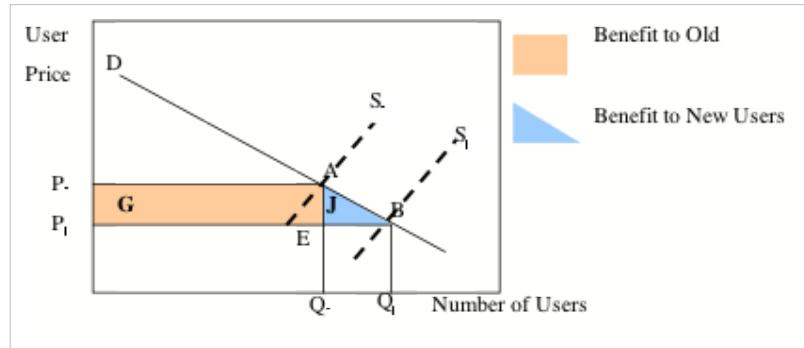


Figure 1: Change in Consumers' Surplus

Table 3: Hourly Benefits

| TYPE | Old trips | New Trips | Total |
|----------|-----------|-----------|----------|
| Peak | \$54,000 | \$9000 | \$63,000 |
| Off-peak | \$9,000 | \$500 | \$9,500 |

Note: Old Trips: For trips which would have been taken anyway the benefit of the project equals the value of the time saved multiplied by the number of trips. New Trips: The project lowers the cost of a trip and public responds by increasing the number of trips taken. The benefit to new trips is equal to one half of the value of the time saved multiplied by the

increase in the number of trips. There are 250 weekdays (excluding holidays) each year and four rush hours per weekday. There are 1000 peak hours per year. With 8760 hours per year, we get 7760 offpeak hours per year. These numbers permit the calculation of annual benefits (shown in Table 4).

Table 4: Annual Travel Time Benefits

| TYPE | Old trips | New Trips | Total |
|----------|---------------|--------------|---------------|
| Peak | \$54,000,000 | \$9,000,000 | \$63,000,000 |
| Off-peak | \$69,840,000 | \$3,880,000 | \$73,720,000 |
| Total | \$123,840,000 | \$12,880,000 | \$136,720,000 |

The safety benefits of the project are the product of the number of lives saved multiplied by the value of life. Typical values of life are on the order of \$3,000,000 in US transportation analyses. We need to value life to determine how to trade off between safety investments and other investments. While your life is invaluable to you (that is, I could not pay you enough to allow me to kill you), you don't act that way when considering chance of death rather than certainty. You take risks that have small probabilities of very bad consequences. You do not invest all of your resources in reducing risk, and neither does society. If the project is expected to save one life per year, it has a safety benefit of \$3,000,000. In a more complete analysis, we would need to include safety benefits from non-fatal accidents.

The annual benefits of the project are given in Table 5. We assume that this level of benefits continues at a constant rate over the life of the project.

Table 5: Total Annual Benefits

| Type of Benefit | Value of Benefits Per Year |
|-----------------|----------------------------|
| Time Saving | \$136,720,000 |
| Reduced Risk | \$3,000,000 |
| Total | \$139,720,000 |

Costs

Highway costs consist of right-of-way, construction, and maintenance. Right-of-way includes the cost of the land and buildings that must be acquired prior to construction. It does not consider the opportunity cost of the right-of-way serving a different purpose. Let the cost of right-of-way be \$100 million, which must be paid before construction starts. In principle, part of the right-of-way cost can be recouped if the highway is not rebuilt in place (for instance, a new parallel route is constructed and the old highway can be sold for development). Assume that all of the right-of-way cost is recoverable at the end of the thirty-year lifetime of the project. The \$1 billion construction cost is spread uniformly over the first four-years. Maintenance costs \$2 million per year once the highway is completed.

The schedule of benefits and costs for the project is given in Table 6.

Table 6: Schedule Of Benefits And Costs (\$ millions)

| Time (year) | Benefits | Right-of-way costs | Construction costs | Maintenance costs |
|-------------|----------|--------------------|--------------------|-------------------|
| 0 | 0 | 100 | 0 | 0 |
| 1-4 | 0 | 0 | 250 | 0 |
| 5-29 | 139.72 | 0 | 0 | 2 |
| 30 | 139.72 | -100 | 0 | 2 |

Conversion to Present Value

The benefits and costs are in constant value dollars. Assume the real interest rate (excluding inflation) is 2%. The following equations provide the present value of the streams of benefits and costs.

To compute the Present Value of Benefits in Year 5, we apply equation (2) from above.

$$P = A \left[\frac{(1+i)^n - 1}{i(1+i)^n} \right] = 139.72 \left[\frac{(1+0.02)^{26} - 1}{0.02(1+0.02)^{26}} \right] = 2811.31$$

To convert that Year 5 value to a Year 1 value, we apply equation (1)

$$P = \frac{F}{(1+i)^n} = \frac{2811.31}{(1+0.02)^4} = 2597.21$$

The present value of right-of-way costs is computed as today's right of way cost (\$100 M) minus the present value of the recovery of those costs in Year 30, computed with equation (1):

$$P = \frac{F}{(1+i)^n} = \frac{100}{(1+0.02)^{30}} = 55.21$$

$$100 - 55.21 = 44.79$$

The present value of the construction costs is computed as the stream of \$250M outlays over four years is computed with equation (2):

$$P = A \left[\frac{(1+i)^n - 1}{i(1+i)^n} \right] = 250 \left[\frac{(1+0.02)^4 - 1}{0.02(1+0.02)^4} \right] = 951.93$$

Maintenance Costs are similar to benefits, in that they fall in the same time periods. They are computed the same way, as follows: To compute the Present Value of \$2M in Maintenance Costs in Year 5, we apply equation (2) from above.

$$P = A \left[\frac{(1+i)^n - 1}{i(1+i)^n} \right] = 2 \left[\frac{(1+0.02)^{26} - 1}{0.02(1+0.02)^{26}} \right] = 40.24$$

To convert that Year 5 value to a Year 1 value, we apply equation (1)

$$P = \frac{F}{(1+i)^n} = \frac{40.24}{(1+0.02)^4} = 37.18$$

As Table 7 shows, the benefit/cost ratio of 2.5 and the positive net present value of \$1563.31 million indicate that the project is worthwhile under these assumptions (value of time, value of life, discount rate, life of the road). Under a different set of assumptions, (e.g. a higher discount rate), the outcome may differ.

Table 7: Present Value of Benefits and Costs (\$ millions)

| | Present Value |
|--------------------|---------------|
| Benefits | 2,597.21 |
| Costs | |
| Right-of-Way | 44.79 |
| Construction | 951.93 |
| Maintenance | 37.18 |
| Costs SubTotal | 1,033.90 |
| Net Benefit (B-C) | 1,563.31 |
| Benefit/Cost Ratio | 2.5 |

Thought Question

Problem

Is money the only thing that matters in Benefit-Cost Analysis? Is "converted" money the only thing that matters? For example, the value of human life in dollars?

Solution

Absolutely not. A lot of benefits and costs can be converted to monetary value, but not all. For example, you can put a price on human safety, but how can you put a price on, say, aesthetics--something that everyone agrees is beneficial. What else can you think of?

Sample Problems

Problem (Solution)

Key Terms

- Benefit-Cost Analysis
- Profits
- Costs
- Discount Rate
- Present Value
- Future Value

External Exercises

Use the SAND software at the STREET website [3] to learn how to evaluate network performance given a changing network scenario.

End Notes

[1] benefit-cost analysis is sometimes referred to as cost-benefit analysis (CBA)

[2] <http://www.sjsu.edu/faculty/watkins/cba.htm>

[3] <http://street.umn.edu/>

- Aruna, D. Social Cost-Benefit Analysis Madras Institute for Financial Management and Research, pp. 124, 1980.
- Boardman, A. et al., Cost-Benefit Analysis: Concepts and Practice, Prentice Hall, 2nd Ed,

- Dorfman, R, "Forty years of Cost-Benefit Analysis: Economic Theory Public Decisions Selected Essays of Robert Dorfman", pp. 323, 1997.
- Dupuit, Jules. "On the Measurement of the Utility of Public Works R.H. Babcock (trans.)." International Economic Papers 2. London: Macmillan, 1952.
- Ekelund, R., Hebert, R. Secret Origins of Modern Microeconomics: Dupuit and the Engineers, University of Chicago Press, pp. 468, 1999.
- Flyvbjerg, B. et al. Megaprojects and Risk: An Anatomy of Ambition, Cambridge University Press, pp. 207, 2003.
- Gramlich, E., A Guide to Benefit-cost Analysis, Prentice Hall, pp. 273, 1981.
- Hicks, John (1941) "The Rehabilitation of Consumers' Surplus," Review of Economic Studies, pp. 108-116.
- Kaldor, Nicholas (1939) "Welfare Propositions of Economics and Interpersonal Comparisons of Utility," Economic Journal, 49:195, pp. 549-552.
- Layard, R., Glaister, S., Cost-Benefit Analysis, Cambridge University Press; 2nd Ed, pp. 507, 1994.
- Pareto, Vilfredo., (1906) Manual of Political Economy. 1971 translation of 1927 edition, New York: Augustus M. Kelley.
- Perksy, J., Retrospectives: Cost-Benefit Analysis and the Classical Creed Journal of Economic Perspectives, 2001 pp. 526, 2000.
- Treasury Board of Canada "Benefit-cost Analysis Guide", 1998 <http://classwebs.spea.indiana.edu/krutilla/v541/Benfit-Cost%20Guide.pdf>

Fundamentals of Transportation/ Planning

All forecasts are wrong, some forecasts are more wrong than others. - anonymous

Why Model?

Models may be undertaken for a variety of reasons, these include:

- Forecasting traffic
- Testing scenarios (alternative land uses, networks, policies)
- Planning projects/corridor studies
- Regulating land use: Growth management/public facility adequacy
- Managing complexity, when eyeballs are insufficient, (different people have different intuitions)
- Understanding travel behavior
- Influencing decisions
- Estimating traffic in the absence of data

Developing Models

As an engineer or planner you may be given a model to use. But that model was not spontaneously generated, it came from other engineers or planners who undertook a series of steps to translate raw data into a tool that could be used to generate useful information.

The first step, specification, tells us what depends on what (the number of trips leaving a zone may depend on the number of households). The estimation step tells us how strong are these relationships mathematically, (each household generates two trips in the peak hour). Implementation takes those relationships and puts them into a computer program. Calibration looks at the output of the computer program and compares it to other data, if they don't match exactly, adjustments to the models may be made. Validation compares the results to data at another point in time. Finally the model is applied to look at a project (how much traffic will use Granary Road after Washington Avenue is closed to traffic).

- Specification:

- $y = f(X)$

- Estimation:

- $y = mX + b$; $m=1$, $b=2$

- Implementation

- If $Z > W$, Then $y = mX + b$

- Calibration

- $y_{predicted} + k = y_{observed}$

- Validation

- $y_{predicted1990} + k = y_{observed1990}$

- Application

“Four-Step” Urban Transportation Planning Models

We want to answer a number of related questions (who, what, where, when, why, and how):

- Who is traveling or what is being shipped?
- Where are the origin and destination of those trips, shipments?
- When do those trips begin and end (how long do they take, how far away are the trip ends)?
- Why are the trips being made, what is their purpose?
- How are the trips getting there, what routes or modes are they taking?

If we know the answers to those questions, we want to know what are the answers to those questions a function of?

- Cost: Money, Time spent on the trip,
- Cost: Money and Time of alternatives.
- Benefit (utility) of trip (e.g. the activity at the destination)
- Benefit of alternatives

The reason for this is to understand what will happen under various circumstances:

- How much “induced demand” will be generated if a roadway is expanded?
- How many passengers will be lost if bus services are cut back?
- How many people will be “priced off” if tolls are implemented?
- How much traffic will a new development generate?

- How much demand will I lose if I raise shipping costs to my customers?

In short, for urban passenger travel, we are trying to predict the number of trips by:

- Origin Activity,
- Destination Activity,
- Origin Zone,
- Destination Zone,
- Mode,
- Time of Day, and
- Route.

This is clearly a multidimensional problem.

In practice, the mechanism to do this is to employ a "four-step" urban transportation planning model, each step will be detailed in subsequent modules. These steps are executed in turn, though there may be feedback between steps:

- Trip Generation - How many trips are entering or leaving zone i or j
- Trip Distribution or Destination Choice - How many trips are going from zone i to zone j
- Mode Choice - How many trips from i to j are using mode m
- Route Choice - Which links are trips from i to j by mode m using route r

Thought Questions

- Is past behavior reflective of future behavior?
- Can the future be predicted?
- Is the future independent of decisions, or are prophesies self-fulfilling?
- How do we know if forecasts were successful?
- Against what standard are they to be judged?
- What values are embedded in the planning process?
- What happens when values change?

Variables

- msXX - scalar matrix
- moXX - origin vector matrix
- mdXX - destination vector matrix
- mfXX - full vector matrix
- T_{ij} - Total number of trips between origin i and destination j
- K_i - Calibration parameter for origin i
- K_j - Calibration paarmeter for destination j
- T_i - Trips leaving origin i
- T_j - Trips arriving at destination j
- T'_j - Effective Trips arriving at destination j , computed as a result for calibration to the next iteration
- $f(C_{ij})$ - Cost function between origin i and destination j
- TC - Total Cost
- Q - Total flow

Abbreviations

- SOV - single occupant vehicle
- HOV - high occupancy vehicle (2+, 3+, etc.)
- TAZ - transportation analysis zone or traffic analysis zone

Key Terms

- Rational Planning
- Transportation planning model
- Matrix, Full Matrix, Vector Matrix, Scalar Matrix
- Trip table
- Travel time matrix
- Origin, Destination
- Purpose
- Network
- Zone (Traffic Analysis Zone or Transportation Analysis Zone, or TAZ)
- External station or external zone
- Centroid
- Node
- Link
- Turn
- Route
- Path
- Mode

Fundamentals of Transportation/ Operations

Almost anybody can build a road. Its safety and durability would likely come into question if done by an amateur, but roads are not a difficult thing to deploy. Management of these roads, however, is something that is far from simple. Safety, efficiency, and sensibility are all elements that come into play when managing a road. It is seldom possible to find a managing strategy that satisfies all three of these categories, but often a "good" strategy can be found by balancing tradeoffs. This section introduces the fundamentals of traffic management and operation, discussing everything from queueing theory to traffic signals. The discussions here cover the fundamentals for traffic operations. To this day, research is still being conducted to perfect these elements.

Queueing

Queueing is the study of traffic behavior near a certain section where demand exceeds available capacity. Queues can be seen in a variety of situations and transportation systems. Its presence is one of the causes of driver delay.

Traffic Flow

Traffic Flow is the study of the movement of individual drivers and vehicles between two points and the interactions they make with one another. While traffic flow cannot be predicted with one-hundred percent certainty due to unpredictable driver behavior, drivers tend to behave within a reasonable range that can be represented mathematically. Relationships have been established between the three main traffic flow characteristics:

- Flow
- Density
- Velocity

Queueing and Traffic Flow

Queueing and Traffic Flow is the study of traffic flow behaviors surrounding queueing events. They merge concepts learned in the prior two sections and apply them to bottlenecks.

Shockwaves

Shockwaves are the byproducts of traffic congestion and queueing. They are defined as transition zones between two traffic states that are dynamic, meaning they generally have the ability to move. Most drivers can identify shockwaves as the propagation of brake lights stemming from a given incident.

Traffic Signals

Traffic Signals are one of the more familiar types of traffic control. Using either a fixed or adaptive schedule, traffic signals allow certain parts of the intersection to move while forcing other parts to wait, delivering instructions to drivers through a set of colorful lights. While many benefits come with traffic signals, they also come with a series of costs, such as increasing delay during the off-peak. But, traffic signals are generally a well-accepted form of traffic control for busy intersections and continue to be deployed.

Additional Areas of Study

Level of Service

One of the basic assessments of roadway performance is by determining its Level of Service (LOS). These Levels of Service, typically ranging between A and F (A being good performance, F being poor performance), are assessed by certain predetermined thresholds for any of the three characteristics of traffic flow (flow, density, and/or velocity). Finding the correct LOS often requires reference to a table.

When computing a value of flow to determine LOS, the following formula is used:

$$v_p = \frac{V}{PHF * N * f_{HV} * f_p}$$

Where:

- v_p = Actual Volume per Lane
- V = Hourly Volume
- PHF = Peak Hour Factor
- N = Number of Lanes
- f_{HV} = Heavy Vehicle Factor
- f_p = Driver Familiarity Factor (generally, a value of 1 for commuters, less for out-of-town drivers)

Peak Hour Factor is found to be:

$$PHF = V/(4V_{15})$$

Where:

- V = Hourly Volume
- V_{15} = Peak 15-Minute Flow

The Heavy Vehicle Factor is found to be:

$$f_{HV} = \frac{1}{1 + P_T(E_T - 1) + P_R(E_R - 1)}$$

Where:

- P_T = Percentage of Trucks, in decimal
- E_T = Passenger Car Equivalent of Trucks (found through table)
- P_R = Percentage of Recreational Vehicles, in decimal
- E_R = Passenger car Equivalent of Recreational Vehicles (found through table)

Similarly, LOS can be analyzed the anticipated speed of the corridor. The equation below produces the estimated free-flow speed for a strip of rural road, given certain characteristics about it:

$$FFS = BFFS - f_{LW} - f_{LC} - f_N - f_{ID}$$

Where:

- FFS = Estimated Free Flow Speed
- $BFFS$ = Base Free Flow Speed
- f_{LW} = Adjustment for Lane Width (found in tables)
- f_{LC} = Adjustment for Lateral Clearance (found in tables)
- f_N = Adjustment for Number of Lanes (found in tables)
- f_{ID} = Adjustment for Interchange Density (found in tables)

For multilane urban highways, free-flow speed is computed with a similar equation:

$$FFS = BFFS - f_{LW} - f_{LC} - f_M - f_A$$

Where:

- FFS = Estimated Free Flow Speed
- $BFFS$ = Base Free Flow Speed
- f_{LW} = Adjustment for Lane Width (found in tables)
- f_{LC} = Adjustment for Lateral Clearance (found in tables)
- f_M = Adjustment for Median Type (found in tables)
- f_A = Adjustment for Access Points (found in tables)

Lateral Clearance is computed:

$$TLC = LC_R + LC_L$$

Where:

- TLC = Total Lateral Clearance
- LC_R = Lateral Clearance on the Right Side of the Travel Lanes to Obstructions
- LC_L = Lateral Clearance on the Left Side of the Travel Lanes to Obstructions

Once TLC is computed, a table can provide the value for the Lateral Clearance Adjustment.

Estimating Gaps

One area of transportation operations that is still being considered is the estimation of gaps between vehicles. This is applicable in a variety of situations, primarily dealing with arrivals of vehicles at a given intersection. Gaps can only be predicted through a probability, based on several known characteristics occurring at a given time or location.

One of the more popular models used for predicting arrivals is the Poisson Model. This model is shown here:

$$P(x) = \frac{\lambda t^x e^{-\lambda t}}{x!} \text{ for } x = 0, 1, 2, \infty$$

Where:

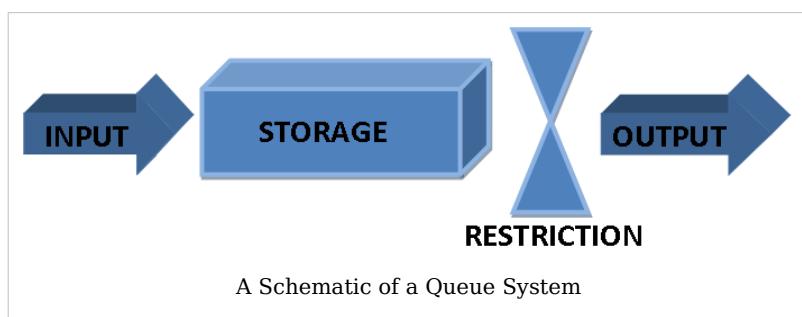
- $P(x)$ = Probability of having x vehicles arrive in time t
- λ = Average vehicle flow or arrival rate
- t = Duration of the time interval over which vehicles are counted
- x = Number of vehicles to arrive in time t

Variables

- v_p - Actual Volume per Lane
- V - Hourly Volume
- PHF - Peak Hour Factor
- N - Number of Lanes
- f_{HV} - Heavy Vehicle Factor
- f_p - Driver Familiarity Factor (generally, a value of 1 for commuters, less for out-of-town drivers)
- V_{15} - Peak 15-Minute Flow
- P_T - Percentage of Trucks, in decimal
- E_T - Passenger Car Equivalent of Trucks (found through table)
- P_R - Percentage of Recreational Vehicles, in decimal
- E_R - Passenger car Equivalent of Recreational Vehicles (found through table)
- FFS - Estimated Free Flow Speed
- $BFFS$ - Base Free Flow Speed
- f_{LW} - Adjustment for Lane Width (found in tables)
- f_{LC} - Adjustment for Lateral Clearance (found in tables)
- f_N - Adjustment for Number of Lanes (found in tables)
- f_{ID} - Adjustment for Interchange Density (found in tables)
- f_M - Adjustment for Median Type (found in tables)
- f_A - Adjustment for Access Points (found in tables)
- TLC - Total Lateral Clearance
- LC_R - Lateral Clearance on the Right Side of the Travel Lanes to Obstructions
- LC_L - Lateral Clearance on the Left Side of the Travel Lanes to Obstructions
- $P(x)$ - Probability of having x vehicles arrive in time t
- λ - Average vehicle flow or arrival rate
- t - Duration of the time interval over which vehicles are counted
- x - Number of vehicles to arrive in time t

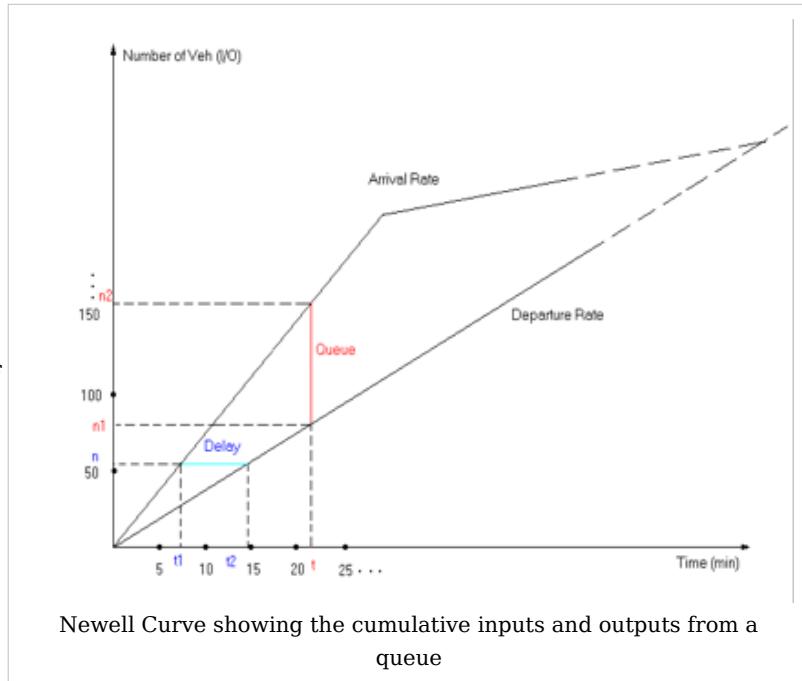
Fundamentals of Transportation/ Queueing

Queueing^[1] is the study of traffic behavior near a certain section where demand exceeds available capacity. Queues can be seen in many common situations: boarding a bus or train or plane, freeway bottlenecks, shopping checkout, exiting a doorway at the end of class, waiting for a computer in the lab, a hamburger at McDonald's, or a haircut at the barber. In transportation engineering, queueing can occur at red lights, stop signs, bottlenecks, or any design-based or traffic-based flow constriction. When not dealt with properly, queues can result in severe network congestion or "gridlock" conditions, therefore making them something important to be studied and understood by engineers.



Cumulative Input-Output Diagram (Newell Curve)

Based on the departure rate and arrival rate pair data, the delay of every individual vehicle can be obtained. Using the input-output (I/O) queueing diagram shown in the side figure, it is possible to find the delay for every individual vehicle: the delay of the i^{th} vehicle is time of departure - time of arrival ($t_2 - t_1$). Total delay is the sum of the delays of each vehicle, which is the area in the triangle between the arrival ($A(t)$) and departure ($D(t)$) curves.



Distributions

Arrival Distribution - Deterministic (uniform) OR Random (e.g. Poisson)

Service Distribution - Deterministic OR Random

Service Method:

- First Come First Serve (FCFS) or First In First Out (FIFO)
- Last Come First Served (LCFS) or Last In First Out (LIFO)

- Priority (e.g. HOV bypasses at ramp meters)

Characterizing Queues

Queue Length Characteristics - Finite or Infinite

Number of Channels - Number of Waiting Lines (e.g. Ramps = 2, Supermarket = 12)

We use the following notation:

- Arrival Rate = λ
- Departure Rate = μ

$$\text{Service time } \rho = \frac{\lambda}{\mu}$$

Degree of Saturation

Oversaturated: $\lambda > \mu$

Undersaturated $\lambda < \mu$

Saturated $\lambda = \mu$

Little's Formula

Little's Formula: $Q = \lambda w$

This means that the average queue size (measured in vehicles) equals the arrival rate (vehicles per unit time) multiplied by the average waiting time (in units of time). This result is independent of particular arrival distributions and, while perhaps obvious, is an important fundamental principle that was not proven until 1961.

See Wikipedia article [2] for more applications.



A Traffic Queue

Uncapacitated Queues (M/D/1) and (M/M/1)

It has been shown that queue sizes, waiting times, and delays differ between M/D/1 and M/M/1 queueing. These differences are represented in formulas and shown below. Note the minor differences between the two.

Comparison of M/D/1 and M/M/1 queue properties

| | M/D/1 | M/M/1 |
|----------------------------|-----------------------------------|--|
| Q (average queue size (#)) | $Q = \frac{\rho^2}{2(1-\rho)}$ | $Q = \frac{\rho^2}{1-\rho}$ |
| w (average waiting time) | $w = \frac{\rho}{2\mu(1-\rho)}$ | $w = \frac{\lambda}{\mu(\mu-\lambda)}$ |
| t (average total delay) | $t = \frac{2-\rho}{2\mu(1-\rho)}$ | $t = \frac{1}{(\mu-\lambda)}$ |

Notes:

- Average queue size includes customers currently being served (in number of units)
- Average wait time excludes service time
- Average delay time is (wait time + service time)

Uncapacitated queues (M/M/1) (random arrival and random service)

In addition to the properties stated before, M/M/1 queueing have a few additional ones of which to take note.

Additional M/M/1 queue properties

| | M/M/1 |
|--|---|
| Probability of n units in the system | $P(n) = \rho^n (1 - \rho)$ |
| Expected number of units in the system | $E(n) = \frac{\lambda}{\mu - \lambda}$ |
| Mean Queue Length | $E(m) = \frac{\lambda^2}{\mu(\mu - \lambda)}$ |
| Average waiting time of arrival, including queue and service | $E(v) = \frac{1}{(\mu - \lambda)}$ |
| Average waiting time in queue | $E(w) = \frac{\lambda}{\mu(\mu - \lambda)}$ |
| Probability of spending time t or less in system | $P(v \leq t) = 1 - e^{-(1-\rho)\lambda t}$ |
| Probability of spending time t or less in queue | $P(w \leq t) = 1 - \rho e^{-(1-\rho)\lambda t}$ |
| Probability of more than N vehicles in queue | $P(n > N) = \rho^{N+1}$ |

Capacitated Queues (Finite)

Capacitated queues permit a maximum number of vehicles to wait, and thus have different properties than uncapacitated queues. For single channel undersaturated finite queues where the maximum number of units in system is specified as N and with random arrivals and departures (λ, μ) we have:

Additional M/M/1 queue properties

| | M/M/1 |
|------------------------------------|---|
| Probability of n Units in System | $P(n) = \frac{(1 - \rho)}{1 - \rho^{N+1}} (\rho)^n$ |
| Expected Number of Units in System | $E(n) = \frac{(\rho)}{(1 - \rho)} \frac{1 - (N + 1)(\rho)^N + N\rho^{N+1}}{1 - \rho^{N+1}}$ |

Real Life Causes of Queue Generation

For Roads:

- Geometric Bottlenecks (lane drops, hard curves, hills)
- Accidents and Incidents
- Traffic Signals and Intersection Controls
- At-Grade Crossings with other Modes (Railroad crossings, drawbridges, etc.)
- Toll Booths
- Ramp Meters
- "Gawker" Effect
- Inclement Weather

For Trains and Transit:

- Platform Capacities
- Fare Gates
- Ticket Windows/Ticket Machines
- Minimum Safe Separation for Trains
- Security Checkpoints
- Efficiency of Trains Entering and Leaving Station (number of tracks, switches, etc.)

For Aviation and Airports:

- Runways
- Designated Minimum Safe Following Distances for Planes (by government)
- Physical Minimum Safe Following Distance for Planes (creation of turbulence, etc.)
- Available Airspace for Approaches and Departures
- Ticketing Counters/Check-in Procedures
- Security Checkpoints
- Baggage Systems
- Terminal Capacity for Planes
- Internal Terminal Capacity for Passengers
- Inclement Weather

For Water and Maritime:

- Locks and Dams
- Port Capacities
- Minimum "Safe" Distances (determined by government and physics)
- Inclement Weather

For Space Flight:

- Launch Capacity
- Minimum Spacings between Orbital Vehicles
- Inclement Weather on Earth
- Unfavorable Celestial Conditions

Examples

Example 1



Problem:

At the Krusty-Burger, if the arrival rate is 1 customer every minute and the service rate is 1 customer every 45 seconds, find the average queue size, the average waiting time, and average total delay. Assume an M/M/1 process.



Solution:

To proceed, we convert everything to minutes.

Service time:

$$\rho = \frac{\lambda}{\mu} = \frac{60/60}{60/45} = \frac{1}{1.33} = 0.75$$

Average queue size (Q):

$$Q = \frac{\rho^2}{1 - \rho} = \frac{(0.75)^2}{1 - 0.75} = 2.25$$

$$Q = \lambda w \text{ (within rounding error)}$$

Average wait time:

$$w = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{1}{1.33(1.33 - 1)} = 2.25$$

Average delay time:

$$t = \frac{1}{(\mu - \lambda)} = \frac{1}{(1.33 - 1)} = 3 = w + \text{servicetime} = 2.25 + 0.75$$

Comparison:

We can compute the same results using the M/D/1 equations, the results are shown in the Table below.

Comparison of M/D/1 and M/M/1 queue properties

| | M/D/1 | M/M/1 |
|----------------------------|-------|-------|
| Q (average queue size (#)) | 1.125 | 2.25 |
| w (average waiting time) | 1.125 | 2.25 |
| t (average total delay) | 1.88 | 3 |

As can be seen, the delay associated with the more random case (M/M/1, which has both random arrivals and random service) is greater than the less random case (M/D/1), which is to be expected.

Example 2


Problem:

How likely was it that Homer got his pile of hamburgers in less than 1, 2, or 3 minutes?


Solution:

$$\begin{aligned} P(t \leq 1) &= 1 - e^{-(1-\rho)\lambda T} = 1 - e^{-(1-0.75)1*1} = 1 - e^{-0.25} = 0.22 \\ P(t \leq 2) &= 1 - e^{-(1-\rho)\lambda T} = 1 - e^{-(1-0.75)1*2} = 1 - e^{-0.5} = 0.39 \\ P(t \leq 3) &= 1 - e^{-(1-\rho)\lambda T} = 1 - e^{-(1-0.75)1*3} = 1 - e^{-0.75} = 0.52 \end{aligned}$$

Example 3


Problem:

Before he encounters the “pimply faced teen” who servers burgers, what is the likelihood that Homer waited more than 3 minutes?


Solution:

$$\begin{aligned} P(w \leq 3) &= 1 - \rho e^{-(1-\rho)\lambda T} = 1 - 0.75e^{-(1-0.75)1*3} = 0.645 \\ P(w > 3) &= 1 - P(w \leq 3) = 1 - 0.645 = 0.355 \end{aligned}$$

Example 4


Problem:

How likely is it that there were more than 5 customers in front of Homer?


Solution:

$$P(n > 5) = \rho^{N+1} = 0.75^6 = 0.178$$

Thought Question

Problem

How does one minimize wait time at a queue?

Solution

Cutting in line always helps, but this problem will be answered without breaking any rules. Think about going out to dinner, only to find a long line at your favorite restaurant. How do you deal with that? Maybe nothing can be done at that time, but the next time you go to that restaurant, you might pick a new time. Perhaps an earlier one to avoid the lunch or dinner rush. Similar decisions can be seen in traffic. People that are tired of being in network queues on their way to work may attempt to leave earlier or (if possible) later than rush hour to decrease their own travel time. This typically works well until all the other drivers figure out the same thing and shift congestion to a different time.

Sample Problems

Sample Problem 1 : Queueing at a Tollbooth

Problem (Solution)

Sample Problem 2 : Queueing at a Ramp Meter

Problem (Solution)

Sample Problem 3 : Queueing at a Ramp Meter

Problem (Solution)

Variables

- $A(t) = \lambda$ - Arrival Rate
- $D(t) = \mu$ - Departure Rate
- $1/\mu$ - service time
- $\rho = \lambda / \mu$ - Utilization
- Q - average queue size including customers currently being served (in number of units)
- w - average wait time
- t - average delay time (queue time + service time)

Key Terms

- Queueing theory
- Cumulative input-output diagram (Newell diagram)
- average queue length
- average waiting time
- average total delay time in system
- arrival rate, departure rate
- undersaturated, oversaturated
- $D/D/1$, $M/D/1$, $M/M/1$
- Channels
- Poisson distribution,

- service rate
- finite (capacitated) queues, infinite (uncapacitated) queues

External Exercises

The assignment seeks to provide students with the opportunity to gain a better understanding of two queuing theories: M/D/1 and M/M/1. Two preformatted spreadsheets have been made available for assistance in computing the values sought after in this exercise. While these spreadsheets provide the computations for these results, the formula is listed below for reference:

$$WT_q = \left(\frac{C_\lambda^2 + C_\mu^2}{2C_\lambda^2} \right) \left(\frac{\rho}{1 - \rho} \right) \frac{1}{\mu}$$

Where:

- WT_q : Average customer delay in the queue
- C_λ : Coefficient of variation (CV) of the arrival distribution
- C_μ : Coefficient of variation of the departure distribution
- CV : Standard deviation/mean; $CV = (1/\text{SqRt (mean)})$ for Poisson process and $CV = 0$ for constant distribution
- μ : Average departure rate
- λ : Average arrival rate
- ρ : Utilization = Arrival rate/service rate ($\rho = \lambda/\mu$)

M/D/1 Queueing

Download the file for M/D/1 Queueing from the University of Minnesota's STREET website: M/D/1 Queue Spreadsheet [3]

With this spreadsheet, run 5 simulations for each of the 10 scenarios, using the arrival and departure information listed in the table below. In other words, program the same data into the spreadsheet 5 different times to capture a changing seed and, thus, produce slightly different answers because of the model's sensitivity. A total of 50 simulations will be run.

| Scenario | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------------|------|-------|------|-----|-----|-----|-----|-----|-----|-----|
| Arrival Rate | 0.01 | 0.025 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| Service Rate | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |

Also, find the utilization for all ten scenarios. Based on the utilization and the distribution variability, use the above equation to compute the average delays for all scenarios with utilization values of less than 1.

Finally, summarize the average-delays obtained both from the simulation and from the WT_q equation in the same delay-utilization plot. Interpret your results. How does the average user delay change as utilization increases? Does the above equation provide a satisfactory approximation of the average delays?

M/M/1 Queueing

Download the file for M/M/1 Queueing from the University of Minnesota's STREET website: M/M/1 Queue Spreadsheet [4]

With this spreadsheet, run 5 simulations for each of the 10 scenarios, using the arrival and departure information listed in the table below. In other words, program the same data into the spreadsheet 5 different times to capture a changing seed and, thus, produce slightly

different answers because of the model's sensitivity. A total of 50 simulations will be run.

| Scenario | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
| Arrival Rate | 0.01 | 0.025 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| Service Rate | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |

Again, you need to use five different random seeds for each scenario. Summarize the results in a delay-utilization plot. Interpret your results (You do NOT need to use the WTq equation given above to compute delays for the M/M/1 queue).

Additional Questions

- Finally compare the M/M/1 queue and the M/D/1 queue. What conclusion can you draw? (For the same average arrival rate, do users experience the same delays in the two queuing systems? Why or why not?).
- Provide a brief example where M/M/1 might be the appropriate model to use.
- Provide a brief example where M/D/1 might be the appropriate model to use.

End Notes

Queueing is the only common English word with 5 vowels in a row.

It has been posited: Cooeeing - To call out "cooee," which is apparently something done in Australia.

The uncommon word: archaeoaelotrophic has 6 vowels - a prehistorical item that is unequally elastic in different directions - One suspects it is just made up to have a word with 6 vowels in a row though, and the "ae" is questionable anyway.

- [2] http://en.wikipedia.org/wiki/Little's_Law
- [3] <http://street.umn.edu/FOT/MD1Queue.xls>
- [4] <http://street.umn.edu/FOT/MM1Queue.xls>

Fundamentals of Transportation/ Queueing/Problem1


Problem:

Application of Single-Channel Undersaturated Infinite Queue Theory to Tollbooth Operation. Poisson Arrival, Negative Exponential Service Time

- Arrival Rate = 500 vph,
- Service Rate = 700 vph

Determine

- Percent of Time operator will be free
- Average queue size in system
- Average wait time for vehicles that wait

Note: For operator to be free, vehicles must be 0

- Solution

Fundamentals of Transportation/ Queueing/Solution1


Problem:

Application of Single-Channel Undersaturated Infinite Queue Theory to Tollbooth Operation. Poisson Arrival, Negative Exponential Service Time

- Arrival Rate = 500 vph,
- Service Rate = 700 vph

Determine

- Percent of Time operator will be free
- Average queue size in system
- Average wait time for vehicles that wait

Note: For operator to be free, vehicles must be 0



$$P(n) = \rho^n (1 - \rho) = (0.714)^0 (1 - 0.714) = 28.6\%$$

$$Q = \frac{\rho^2}{1 - \rho} = \frac{0.510}{0.286} = 1.78$$

$$w = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{500}{700(700 - 500)} = 0.003571 \text{ hours} = 12.8571 \text{ sec}$$

$$Q = \lambda w = 500 * 0.00357 = 1.785$$

$$\text{ServiceTime} = \frac{1}{\mu} = \frac{1}{700} = 0.001429 \text{ hours} = 5.142 \text{ sec}$$

$$t = \frac{1}{(\mu - \lambda)} = \frac{1}{(700 - 500)} = 0.005 \text{ hours} = 18 \text{ seconds}$$

Fundamentals of Transportation/ Queueing/Problem2


Problem:

A ramp has an arrival rate of 200 cars an hour and the ramp meter only permits 250 cars per hour, while the ramp can store 40 cars before spilling over.

- (A) What is the probability that it is half-full, empty, full?
- (B) How many cars do we expect on the ramp?

- Solution

Fundamentals of Transportation/ Queueing/Solution2


Problem:

A ramp has an arrival rate of 200 cars an hour and the ramp meter only permits 250 cars per hour, while the ramp can store 40 cars before spilling over.

- (A) What is the probability that it is half-full, empty, full?
- (B) How many cars do we expect on the ramp?

Solution:


$$\rho = 200/250 = 0.8$$

Part (A)

What is the probability that it is half-full, empty, full?

Half-full

$$P(n) = \frac{(1 - \rho)}{1 - \rho^{N+1}} (\rho)^n \rightarrow P(20) = \frac{(1 - 0.8)}{1 - 0.8^{40+1}} (0.8)^{20} = 0.0023$$

Empty

$$P(n) = \frac{(1 - \rho)}{1 - \rho^{N+1}} (\rho)^n \rightarrow P(0) = \frac{(1 - 0.8)}{1 - 0.8^{40+1}} (0.8)^0 = 0.20$$

Full

$$P(n) = \frac{(1 - \rho)}{1 - \rho^{N+1}} (\rho)^n \rightarrow P(40) = \frac{(1 - 0.8)}{1 - 0.8^{40+1}} (0.8)^{40} = 2.65 * 10^{-5}$$

Part (B)

How many cars do we expect on the ramp?

$$E(n) = \frac{(\rho)}{(1 - \rho)} \frac{1 - (N + 1)(\rho)^N + N\rho^{N+1}}{1 - \rho^{N+1}} = \frac{(0.8)}{(1 - 0.8)} \frac{1 - (40 + 1)(0.8)^{40} + 40(0.8)^{40+1}}{1 - (0.8)^{40+1}} = 4$$

Fundamentals of Transportation/ Queueing/Problem3


Problem:

In this problem we apply the properties of capacitated queues to an expressway ramp. Note the following

- Ramp will hold 15 vehicles
- Vehicles can enter expressway at 1 vehicle every 6 seconds
- Vehicles arrive at ramp at 1 vehicle every 8 seconds

Determine:

- (A) Probability of 5 cars,
- (B) Percent of Time Ramp is Full,
- (C) Expected number of vehicles on ramp in peak hour.
- Solution

Fundamentals of Transportation/ Queueing/Solution3


Problem:

In this problem we apply the properties of capacitated queues to an expressway ramp. Note the following:

- Ramp will hold 15 vehicles
- Vehicles can enter expressway at 1 vehicle every 6 seconds
- Vehicles arrive at ramp at 1 vehicle every 8 seconds

Determine:

- (A) Probability of 5 cars,
- (B) Percent of Time Ramp is Full,
- (C) Expected number of vehicles on ramp in peak hour.


Solution:

$$\lambda = \frac{3600}{8} = 450$$

$$\mu = \frac{3600}{6} = 600$$

$$\rho = \frac{\lambda}{\mu} = 0.75$$

Part A

Probability of 5 cars,

$$P(n) = \frac{(1 - \rho)}{1 - \rho^{N+1}} (\rho)^n = \frac{(1 - 0.75)}{1 - 0.75^{16}} (0.75)^5 = 0.06 = 6\%$$

Part B

Percent of time ramp is full (i.e. 15 cars),

$$P(n) = \frac{(1 - \rho)}{1 - \rho^{N+1}} (\rho)^n = \frac{(1 - 0.75)}{1 - 0.75^{16}} (0.75)^{15} = 0.0033 = 0.33\%$$

Part C

Expected number of vehicles on ramp in peak hour.

$$E(n) = \frac{(\rho)}{(1 - \rho)} \frac{1 - (N + 1)(\rho)^N + N\rho^{N+1}}{1 - \rho^{N+1}} = \frac{(0.75)}{(1 - 0.75)} \frac{1 - (15 + 1)(\rho)^{15} + 15\rho^{16}}{1 - \rho^{16}} = 2.81 \approx 3$$

Conclusion, ramp is large enough to hold most queues, though 12 seconds an hour, there will be some ramp spillover. It is a policy question as to whether that is acceptable.

Fundamentals of Transportation/ Traffic Flow

Traffic Flow is the study of the movement of individual drivers and vehicles between two points and the interactions they make with one another. Unfortunately, studying traffic flow is difficult because driver behavior is something that cannot be predicted with one-hundred percent certainty. Fortunately, drivers tend to behave within a reasonably consistent range and, thus, traffic streams tend to have some reasonable consistency and can be roughly represented mathematically. To better represent traffic flow, relationships have been established between the three main characteristics: (1) flow, (2) density, and (3) velocity. These relationships help in planning, design, and operations of roadway facilities.

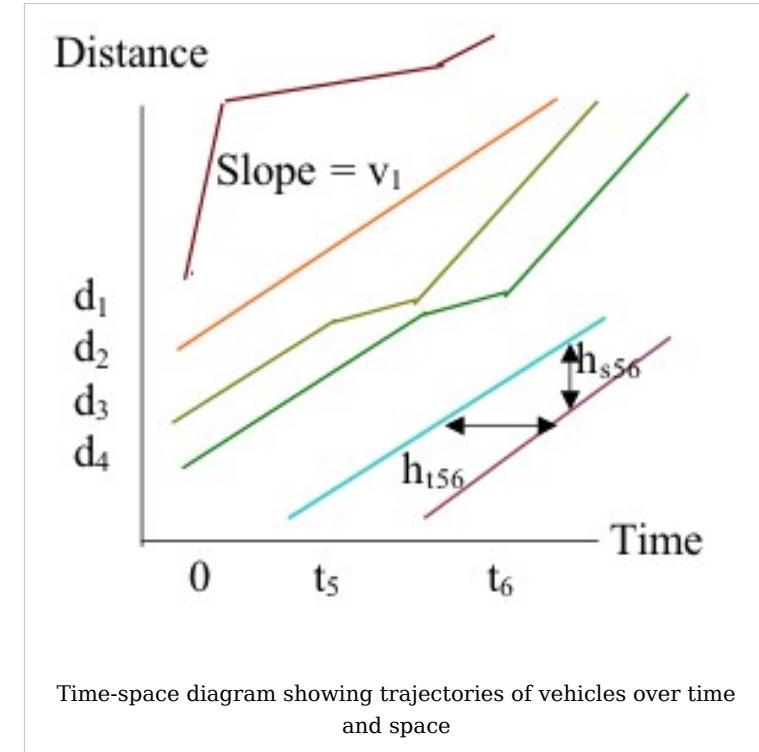
Traffic flow theory

Time-Space Diagram

Traffic engineers represent the location of a specific vehicle at a certain time with a time-space diagram. This two-dimensional diagram shows the trajectory of a vehicle through time as it moves from a specific origin to a specific destination. Multiple vehicles can be represented on a diagram and, thus, certain characteristics, such as flow at a certain site for a certain time, can be determined.

Flow and Density

Flow (q) = the rate at which vehicles pass a fixed point (vehicles per hour)



$$q = \frac{3600N}{t_{measured}}$$

Density (Concentration) (k) = number of vehicles (N) over a stretch of roadway (L) (in units of vehicles per kilometer)^[1]

$$k = \frac{N}{L}$$

where:

- N = number of vehicles passing a point in the roadway in $t_{measured}$ sec
- q = equivalent hourly flow
- L = length of roadway
- k = density

Speed

Measuring speed of traffic is not as obvious as it may seem; we can average the measurement of the speeds of individual vehicles over time or over space, and each produces slightly different results.

Time mean speed

Time mean speed (\bar{v}_t) = arithmetic mean of speeds of vehicles passing a point

$$\bar{v}_t = \frac{1}{N} \sum_{n=1}^N v_n$$

Space mean speed

Space mean speed (\bar{v}_s) is defined as the harmonic mean of speeds passing a point during a period of time. It also equals the average speeds over a length of roadway

$$\bar{v}_s = \frac{N}{\sum_{n=1}^N \frac{1}{v_n}}$$

Headway

Time headway

Time headway (h_t) = difference between the time when the front of a vehicle arrives at a point on the highway and the time the front of the next vehicle arrives at the same point (in seconds)

Average Time Headway (\bar{h}_t) = Average Travel Time per Unit Distance * Average Space Headway

$$\bar{h}_t = \bar{t} * \bar{h}_s$$

Space headway

Space headway (h_s) = difference in position between the front of a vehicle and the front of the next vehicle (in meters)

Average Space Headway (\bar{h}_s) = Space Mean Speed * Average Time Headway

$$\bar{h}_s = \bar{v}_s * \bar{h}_t$$

Note that density and space headway are related:

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

Fundamental Diagram of Traffic Flow

The variables of flow, density, and space mean speed are related definitionally as:

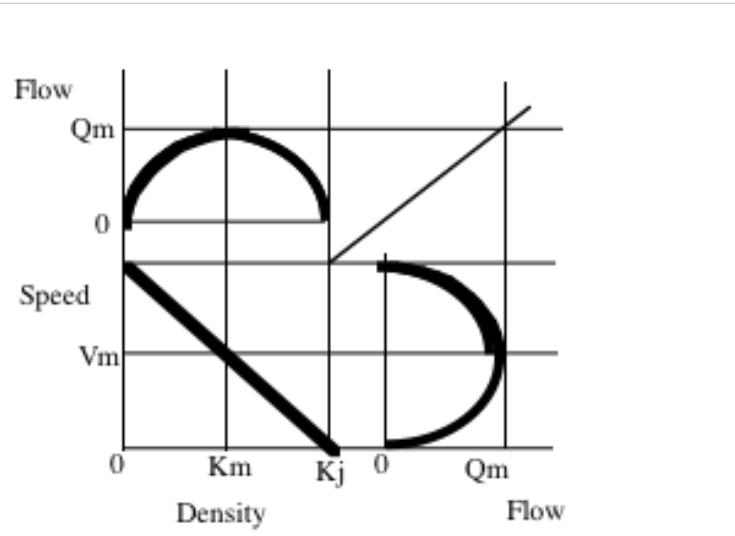
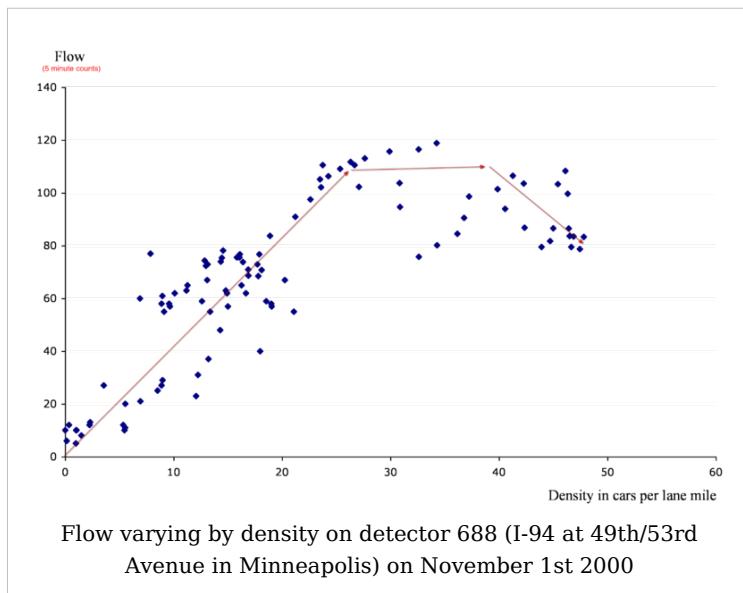


FIGURE 2: Flow-density, and speed-concentration curves
(assuming single-regime, linear speed-concentration model)

Traditional fundamental diagram of traffic



$$q = k\bar{v}_s$$

Traditional Model (Parabolic)

Properties of the traditional fundamental diagram.

- When density on the highway is zero, the flow is also zero because there are no vehicles on the highway
- As density increases, flow increases
- When the density reaches a maximum jam density (k_j), flow must be zero because vehicles will line up end to end
- Flow will also increase to a maximum value (q_m), increases in density beyond that point result in reductions of flow.
- Speed is space mean speed.
- At density = 0, speed is freeflow (v_f). The upper half of the flow curve is uncongested, the lower half is congested.
- The slope of the flow density curve gives speed. Rise/Run = Flow/Density = Vehicles per hour / Vehicles per km = km / hour.

Observation (Triangular or Truncated Triangular)

Actual traffic data is often much noisier than idealized models suggest. However, what we tend to see is that as density rises, speed is unchanged to a point (capacity) and then begins to drop if it is affected by downstream traffic (queue spillbacks). For a single link, the relationship between flow and density is thus more triangular than parabolic. When we aggregate multiple links together (e.g. a network), we see a more parabolic shape.

Microscopic and Macroscopic Models

Models describing traffic flow can be classed into two categories: microscopic and macroscopic. Ideally, macroscopic models are aggregates of the behavior seen in microscopic models.

Microscopic Models

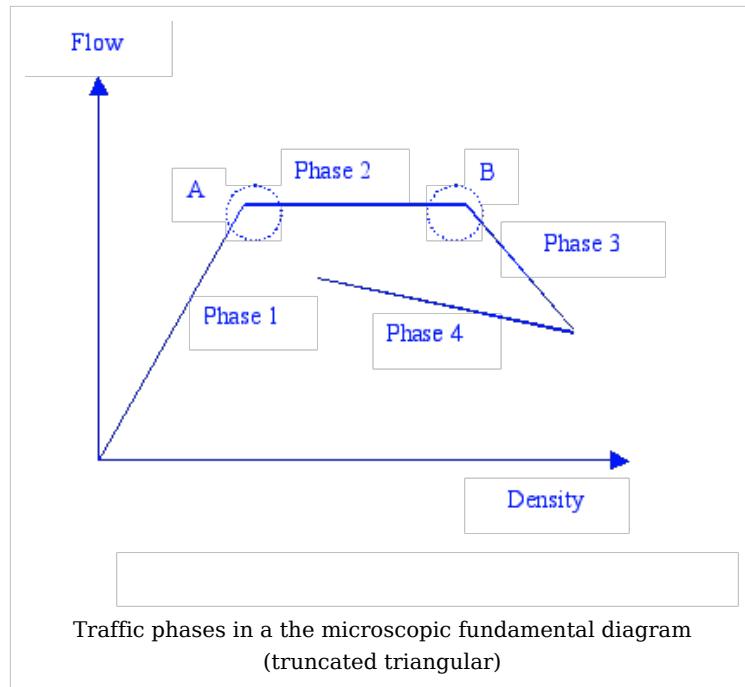
Microscopic models predict the following behavior of cars (their change in speed and position) as a function of the behavior of the leading vehicle.

Macroscopic Models

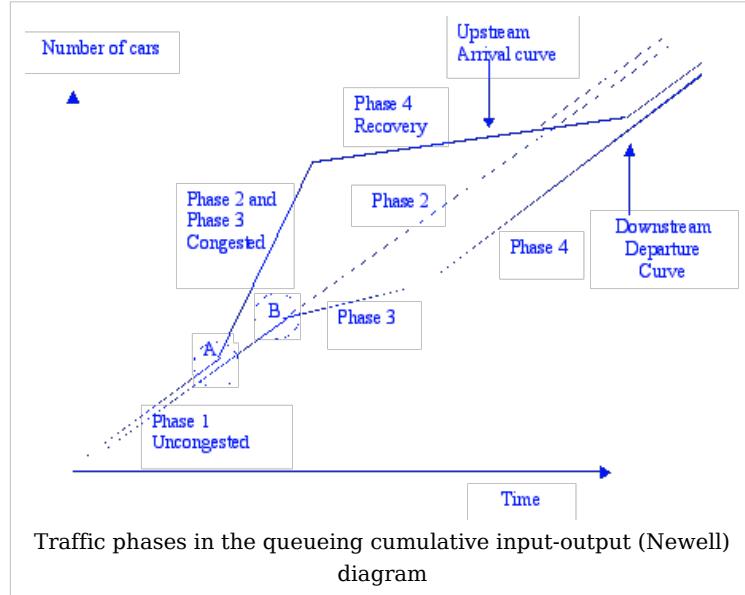
Macroscopic traffic flow theory relates traffic flow, running speed, and density. Analogizing traffic to a stream, it has principally been developed for limited access roadways (Leutzbach 1988). The fundamental relationship “ $q=kv$ ” (flow (q) equals density (k) multiplied by speed (v)) is illustrated by the fundamental diagram. Many empirical studies have quantified the component bivariate relationships (q vs. v , q vs. k , k vs. v), refining parameter estimates and functional forms (Gerlough and Huber 1975, Pensaud and Hurdle 1991; Ross 1991; Hall, Hurdle and Banks 1992; Banks 1992; Gilchrist and Hall 1992; Disbro and Frame 1992).

The most widely used model is the Greenshields model, which posited that the relationships between speed and density is linear. These were most appropriate before the advent of high-powered computers enabled the use of microscopic models. Macroscopic properties like flow and density are the product of individual (microscopic) decisions. Yet those microscopic decision-makers are affected by the environment around them, i.e. the macroscopic properties of traffic.

While traffic flow theorists represent traffic as if it were a fluid, queueing analysis essentially treats traffic as a set of discrete particles. These two representations are not-necessarily inconsistent. The figures to the right show the same 4 phases in the



Traffic phases in a the microscopic fundamental diagram
(truncated triangular)



Traffic phases in the queueing cumulative input-output (Newell)
diagram

fundamental diagram and the queueing input-output diagram. This is discussed in more detail in the next section.

Examples

Example 1: Time-Mean and Space-Mean Speeds



Problem:

Given five observed velocities (60 km/hr, 35 km/hr, 45 km/hr, 20 km/hr, and 50 km/hr), what is the time-mean speed and space-mean speed?



Solution:

Time-Mean Speed:

$$\bar{v}_t = \frac{1}{5}(60 + 35 + 45 + 20 + 50) = 42$$

Space-Mean Speed:

$$\bar{v}_s = \frac{N}{\sum_{n=1}^N \frac{1}{v_n}} = \frac{5}{\frac{1}{60} + \frac{1}{35} + \frac{1}{45} + \frac{1}{20} + \frac{1}{50}} = 36.37$$

The time-mean speed is 42 km/hr and the space-mean speed is 36.37 km/hr.

Example 2: Computing Traffic Flow Characteristics



Problem:

Given that 40 vehicles pass a given point in 1 minute and traverse a length of 1 kilometer, what is the flow, density, and time headway?



Solution:

Compute flow and density:

$$q = \frac{3600(40)}{60s} = 2400 \text{ veh/hr}$$

$$k = \frac{40}{1} = 40 \text{ veh/km}$$

Find space-mean speed:

$$q = k\bar{v}_s = 2400 = 40\bar{v}_s$$

$$\bar{v}_s = 60 \text{ km/hr}$$

Compute space headway:

$$k = 40 = \frac{1}{\bar{h}_s}$$

$$\bar{h}_s = 0.025 \text{ km} = 25 \text{ m}$$

Compute time headway:

$$\overline{h_s} = \overline{v_s} * \overline{h_t} = 25 = (60 * 1000 / 3600) \overline{h_t}$$

$$\overline{h_t} = 1.5s$$

The time headway is 1.5 seconds.

Thought Question

Problem

Microscopic traffic flow simulates the behaviors of individual vehicles while macroscopic traffic flow simulates the behaviors of the traffic stream overall. Conceptually, it would seem that microscopic traffic flow would be more accurate, as it would be based off of driver behavior than simply flow characteristics. Assuming microscopic simulation could be calibrated to truly account for driver behaviors, what is the primary drawback to simulating a large network?

Solution

Computer power. To simulate a very large network with microscopic simulation, the number of vehicles that needed to be assessed is very large, requiring a lot of computer memory. Current computers have issues doing very large microscopic networks in a timely fashion, but perhaps future advances will do away with this issue.

Sample Problem

- Problem (Solution)

Variables

- d_n = distance of n^{th} vehicle
- t_n = travel time of n^{th} vehicle
- v_n = speed (velocity) of n^{th} vehicle
- $h_{t,nm}$ = time headway between vehicles n and m
- $h_{s,nm}$ = space (distance) headway between vehicles n and m
- q = flow past a fixed point (vehicles per hour)
- N = number of vehicles
- t_{measured} = time over which measurement takes place (number of seconds)
- t = travel time
- k = density (vehicles per km)
- L = length of roadway section (km)
- v_t = time mean speed
- v_s = space mean speed
- v_f = freeflow (uncongested speed)
- k_j = jam density
- q_m = maximum flow

Key Terms

- Time-space diagram
- Flow, speed, density
- Headway (space and time)
- Space mean speed, time mean speed
- Microscopic, Macroscopic

Supplementary Reading

- Revised Monograph on Traffic Flow Theory ^[2]

References

- Banks, J.H. (1991). Two Capacity Phenomenon at Freeway Bottlenecks: A Basis for Ramp Metering? *Transportation Research Record* 1320, pp. 83-90.
- Banks, James H. (1992). "Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations." *Transportation Research Record* 1225:53-60.
- Cassidy, M.J. and R.L. Bertini (1999) Some Traffic Features at Freeway Bottlenecks. *Transportation Research Part B Vol. 33*, pp. 25-42
- Disbro, John E. and Frame, Michael. (1992). "Traffic Flow Theory and Chaotic Behavior." *Transportation Research Record*. 1225: 109-115.
- Gerlough, Daniel L. and Huber, Matthew J. (1975). *Traffic Flow Theory: A Monograph* TRB Special Report 165. Transportation Research Board, Washington DC.
- Gilchrist, Robert S. and Fred L. Hall. (1992). "Three-Dimensional Relationships Among Traffic Flow Theory Variables." *Transportation Research Record*. 1225:99-108.
- Hall, F.L. and K. Agyemang-Duah (1991). Freeway Capacity Drop and the Definition of Capacity. *Transportation Research Record* 1320, pp. 91-98
- Hall, Fred L., Hurdle, V. F., and Banks, James H. (1992). "A Synthesis of Recent Work on the Nature of Speed-Flow and Flow-Occupancy (or Density) Relationships on Freeways". presented at TRB 71st Annual Meeting, Washington DC.
- Pensaud, B.N. and Hurdle, V. F. (1991). "Some New Data That Challenge Some Old Ideas About Speed-Flow Relationships." *Transportation Research Record*. 1194: 191-8.
- Ross, Paul. (1991). "Some Properties of Macroscopic Traffic Models." *Transportation Research Record*. 1194: 129-34.

End Notes

[1] Note: We use k because the word is Konzentration in German

[2] <http://www.tfhrc.gov/its/tft/tft.htm>

Fundamentals of Transportation/ Traffic Flow/Problem


Problem:

Four vehicles are traveling at constant speeds between sections X and Y (28 meters apart) with their positions and speeds observed at an instant in time. An observer at point X observes the four vehicles passing point X during a period of 15 seconds. The speeds of the vehicles are measured as 88, 80, 90, and 72 km/hr respectively. Calculate the flow, density, time mean speed, and space mean speed of the vehicles.

- Solution

Fundamentals of Transportation/ Traffic Flow/Solution


Problem:

Four vehicles are traveling at constant speeds between sections X and Y (28 meters apart) with their positions and speeds observed at an instant in time. An observer at point X observes the four vehicles passing point X during a period of 15 seconds. The speeds of the vehicles are measured as 88, 80, 90, and 72 km/hr respectively. Calculate the flow, density, time mean speed, and space mean speed of the vehicles.


Solution:
Flow

$$q = N \left(\frac{3600}{t_{measured}} \right) = 4 \left(\frac{3600}{15} \right) = 960 \text{veh/hr}$$

Density

$$k = \frac{N}{L} = \frac{4}{28} = 142 \text{veh/km}$$

Time Mean Speed

$$\bar{v}_t = \frac{1}{N} \sum_{n=1}^N v_n = \frac{1}{4} (72 + 90 + 80 + 88) = 82.5 \text{km/hr}$$

Space Mean Speed

$$\bar{v}_s = \frac{N}{\sum_{n=1}^N \frac{1}{v_i}} = \frac{4}{\frac{1}{72} + \frac{1}{90} + \frac{1}{80} + \frac{1}{88}} = 81.86$$

$$t_i = L/v_i$$

$$t_A = L/v_A = .028/88 = 0.000318hr$$

$$t_B = L/v_B = .028/80 = 0.000350hr$$

$$t_C = L/v_C = .028/90 = 0.000311hr$$

$$t_D = L/v_D = .028/72 = 0.000389hr$$

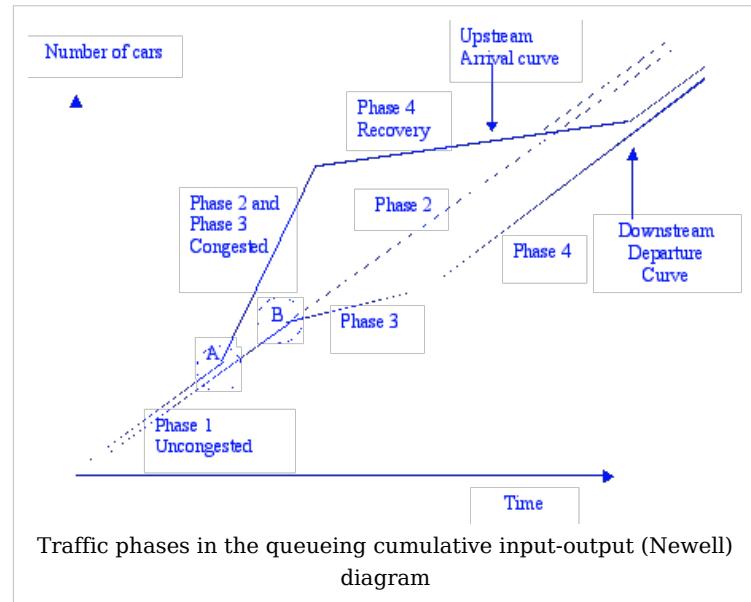
$$\bar{n}_s = \frac{NL}{\sum_{n=1}^N t_n} = \frac{4*0.028}{(0.000318+0.000350+0.000311+0.000389)} = 81.87km/hr$$

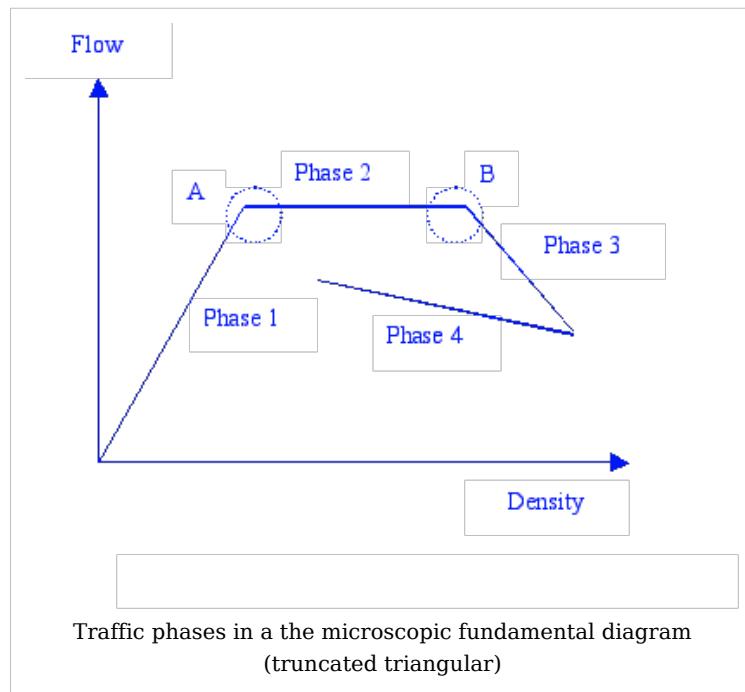
Fundamentals of Transportation/ Queueing and Traffic Flow

Queueing and Traffic Flow is the study of traffic flow behaviors surrounding queueing events. It combines elements from the previous two sections, Queueing and Traffic Flow.

Queueing and Traffic Flow

The first side figure illustrates a traffic bottleneck that drops the roadway from two lanes to one. It allows us to illustrate the changes in capacity, the changes in lane flow, and the stability in the total section flow. Over an extended period of time, by laws of conservation, flow through the bottleneck (q) must equal flow through the upstream section (Q).





$$(1) \sum_t q = \sum_t Q$$

A bottleneck causes lane flow (Q_l) to drop, but not section flow (Q), where:

$$(2) Q = \sum_t Q_l$$

When researchers observe a “backward-bending” flow-travel time curve, (which we will tackle later in the section) this is occasionally what they are seeing.

$$(3) \sum_t q = \sum_t Q \geq \sum_t Q_l$$

The lane flow drops upstream of a bottleneck when demand for a bottleneck exceeds the capacity of the bottleneck. Thus, the same level of flow can be observed at two different speeds on an upstream lane. The first is at freeflow speed in uncongested conditions and the second is at a lower speed when the downstream bottleneck is at capacity. In other words, congestion (a queue) forms when $Q > q_{max}$ for any period of time. Ultimately, what enters the queue must eventually exit; otherwise the queue will grow to infinity. This is illustrated with the queueing input-output (IO) diagram shown in the second side figure. We can identify four distinct phases in traffic.

- Phase 1 is the uncongested phase when there is no influence of the increasing density on the speeds of the vehicles. The speed does not drop with the introduction of newer vehicles onto the freeway.
- Phase 2 finds the freeway cannot sustain the speed with injection of newer vehicles into the traffic stream. The density increases while speed falls, maintaining the flow.
- Phase 3 shows decreased speed and decreased flows. This is caused by very low speeds cause the queue discharge to drop (slightly) at an active bottleneck, or a queue from a downstream bottleneck may be constraining the flow. As reported by many researchers (Banks, 1991; Hall and Agyemang-Duah, 1991; Persaud and Hurdle, 1991; Cassidy and Bertini, 1999), the reduced bottleneck capacity after breakdown ranges from 0% to 8%. A majority of papers conclude that the section capacity reduction a queue presents is

negligible. (As noted above, lane flow upstream of a bottleneck does drop, but this is due to the downstream bottleneck). This is a second source of the “backward-bending” phenomenon.

- Phase 4 is the recovery phase. During this phase, the density of traffic starts decreasing and speed starts increasing.

The circled region, ‘A’ (in the third side figure) is typically where we start observing ‘freeway breakdown’. In other words, this region occurs when upstream flow exceeds some critical downstream capacity at a specific point and there is a drop in speed. In the circled region, ‘B’, flow drops are associated with very low speeds. The phenomenon is particularly evidenced by the formation of queues upstream of where the breakdown occurs and a low discharge rate of vehicles due to sustained low speeds.

Traditional queues have “servers”. For example, the check-out line at the grocery store can serve one customer every 5 minutes, or one item every 10 seconds, etc. A conveyor belt can serve so many packages per hour. Capacity is often referred to as belonging to the road. However when we talk about road capacity, it is really a misnomer, as capacity is located in the driver, more precisely in the driver’s willingness and ability to drive behind the driver ahead. If drivers were willing and able to drive behind the vehicle ahead with no gap (spacing between vehicles), and that driver was driving behind the driver ahead of him with no gap, and so on, at high speeds, many more vehicles per hour could use the road. However, while some compression of vehicles occurs in heavy traffic, this situation is unstable because a driver will tap the brakes or even let-up on the accelerator for any number of reasons (to change lanes, to respond to someone else trying to change lanes, because he sees an object in the road, to limit the forces when rounding a corner, etc). This by definition lowers his speed, which in turn will lower the flow. Risk-averse drivers behind him will slow down even more (the driver has established some unpredictability in behavior, it is reasonable for other drivers to establish an even larger gap to accommodate the unpredictable driver’s behavior, especially given there is a reaction time between the lead driver’s actions, and the following driver’s perception (lead vehicle is slowing), decision (must brake), action (tap the brakes), the vehicle’s response to the action (tighten brakes on wheel)). In this way, maximum flow possible, our capacity (q_{max}), is a function of the drivers. The road shapes the driver’s willingness to take risks. Drivers will slow down around curves, vehicles may have difficulty accelerating uphill, or even from a slower speed (and even if they don’t, driver’s may provide insufficient fuel before they realize they are going too slow by not giving the vehicle enough gas), merges take time to avoid collision, etc.

Clearly, different drivers and different vehicles (for instance racecars or taxis) could increase the maximum flow through the bottleneck (q_{max}). It is better to think of capacity as a maximum sustainable flow (over an extended period of time), given typical drivers’ willingness to follow (subject to highway geometrics and environmental conditions) and their vehicles’ ability to respond to decisions. A series of aggressive drivers may exceed this ‘capacity’ for a short period of time, but eventually more cautious drivers will even out the function.

The IO diagram lets us understand delay in a way that the fundamental diagrams of traffic flow don’t easily allow. The first point to note about the IO diagram is that delay differs for each driver. The average delay can be measured easily (the total area in the triangle is the total delay, the average delay is just that triangle divided by the number of vehicles). The

variation (or standard deviation) can also be measured with some more statistics. As the total number of vehicles increases, the average delay increases.

The second point to note about the IO diagram is that the total number of queued vehicles (the length of the queue) can also be easily measured. This also changes continuously; the length of the queue rises and falls with changes in the arrival rate (the tail of the queue edge of the shockwave, where travel speed changes suddenly). Shockwaves indicate a change in state, or speed that suddenly occurs. One such shockwave is found where vehicles reach the back of queues. The arrival rate, coupled with the queue clearance rate, will tell you where the back of the queue is. This is interesting and is where the traveler first suffers delay - but it is not the source of the delay.

Only if the arrival rate exactly equals the departure rate would we expect to see a fixed length of the queue. If the queue results from a management practice (such as ramp metering), we can control the departure rate to match the arrival rate, and ensure that the queue remains on the ramp and doesn't spill over to neighboring arterials. However that observation suggests that congested "steady state" is likely to be a rare phenomenon, since in general the arrival rate does not equal the service rate.

Queueing analysts often make a simplifying assumption that vehicles stack vertically (queues take place at a point). This is of course wrong, but not too wrong. The resulting travel time is almost the same as if the queue were measured over space. The difference is that the time required to cover distance is included when we make the better assumption, even under freeflow conditions it takes time to travel from the point where a vehicle entered the back of the queue to the point where it exits the front. We can make that correction, but when queueing is taking place, that time is often small compared to the time delayed by the queue. We also assume a first-in, first-out logic, though again this can be relaxed without distracting from the main point. Another assumption we will make for exposition is that this is a deterministic process; vehicles arrive in a regular fashion and depart in a regular fashion. However sometimes vehicles bunch up (drivers are not uniform), which leads to stochastic arrivals and departures. This stochastic queueing can be introduced, and will in general increase the measured delay.

Hypercongestion

It has been observed that the same flow can be achieved on many links at two different speeds. Some call this the "backward-bending" phenomenon (Hau 1998, Crozet and Marlot 2001). The queueing analysis framework also has implications for "hypercongestion" or "backward-bending" flow-travel time curves, such as shown in the figure to the right. Recall we identified two sources for "backward-bending" speed-flow relationships. The first has to do with the point of observation. Observing the lane flow upstream of a bottleneck gives the impression of a backward bending relationship, but this disappears at the bottleneck itself. Under any given demand pattern, flow and speed are a unique pair. When demand is below the downstream active bottleneck's capacity, a flow on an upstream link can be achieved at high speed. When demand is above the downstream active bottleneck's capacity, the same flow on the upstream link can only be achieved at a low speed because of queueing. The second has to do with a capacity drop at the bottleneck itself under congested conditions. However, much research reports that this drop is slight to non-existent.

As in the bottleneck, we define lowercase q to be flow (vehicles per hour) departing the front of the bottleneck and uppercase Q to be flow arriving at the back of the bottleneck.

We also define k to be density (vehicles per kilometer), v to be speed (kilometers per hour), and s to be service rate (seconds per vehicle). The fundamental diagrams of traffic flow (q - k - v curves) represent a model of traffic flow as stylized in the traditional textbook representation of the fundamental diagram of traffic flow. The reasons why q should drop as k increases beyond a certain point at an isolated bottleneck are unclear. In other words, why should flow past a point drop just because the number of vehicles behind that point increases? Why should leading traffic be influenced by the behavior of following traffic?

If traffic behaves as a queue through a bottleneck (illustrated above), we should consider reasons why traffic flow departing the queue would not stay at its maximum.

One reason is if the vehicles in the queue could not travel fast enough so that the front of the following car could not reach the point of the front of the leading car in the time allotted the service rate. If a lane serves 1800 vehicles per hour, it serves 1 vehicle every 2 seconds, we say there is a 2 second service rate. If there were a 10 meter spacing between the vehicles (including the vehicle length plus a physical gap), this implies that the service rate would be worse than 1 vehicle every 2 seconds only if it took longer than 2 seconds to travel 10 meters (i.e. speed < 18 km /hour). This is very slow traffic and may properly be called hypercongestion. Why would traffic get that slow if freeflow speed is 120 km / hour? In general traffic departing the front of the queue won't be traveling that slowly, as the shockwave that reduces speed (the wave of red brake lights) has moved to the back of the queue.

A second reason the bottleneck flow might drop is if the departure flow is affected by external sources, namely a bottleneck further downstream spilling backwards. If the queue at a downstream bottleneck becomes sufficiently long, it will reduce the number of vehicles that can depart the upstream bottleneck. This is because the first bottleneck is no longer controlling, the downstream bottleneck is.

A third reason that bottleneck flow might drop is if the bottleneck is not being fully served (slots for cars are going unfilled). If vehicles are separated by large time headways, then the bottleneck might lose capacity. Thus, if vehicles choose large gaps, the bottleneck flow might drop. However, in congested situations, vehicles tend to follow more closely, not less closely.

For an isolated bottleneck, the departure flow remains (essentially) a constant and arrival flow varies. As more vehicles arrive at the back of the queue, the expected wait at the queue increases. All vehicles will eventually be served. In general, there is no practical constraint on the number of vehicles arriving at the back of the queue, but there is a maximum output flow in vehicles per unit time. Examining traffic upstream of the bottleneck is interesting, but does not get to the root of the problem - the bottleneck itself. This view of hypercongestion is thus not inconsistent with the conclusion drawn by Small and Chu (1997) that the hypercongested region is unsuitable for use as a supply curve in congestion pricing analyses.

Thought Question

Problem

Bottlenecks are shown here to be the main cause of traffic flow congestion and queueing. The examples given are of lane drops, which are not very common (some drivers think they're more common than they actually are). Is this the only way a bottleneck can form?

Solution

Absolutely not. A bottleneck is defined as a constriction of traffic flow, where demand exceeds available capacity. Infrastructure-wise, a bottleneck would be where a lane drop is present, where lane widths decrease, or a sight obstruction lowers the natural free-flow speed. However, bottlenecks can also form because of traffic. Chaotic traffic, primarily where lane changing is occurring, is a prime source of a bottleneck. On-ramps, off-ramps, and weaving areas are the most common examples, but there are many more. Think about how often you have been stuck in standstill traffic where a lane drop is not present.

Variables

- q - Flow through the bottleneck
- Q - Flow through the upstream section
- Q_l - Lane Flow
- t - time

Key Terms

- Bottleneck
- Queueing
- Traffic Flow
- Lane Flow
- Lane Flow Drops
- Flow-Travel Time Curve
- "Backwards-Bending"
- IO
- Hypercongestion

References

References on Queueing and Traffic Flow

- Banks, J.H. (1991). Two Capacity Phenomenon at Freeway Bottlenecks: A Basis for Ramp Metering? *Transportation Research Record* 1320, pp. 83-90.
- Banks, James H. (1992). "Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations." *Transportation Research Record* 1225:53-60.
- Cassidy, M.J. and R.L. Bertini (1999) Some Traffic Features at Freeway Bottlenecks. *Transportation Research Part B Vol. 33*, pp. 25-42
- Hall, F.L. and K. Agyemang-Duah (1991). Freeway Capacity Drop and the Definition of Capacity. *Transportation Research Record* 1320, pp. 91-98
- Persaud, B.N. and Hurdle, V. F. (1991). "Some New Data That Challenge Some Old Ideas About Speed-Flow Relationships." *Transportation Research Record*. 1194: 191-8.

- Persaud, B.N. and V.F. Hurdle (1991). Freeway Capacity: Definition and Measurement Issues. Proceedings of the International Symposium of Highway Capacity, A.A. Balkema press, Germany, pp.289-307

References on Hypercongestion

- Crozet, Yves Marlot, Gregoire (2001) "Congestion and Road Pricing: Where Is The 'Bug'?" 9th World Conference on Transport Research
- Hau, T.D. (1998) "Congestion pricing and road investment", in: K.J. Button and E.T. Verhoef, eds., *Road pricing, traffic congestion and the environment: Issues of efficiency and social feasibility*. Cheltenham: Edward Elgar.
- Small, Kenneth and Xuehao Chu. 1997. "Hypercongestion", Working Paper No. 96-97-11, Department of Economics, University of California, Irvine.

Fundamentals of Transportation/ Shockwaves

Shockwaves are byproducts of traffic congestion and queueing. They are transition zones between two traffic states that move through a traffic environment like, as their name states, a propagating wave. On the urban freeway, most drivers can identify them as a transition from a flowing, speedy state to a congested, standstill state. However, shockwaves are also present in the opposite case, where drivers who are idle in traffic suddenly are able to accelerate. Shockwaves are one of the major safety concerns for transportation agencies because the sudden change of conditions drivers experience as they pass through a shockwave often can cause accidents.

Visualization

While most people have probably experienced plenty of traffic congestion first hand, it is useful to see it systematically from three different perspectives: (1) That of the driver (with which most people are familiar), (2) a birdseye view, and (3) a helicopter view. Some excellent simulations are available here, please see the movies:

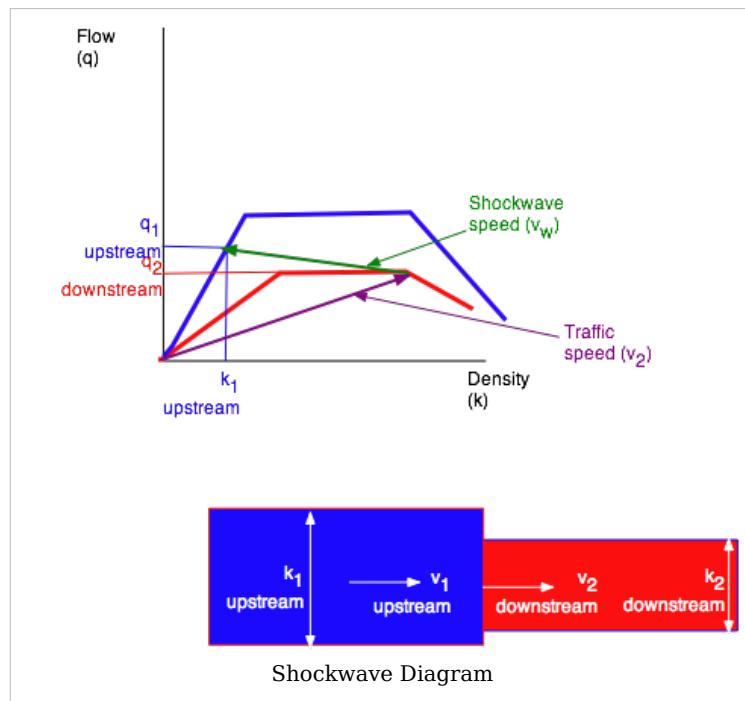
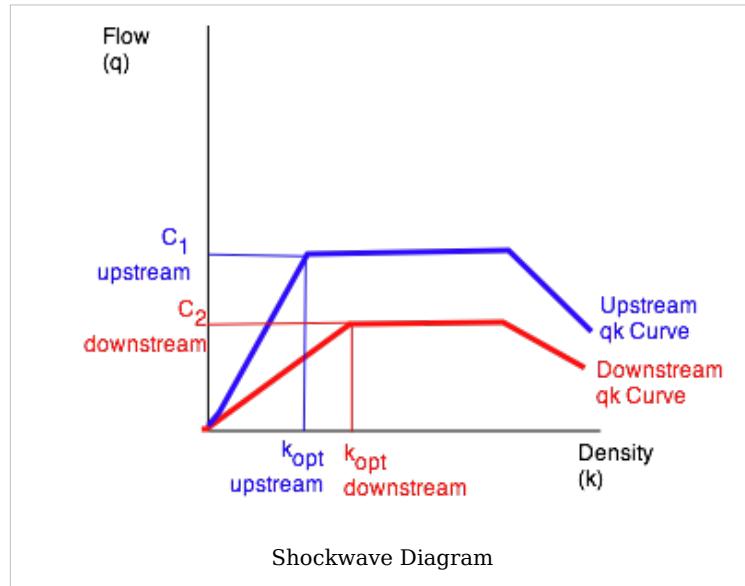
Visualization with the TU-Dresden 3D Traffic Simulator ^[1]

This movie shows traffic jams without an "obvious source" such as an on-ramp, but instead due to randomness in driver behavior: Shockwave traffic jams recreated for first time ^[2]

Analysis of shockwaves

Shockwaves can be seen by the cascading of brake lights upstream along a highway. They are often caused by a change in capacity on the roadways (a 4 lane road drops to 3), an incident, a traffic signal on an arterial, or a merge on freeway. As seen above, just heavy traffic flow alone (flow above capacity) can also induce shockwaves. In general, it must be remembered that capacity is a function of drivers rather than just being a property of the roadway and environment. As the capacity (maximum flow) drops from C_1 to C_2 , optimum density also changes. Speeds of the vehicles passing the bottleneck will of course be reduced, but the drop in speed will cascade upstream as following vehicles also have to decelerate.

The figures illustrate the issues. On the main road, far upstream of the bottleneck, traffic moves at density k_1 , below capacity (k_{opt}). At the bottleneck, density increases to accommodate the most of the flow, but speed drops.



Shockwave Math

Shockwave speed

If the flow rates in the two sections are q_1 and q_2 , then $q_1 = k_1 v_1$ and $q_2 = k_2 v_2$.

$$v_w = \frac{q_2 - q_1}{k_2 - k_1}$$

Relative speed

With v_1 equal to the space mean speed of vehicles in area 1, the speed relative to the line w is:

$$v_{r1} = v_1 - v_w$$

The speed of vehicles in area 2 relative to the line w is $v_{r2} = v_2 - v_w$

Boundary crossing

The number of vehicles crossing line 2 from area 1 during time period t is

$$N_1 = v_{r1}k_1t = (v_1 - v_w)k_1t$$

and similarly

$$N_2 = v_{r2}k_2t = (v_2 - v_w)k_2t$$

By conservation of flow, the number of vehicles crossing from left equals the number that crossed on the right

$$N_1 = N_2$$

so:

$$v_2k_2 - v_1k_1 = v_w(k_2 - k_1)$$

or

$$q_2 - q_1 = v_w(k_2 - k_1)$$

which is equivalent to

$$v_w = \frac{q_2 - q_1}{k_2 - k_1}$$

Examples

Example 1



- (A) What is the wave speed (v_w)?
- (B) What is the rate at which the queue grows, in units of vehicles per hour (q)?



Solution:

(A) At what rate does the queue increase?

1. Identify Unknowns:

$$k_1 = q/v_1 = 2000/80 = 25 \text{ veh/km}$$

$$v_2 = 0, q_2 = k_2v_2 = 0$$

2. Solve for wave speed (v_w)

$$v_w = \frac{q_2 - q_1}{k_2 - k_1} = \frac{0 - 2000}{275 - 25} = -8 \text{ km/hr}$$

Conclusion: the queue grows against traffic

(B) What is the rate at which the queue grows, in units of vehicles per hour?

$$\begin{aligned}
 N_1 &= (v_1 - v_w) k_1 t = (v_2 - v_w) k_2 t = N_2 \\
 \text{dropping } t \text{ (let } t = 1) \\
 v_1 k_1 - v_w k_1 &= v_2 k_2 - v_w k_2 \\
 q_1 - v_w k_1 &= q_2 - v_w k_2 \\
 2000 - (-8) * 25 &= 0 - (-8) * 275 \\
 2200 \text{veh/hr} &= 2200 \text{veh/hr}
 \end{aligned}$$

Thought Question

Problem

Shockwaves are generally something that transportation agencies would like to minimize on their respective corridor. Shockwaves are considered a safety concern, as the transition of conditions can often lead to accidents, sometimes serious ones. Generally, these transition zones are problems because of the inherent fallibility of human beings. That is, people are not always giving full attention to the road around them, as they get distracted by a colorful billboard, screaming kids in the backseat, or a flashy sports car in the adjacent lanes. If people were able to give full attention to the road, would these shockwaves still be causing accidents?

Solution

Yes, but not to the same extent. While accidents caused by driver inattentiveness would decrease nearly to zero, accidents would still be occurring between different vehicle types. For example, in a case where conditions change very dramatically, a small car (say, a Beetle) would be able to stop very quickly. A semi truck, however, is a much heavier vehicle and would require a longer distance to stop. If both were moving at the same speed when encountering the shockwave, the truck may not be able to stop in time before smashing into the vehicle ahead of them. That is why most trucks are seen creeping along through traffic with very big gaps ahead of them.

Sample Problem

- Problem (Solution)

Variables

- q - flow
- c - capacity (maximum flow)
- k - density
- v - speed
- v_r - relative speed (travel speed minus wave speed)
- v_w - wave speed
- N - number of vehicles crossing wave boundary

Key Terms

- Shockwaves
- Time lag, space lag

References

[1] <http://www.vwi.tu-dresden.de/~treiber/movie3d/index.html>

[2] <http://www.youtube.com/watch?v=Suugn-p5C1M>

Fundamentals of Transportation/ Shockwaves/Problem



Problem:

Flow on a road is $q_1 = 1800 \text{veh/hr/lane}$, and the density of $k_1 = 14.4 \text{veh/km/lane}$. To reduce speeding on a section of highway, a police cruiser decides to implement a rolling roadblock, and to travel in the left lane at the speed limit ($v_2 = 88 \text{km/hr}$) for 10 km. No one dares pass. After the police cruiser joins, the platoon density increases to 20 veh/km/lane and flow drops. How many vehicles (per lane) will be in the platoon when the police car leaves the highway?

- Solution

Fundamentals of Transportation/ Shockwaves/Solution



Problem:

Flow on a road is $q_1 = 1800 \text{veh/hr/lane}$, and the density of $k_1 = 14.4 \text{veh/km/lane}$. To reduce speeding on a section of highway, a police cruiser decides to implement a rolling roadblock, and to travel in the left lane at the speed limit ($v_2 = 88 \text{km/hr}$) for 10 km. No one dares pass. After the police cruiser joins, the platoon density increases to 20 veh/km/lane and flow drops. How many vehicles (per lane) will be in the platoon when the police car leaves the highway?



Solution:

Step 0

Solve for Unknowns:

Original speed

$$v_1 = \frac{q_1}{k_1} = \frac{1800}{14.4} = 125 \text{km/h}$$

Flow after police cruiser joins

$$q_2 = k_2 v_2 = 88 * 20 = 1760 \text{veh/h}$$

Step 1

Calculate the wave velocity:

$$v_w = \frac{q_2 - q_1}{k_2 - k_1} = \frac{1760 - 1800}{20 - 14.4} = -7.14 \text{km/hr}$$

Step 2

Determine the growth rate of the platoon (relative speed)

$$v_{r2} = v_2 - v_w = 88 - (-7.14) = 95.1 \text{km/hr}$$

Step 3

Determine the time spent by the police cruiser on the highway

$$t = d/v = 10 \text{km}/88 \text{km/hr} = 0.11 \text{hr} = 6.8 \text{minutes}$$

Step 4

Calculate the Length of platoon (not a standing queue)

$$L = vt = 95.1 \text{km/hr} * 0.11 \text{hr} = 10.46 \text{km}$$

Step 5

What is the rate at which the queue grows, in units of vehicles per hour?

$$\Delta q = q_1 - k_1 v_w = q_2 - k_2 v_w = 1800 - (14.4 * -7.14) = 1760 - (20 * -7.14) = 1902 \text{veh/hr}$$

Step 6

The number of vehicles in platoon

$$L_p k_2 = 10.46 \text{km} * 20 \text{veh/km} = 209.2 \text{vehicles}$$

OR

$$\Delta qt = 1902 \text{veh/hr} * 0.11 \text{hr} = 209.2 \text{veh}$$

Fundamentals of Transportation/ Traffic Signals

Traffic Signals are one of the more familiar types of intersection control. Using either a fixed or adaptive schedule, traffic signals allow certain parts of the intersection to move while forcing other parts to wait, delivering instructions to drivers through a set of colorful lights (generally, of the standard red-yellow-green format). Some purposes of traffic signals are to (1) improve overall safety, (2) decrease average travel time through an intersection, and (3) equalize the quality of services for all or most traffic streams. Traffic signals provide orderly movement of intersection traffic, have the abilities to be flexible for changes in traffic flow, and can assign priority treatment to certain movements or vehicles, such as emergency services. However, they may increase delay during the off-peak period and increase the probability of certain accidents, such as rear-end collisions. Additionally, when improperly configured, driver irritation can become an issue. Fortunately, traffic signals are generally a well-accepted form of traffic control for busy intersections and continue to be deployed.



A Traffic Light

Intersection Queueing

At an intersection where certain approaches are denied movement, queueing will inherently occur. Of the various queueing models, one of the more common and simple ones is the D/D/1 Queueing Model. This model assumes that arrivals and departures are deterministic and one departure channel exists. D/D/1 is quite intuitive and easily solvable. Using this form of queueing with an arrival rate λ and a departure rate μ , certain useful values regarding the consequences of queues can be computed.

One important piece of information is the duration of the queue for a given approach. This time value can be calculated through the following formula:

$$t_c = \frac{\rho r}{1 - \rho}$$

Where:

- t_c = Time for queue to clear
- ρ = Arrival Rate divided by Departure Rate

- r = Red Time

With this, various proportions dealing with queues can be calculated. The first determines the proportion of cycle with a queue.

$$P_q = \frac{r + t_c}{C}$$

Where:

- P_q = Proportion of cycle with a queue
- C = Cycle Length

Similarly, the proportion of stopped vehicles can be calculated.

$$P_s = \frac{\lambda(r + t_c)}{\lambda(r + g)} = \frac{r + t_c}{C} = P_q$$

$$P_s = \frac{\lambda(r + t_c)}{\lambda(r + g)} = \frac{\mu t_c}{\lambda C} = \frac{\mu t_c}{\rho C}$$

Where:

- P_s = Proportion of Stopped Vehicles
- g = Green Time

Therefore, the maximum number of vehicles in a queue can be found.

$$Q_{\max} = \lambda r$$

Intersection delay

Various models of intersection delay at isolated intersections have been put forward, combining queuing theory with empirical observations of various arrival rates and discharge times (Webster and Cobbe 1966; Hurdle 1985; Hagen and Courage 1992). Intersections on arterials are more complex phenomena, including factors such as signal progression and spillover of queues between adjacent intersections. Delay is broken into two parts: *uniform delay*, which is the delay that would occur if the arrival pattern were uniform, and *overflow delay*, caused by stochastic variations in the arrival patterns, which manifests itself when the arrival rate exceeds the service flow of the intersection for a time period.

Delay can be computed with knowledge of arrival rates, departure rates, and red times. Graphically, total delay is the product of all queues over the time period in which they are present.

$$D_t = \frac{\lambda r^2}{2(1 - \rho)}$$

Similarly, average vehicle delay per cycle can be computed.

$$d_{avg} = \frac{\lambda r^2}{2(1 - \rho)} \frac{1}{\lambda C}$$

$$d_{avg} = \frac{r^2}{2C(1 - \rho)}$$

From this, maximum delay for any vehicle can be found.

$$d_{\max} = r$$

Level of Service

In order to assess the performance of a signalized intersection, a qualitative assessment called Level of Service (LOS) is assessed, based upon quantitative performance measures. For LOS, the performance measured used is average control delay per vehicle. The general procedure for determining LOS is to calculate lane group capacities, calculate delay, and then make a determination.

Lane group capacities can be calculated through the following equation:

$$c = s \frac{g}{C}$$

Where:

- c = Lane Group Capacity
- s = Adjusted Saturation Flow Rate
- g = Effective Green Length
- C = Cycle Length

Average control delay per vehicle, thus, can be calculated by summing the types of delay mentioned earlier.

$$d = (d_1(PF)) + d_2 + d_3$$

Where:

- d = Average Signal Delay per vehicle (sec)
- d_1 = Average Delay per vehicle due to uniform arrivals (sec)
- PF = Progression Adjustment Factor
- d_2 = Average Delay per vehicle due to random arrivals (sec)
- d_3 = Average delay per vehicle due to initial queue at start of analysis time period (sec)

Uniform delay can be calculated through the following formula:

$$d_1 = \frac{0.5C \left(1 - \frac{g}{C}\right)^2}{1 - \left[\min(1, X) \frac{g}{C}\right]}$$

Where:

- X = Volume/Capacity (v/c) ratio for lane group.

Similarly, random delay can be calculated:

$$d_2 = 900T \left[(X - 1) + \sqrt{(X - 1)^2 + \frac{8kIX}{cT}} \right]$$

Where:

- T = Duration of Analysis Period (in hours)
- k = Delay Adjustment Factor that is dependent on signal controller mode
- I = Upstream filtering/metering adjustment factor

Overflow delay generally only applies to densely urban corridors, where queues can sometimes spill over into previous intersections. Since this is not very common (usually the consequence of a poorly timed intersection sequence, the rare increase of traffic demand, or an emergency vehicle passing through the area), it is generally not taken into account for simple problems.

Delay can be calculated for individual vehicles in certain approaches or lane groups. Average delay per vehicle for an approach A can be calculated using the following formula:

$$d_A = \frac{\sum_i d_i v_i}{\sum_i v_i}$$

Where:

- d_A = Average Delay per vehicle for approach A (sec)
- d_i = Average Delay per vehicle for lane group i on approach A (sec)
- v_i = Analysis flow rate for lane group i

Average delay per vehicle for the intersection can then be calculated:

$$d_I = \frac{\sum_A d_A v_A}{\sum_A v_A}$$

Where:

- d_I = Average Delay per vehicle for the intersection (sec)
- d_A = Average Delay per vehicle for approach A (sec)
- v_A = Analysis flow rate for approach A

Critical Lane Groups

For any combination of lane group movements, one lane group will dictate the necessary green time during a particular phase. This lane group is called the Critical Lane Group. This lane group has the highest traffic intensity (v/s) and the allocation of green time for each phase is based on this ratio.

The sum of the flow ratios for the critical lane groups can be used to calculate a suitable cycle length.

$$Y_c = \sum_{i=1}^n \left(\frac{v}{s} \right)_{ci}$$

Where:

- Y_c = Sum of Flow Ratios for Critical Lane Groups
- $(v/s)_{ci}$ = Flow Ratio for Critical Lane Group i
- n = Number of Critical Lane Groups

Similarly, the total lost time for the cycle is also an element that can be used in the calculation of cycle length.

$$L = \sum_{i=1}^n (t_L)_{ci}$$

Where:

- L = Total lost Time for Cycle
- $(t_L)_{ci}$ = Total Lost Time for Critical Lane Group i

Cycle Length Calculation

Cycle lengths are calculated by summing individual phase lengths. Using the previous formulas for assistance, the minimum cycle length necessary for the lane group volumes and phasing plan can be easily calculated.

$$C_{min} = \frac{L * X_c}{X_c - \sum_{i=1}^n Y_i}$$

Where:

- C_{min} = Minimum necessary cycle length
- X_c = Critical v/c ratio for the intersection
- $(v/s)_{ci}$ = Flow Ratio for Critical Lane Group
- n = Number of Critical Lane Groups

This equation calculates the minimum cycle length necessary for the intersection to operate at an acceptable level, but it does not necessarily minimize the average vehicle delay. A more optimal cycle length usually exists that would minimize average delay. Webster (1958) proposed an equation for the calculation of cycle length that seeks to minimize vehicle delay. This optimum cycle length formula is listed below.

$$C_{opt} = \frac{[(1.5L) + 5]}{\left(1.0 - \sum_{i=1}^n Y_i\right)}$$

Where:

- C_{opt} = Optimal Cycle Length for Minimizing Delay

Green Time Allocation

Once cycle length has been determined, the next step is to determine the allocation of green time to each phase. Several strategies for allocating green time exist. One of the more popular ones is to distribute green time such that v/c ratios are equalized over critical lane groups. Similarly, v/c ratios can be found with predetermined values for green time.

$$X_i = \frac{v_i}{c_i} = \frac{v_i}{s_i * g_i / C} = \frac{v_i / s_i}{g_i / C}$$

Where:

- X_i = v/c ratio for lane group i

With knowledge of cycle lengths, lost times, and v/s ratios, the degree of saturation for an intersection can be found.

$$X_c = \sum \frac{v_i}{s_i} \frac{C}{C - L}$$

Where:

- X_c = Degree of saturation for an intersection cycle

From this, the total effective green for all phases can be computed.

$$\sum g_i = \sum \frac{v_i}{s_i} \frac{C}{X_c} = C - L$$

Examples

Example 1: Intersection Queueing


Problem:

An approach at a pretimed signalized intersection has an arrival rate of 0.1 veh/sec and a saturation flow rate of 0.7 veh/sec. 20 seconds of effective green are given in a 60-second cycle. Provide analysis of the intersection assuming D/D/1 queueing.


Solution:

Traffic intensity, ρ , is the first value to calculate.

$$\rho = \frac{\lambda}{\mu} = \frac{0.1}{0.7} = 0.14$$

Red time is found to be 40 seconds ($C - g = 60 - 20$). The remaining values of interest can be easily found.

Time to queue clearance after the start of effective green:

$$t_c = \frac{\rho r}{1 - \rho} = \frac{0.14(40)}{1 - 0.14} = 6.51 \text{ s}$$

Proportion of the cycle with a queue:

$$P_q = \frac{r + t_c}{C} = \frac{40 + 6.51}{60} = 0.775$$

Proportion of vehicles stopped:

$$P_s = \frac{\lambda(r + t_c)}{\lambda(r + g)} = \frac{0.1(40 + 6.51)}{0.1(40 + 20)} = 0.775$$

Maximum number of vehicles in the queue:

$$Q_{\max} = \lambda r = 0.1(40) = 4$$

Total vehicle delay per cycle:

$$D_t = \frac{\lambda r^2}{2(1 - \rho)} = \frac{0.1(40)^2}{2(1 - 0.14)} = 93 \text{veh-s}$$

Average delay per vehicle:

$$d_{avg} = \frac{r^2}{2C(1 - \rho)} = \frac{(40)^2}{2(60)(1 - 0.14)} = 15.5 \text{ s}$$

Maximum delay of any vehicle:

$$d_{\max} = r = 40 \text{ s}$$

Example 2: Total Delay


Problem:

Compute the average approach delay given certain conditions for a 60-second cycle length intersection with 20 seconds of green, a v/c ratio of 0.7, a

progression neutral state ($PF=1.0$), and no chance of intersection spillover delay (overflow delay). Assume the traffic flow accounts for the peak 15-minute period and a lane capacity of 840 veh/hr.


Solution:

Uniform Delay:

$$d_1 = \frac{0.5C \left(1 - \frac{g}{C}\right)^2}{1 - \left[\min(1, X) \frac{g}{C}\right]} = \frac{0.5(60) \left(1 - \frac{20}{60}\right)^2}{1 - \left[\min(1, 0.7) \frac{20}{60}\right]} = 17.39 \text{ s}$$

Random Delay:

$T = 0.25$ (from problem statement)

$X = 0.7$

$k = 0.5$ (for pretimed control)

$I = 1.0$ (isolated intersection)

$c = 840$

$$d_2 = 900T \left[(X - 1) + \sqrt{(X - 1)^2 + \frac{8kIX}{cT}} \right] = 900(0.25) \left[(0.7 - 1) + \sqrt{(0.7 - 1)^2 + \frac{8(0.5)(1)(0.7)}{840(0.25)}} \right] = 4.83 \text{ s}$$

Overflow Delay:

Overflow delay is zero because it is assumed that there is no overflow.

$d_3 = 0$

Total Delay:

$$d = d_1(PF) + d_2 + d_3 = 17.39(1) + 4.83 + 0 = 22.22 \text{ s}$$

The average total delay is 22.22 seconds.

Example 3: Cycle Length Calculation


Problem:

Calculate the minimum and optimal cycle lengths for the intersection of Oak Street and Washington Avenue, given that the critical v/c ratio is 0.9, the two critical approaches have a v/s ratio of 0.3, and the Lost Time equals 15 seconds.


Solution:

Minimum Cycle Length:

$$C_{min} = \frac{L * X_c}{X_c - \sum_{i=1}^n Y_i} = \frac{15 * 0.9}{[0.9 - (2(0.3))]} = 45 \text{ s}$$

Optimal Cycle Length:

$$C_{opt} = \frac{[(1.5L) + 5]}{\left(1.0 - \sum_{i=1}^n Y_i\right)} = \frac{1.5(15) + 5}{1.0 - 2(0.3)} = 68.75 \text{ s}$$

The minimum cycle length is 45 seconds and the optimal cycle length is 68.75 seconds.

Thought Question

Problem

Why don't signalized intersections perform more efficiently than uncontrolled intersections?

Solution

The inherent lost time that comes from each signal change is wasted time that does not occur when intersections are uncontrolled. It comes at quite a surprise to most of the Western World, where traffic signals are plentiful, but there are intersections that perform quite well without any form of control. There is an infamous video on Youtube that shows an uncontrolled intersection in India where drivers somehow navigate through a busy, chaotic environment smoothly and efficiently. The video is humorous to watch, but it shows a valid point that uncontrolled intersections can indeed work and are quite efficient. However, the placement of traffic signals is for safety, as drivers entering an uncontrolled intersection have a higher likelihood of being involved in a dangerous accident, such as a T-bone or head-on collision.

Sample Problem

Problem (Solution)

Variables

- t_c - Time for queue to clear
- ρ - Arrival Rate divided by Departure Rate
- r - Red Time
- P_q - Proportion of cycle with a queue
- P_s - Proportion of Stopped Vehicles
- c - Lane Group Capacity
- s - Adjusted Saturation Flow Rate
- g - Effective Green Length
- C - Cycle Length
- d - Average Signal Delay per vehicle (sec)
- d_1 - Average Delay per vehicle due to uniform arrivals (sec)
- PF - Progression Adjustment Factor
- d_2 - Average Delay per vehicle due to random arrivals (sec)
- d_3 - Average delay per vehicle due to initial queue at start of analysis time period (sec)
- X - Volume/Capacity (v/c) ratio for lane group.
- T - Duration of Analysis Period (in hours)
- k - Delay Adjustment Factor that is dependent on signal controller mode
- I - Upstream filtering/metering adjustment factor
- d_A - Average Delay per vehicle for approach A (sec)

- d_i - Average Delay per vehicle for lane group i on approach A (sec)
- v_i - Analysis flow rate for lane group i
- d_I - Average Delay per vehicle for the intersection (sec)
- v_A - Analysis flow rate for approach A
- Y_c - Sum of Flow Ratios for Critical Lane Groups
- $(v/c)_{ci}$ - Flow Ratio for Critical Lane Group i
- n - Number of Critical Lane Groups
- C_{min} - Minimum necessary cycle length
- X_c - Critical v/c ratio for the intersection
- $(v/s)_{ci}$ - Flow Ratio for Critical Lane Group
- C_{opt} - Optimal Cycle Length for Minimizing Delay
- X_i - v/c ratio for lane group i
- X_c - Degree of saturation for an intersection cycle

Key Terms

- Delay
- Total Delay
- Average Delay
- Uniform Delay
- Random Delay
- Overflow Delay
- Cycle Length
- v/c ratio
- v/s ratio
- Saturation Flow Rate
- Red Time
- Effective Green
- Minimum Cycle Length
- Optimal Cycle Length
- Critical Lane Group
- Degree of Saturation
- Progression Adjustment Factor
- Lost Time
- Queue

External Exercises

Use the GAME software at the STREET website ^[1] to learn how to coordinate traffic signals to reduce delay.

Use the OASIS software at the STREET website ^[2] to study how signals change when given information about time-dependent vehicle arrivals.

References

- Hagen, Lawrence T. and Courage, Kenneth. (1992). "Comparison of Macroscopic Models for Signalized Intersection Analysis." *Transportation Research Record*. 1225: 33-44.
- Hurdle, V. F. (1982). "Signalized Intersections: A Primer for the Uninitiated." *Transportation Research Record* 971:96-105.
- Webster, F.V. (1958). Traffic Signal Settings. Road Research Technical Paper No. 39. London: Great Britain Road Research Laboratory.
- Webster, F.V. and Cobbe, B. M. (1966). Traffic Signals. Road Research Technical Paper No. 56. HMSO London UK.

References

[1] <http://street.umn.edu/>

[2] <http://street.umn.edu/>

Fundamentals of Transportation/ Traffic Signals/Problem



Problem:

An approach at a pretimed signalized intersection has an arrival rate of 500 veh/hr and a saturation flow rate of 3000 veh/hr. 30 seconds of effective green are given in a 100-second cycle. Provide analysis of the intersection assuming D/D/1 queueing by describing the proportion of the cycle with a queue, the maximum number of vehicles in the queue, the total and average delay, and the maximum delay.

- Solution

Fundamentals of Transportation/ Traffic Signals/Solution



Problem:

An approach at a pretimed signalized intersection has an arrival rate of 500 veh/hr and a saturation flow rate of 3000 veh/hr. 30 seconds of effective green are given in a 100-second cycle. Provide analysis of the intersection assuming D/D/1 queueing by describing the proportion of the cycle with a queue, the maximum number of vehicles in the queue, the total and average delay, and the maximum delay.



Solution:

With the statements in the problem, we know:

- Green Time = 30 seconds
- Red Time = 70 seconds
- Cycle Length = 100 seconds
- Arrival Rate = 500 veh/hr (0.138 veh/sec)
- Departure Rate = 3000 veh/hr (0.833 veh/sec)

Traffic intensity, ρ , is the first value to calculate.

$$\rho = \frac{\lambda}{\mu} = \frac{500}{3000} = 0.167$$

Time to queue clearance after the start of effective green:

$$t_c = \frac{\rho r}{1 - \rho} = \frac{0.167(70)}{1 - 0.167} = 14.03 \text{ s}$$

Proportion of the cycle with a queue:

$$P_q = \frac{r + t_c}{C} = \frac{70 + 14.03}{100} = 0.84$$

Proportion of vehicles stopped:

$$P_s = \frac{\lambda(r + t_c)}{\lambda(r + g)} = \frac{0.138(70 + 14.03)}{0.138(70 + 30)} = 0.84$$

Maximum number of vehicles in the queue:

$$Q_{\max} = \lambda r = 0.138(70) = 9.66$$

Total vehicle delay per cycle:

$$D_t = \frac{\lambda r^2}{2(1 - \rho)} = \frac{0.138(70^2)}{2(1 - 0.167)} = 406 \text{ veh-s}$$

Average delay per vehicle:

$$d_{avg} = \frac{r^2}{2C(1 - \rho)} = \frac{(70)^2}{2(100)(1 - 0.167)} = 29.41 \text{ s}$$

Maximum delay of any vehicle:

$$d_{\max} = r = 70 \text{ s}$$

Thus, the solution can be determined:

- Proportion of the cycle with a queue = 0.84
- Maximum number of vehicles in the queue = 9.66

- Total Delay = 406 veh-sec
- Average Delay = 29.41 sec
- Maximum Delay = 70 sec

Fundamentals of Transportation/ Design

In order to have a fully functional transportation system, the links that connect the various origins and destinations need to be designed to a level of quality that allows the safe and efficient movement of all vehicles that use them. This level of quality is reflective to the accurate installment of a geometric design. Such a design needs to take various elements into consideration, including number of lanes, lane width, median type, length of acceleration and deceleration lanes, curve radii, and many more. The detailed work of design has made lifelong careers for many engineers in the past. Today, despite advances in computer software, the basic fundamental understanding of building a highway still needs to be understood to guarantee that intuitive roads continue to be built in the future.

Highway design itself is made up of a spectrum of considerations. Some of the main points are discussed below. Details can be found in their respective sections.

Sight Distance

Sight Distance is the distance a driver can see from his or her vehicle. This becomes important when determining design speed, as it would be unsafe to allow a driver to drive faster and not be able to stop in time for a potential, unforeseen hazard. Sight distance is applied to two main categories:

- Stopping Sight Distance (SSD)
- Passing Sight Distance (PSD)

Grade

Grade is the slope, either upward or downward, of a road. This is especially critical because the terrain on which a road is built is seldom flat, thus requiring inherent "hills" to be present on the road to keep costs down from digging canyons or tunnels. However, too steep of a grade would make it difficult for vehicles to travel along that route. Since a road is built to provide a service to travelers, it is undesirable to have an impassable road.

An appropriate grade is dependent on the engine power in the vehicles that are expected to use the road. This breaks down into a force balance equation, where engine power is countered by various resistances. These include:

- Aerodynamic Resistance
- Rolling Resistance
- Grade Resistance

Horizontal Curves

Horizontal Curves are semicircular curves designed on the horizontal plane to allow roads to weave around obstacles, such as towns, mountains, or lakes. They allow a smooth transition to occur between two nonparallel roads instead of a sharp, pointed turn. When designing horizontal curves, designers must consider the intended design speed, as centripetal force is required to ensure that vehicles can successfully negotiate the curve. Such elements to consider for design include:

- Curve Radius
- Superelevation

Vertical Curves

Vertical Curves are placed to allow the road to follow the terrain, whether it be hills or valleys. Vertical curves are primarily designed as parabolas, using the general form of the parabolic equation with coefficients corresponding to known elements for the road in question. Designers can design a road by adjusting these elements and minimizing costs. These elements include:

- Inbound and Outbound Grades
- Curve Length
- Rate of curvature

Cross Sections

Roadway cross sections are important in design primarily for drainage. If roads were flat and level, water would congregate on the surface, reduce the coefficient of friction, and become a safety concern. Therefore, designers implement a "crown" cross section, which has a high elevation along the road's centerline and then tapers off as it leads to the shoulders. This slope is generally very small and unnoticeable by drivers, but succeeds in allowing the water to run into the ditches or storm sewers.

Earthwork

Earthwork plays into highway design as primarily a budgetary concern. The movement of earth, regardless if it is to a site, away from a site, or around a site, is an expensive venture. Designers often make roadway designs with a minimal amount of necessary earthwork to keep project costs down. However, knowing how to minimize dirt movement requires an understanding of the process.

Pavement Design

Pavement design generally is not considered part of geometric design, but is still an important part of the design process. In order to facilitate efficient traffic flow, the road beneath the vehicles' tires needs to be sturdy, stable, and smooth. Pavement engineers are responsible for determining appropriate pavement depths necessary to allow traffic to pass on a certain road. Inadequate engineering would result in cracking, formation of potholes, and total degradation of the roadway surface.

External Exercises

The calculation and design process of roadway geometry design are often cumbersome and time consuming. The University of Minnesota has an online roadway geometry design tool that was created to assist transportation students in this process. This tool, titled Roadway Online Application for Design (ROAD), allows students to design a roadway on a computer screen using a contour map in the background as reference. This tool makes the design process more efficient and effective over the traditional paper and pencil method. It can be found on the link listed below, is free to use, and covers the topics discussed in the Design section. It can be found at the STREET website^[1].

References

[1] http://street.umn.edu/ROAD_about.html

Fundamentals of Transportation/ Sight Distance

Sight Distance is a length of road surface which a particular driver can see with an acceptable level of clarity. Sight distance plays an important role in geometric highway design because it establishes an acceptable design speed, based on a driver's ability to visually identify and stop for a particular, unforeseen roadway hazard or pass a slower vehicle without being in conflict with opposing traffic. As velocities on a roadway are increased, the design must be catered to allowing additional viewing distances to allow for adequate time to stop. The two types of sight distance are (1) stopping sight distance and (2) passing sight distance.

Derivations

Stopping Sight Distance

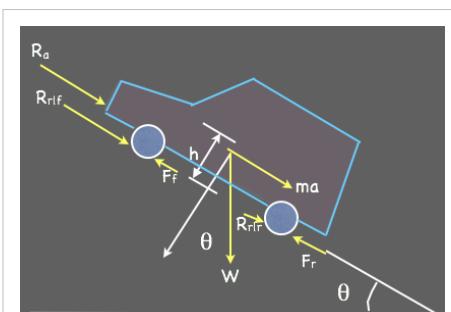
Stopping Sight Distance (SSD) is the viewable distance required for a driver to see so that he or she can make a complete stop in the event of an unforeseen hazard. SSD is made up of two components: (1) Braking Distance and (2) Perception-Reaction Time.

For highway design, analysis of braking is simplified by assuming that deceleration is caused by the resisting force of friction against skidding tires. This is applicable to both an uphill or a downhill situation. A vehicle can be modeled as an object with mass m sliding on a surface inclined at angle θ .

While the force of gravity pulls the vehicle down, the force of friction resists that movement. The forces acting this vehicle can be simplified to:

$$F = W (\sin(\theta) - f \cos(\theta))$$

where



Forces acting on a vehicle that is
braking

- $W = mg$ = object's weight,
- f = coefficient of friction.

Using Newton's second law we can conclude then that the acceleration (a) of the object is

$$a = g(\sin(\theta) - f \cos(\theta))$$

Using our basic equations to solve for braking distance (d_b) in terms of initial speed (v_i) and ending speed (v_e) gives

$$d_b = \frac{v_i^2 - v_e^2}{-2a}$$

and substituting for the acceleration yields

$$d_b = \frac{v_i^2 - v_e^2}{2g(f \cos(\theta) - \sin(\theta))}$$



Ample Stopping Sight Distance

$$d_b = \frac{v_i^2 - v_e^2}{2g(f - G)}$$

Using simply the braking formula assumes that a driver reacts instantaneously to a hazard. However, there is an inherent delay between the time a driver identifies a hazard and when he or she mentally determines an appropriate reaction. This amount of time is called perception-reaction time. For a vehicle in motion, this inherent delay translates to a distance covered in the meanwhile. This extra distance must be accounted for.

For a vehicle traveling at a constant rate, distance d_r covered by a specific velocity v and a certain perception-reaction time t_r can be computed using simple dynamics:

$$d_r = (vt_r)$$

Finally, combining these two elements together and incorporating unit conversion, the AASHTO stopping sight distance formula is produced. The unit conversions convert the problem to metric, with v_i in kilometers per hour and d_s in meters.

$$d_s = d_r + d_b = 0.278t_r v_i + \frac{(0.278v_i)^2}{19.6(f \pm G)}$$

Passing Sight Distance

Passing Sight Distance (PSD) is the minimum sight distance that is required on a highway, generally a two-lane, two-directional one, that will allow a driver to pass another vehicle without colliding with a vehicle in the opposing lane. This distance also allows the driver to abort the passing maneuver if desired. AASHTO defines PSD as having three main distance components: (1) Distance traveled during perception-reaction time and acceleration into the opposing lane, (2) Distance required to pass in the opposing lane, (3) Distance necessary to clear the slower vehicle.

The first distance component d_1 is defined as:

$$d_1 = 1000t_1 \left(u - m + \frac{at_1}{2} \right)$$

where

- t_1 = time for initial maneuver,
- a = acceleration (km/h/sec),
- u = average speed of passing vehicle (km/hr),
- m = difference in speeds of passing and impeded vehicles (km/hr).

The second distance component d_2 is defined as:

$$d_2 = (1000ut_2)$$

where

- t_2 = time passing vehicle is traveling in opposing lane,
- u = average speed of passing vehicle (km/hr).

The third distance component d_3 is more of a rule of thumb than a calculation. Lengths to complete this maneuver vary between 30 and 90 meters.

With these values, the total passing sight distance (PSD) can be calculated by simply taking the summation of all three distances.

$$d_p = (d_1 + d_2 + d_3)$$

Examples

Example 1: Stopping Distance



Problem:

A vehicle initially traveling at 88 km/h skids to a stop on a 3% downgrade, where the pavement surface provides a coefficient of friction equal to 0.3. How far does the vehicle travel before coming to a stop?



Solution:

$$d_b = \frac{\left(88 * \left(\frac{1000}{3600}\right)\right)^2 - (0)^2}{2 * (9.8) * (0.3 - 0.03)} = 112.9m$$

Example 2: Coefficient of Friction



Problem:

A vehicle initially traveling at 150 km/hr skids to a stop on a 3% downgrade, taking 200 m to do so. What is the coefficient of friction on this surface?



Solution:

$$d_b = \frac{\left(150 * \left(\frac{1000}{3600}\right)\right)^2 - (0)^2}{2 * (9.8) * (f - 0.03)} = 200m$$

$$(f - 0.03) = \frac{\left(150 * \left(\frac{1000}{3600}\right)\right)^2 - (0)^2}{2 * (9.8) * 200}$$

$$f = 0.47$$

Example 3: Grade



Problem:

What should the grade be for the previous example if the coefficient of friction is 0.40?



Solution:

$$d_b = \frac{\left(150 * \left(\frac{1000}{3600}\right)\right)^2 - (0)^2}{2 * (9.8) * (0.40 - G)} = 200m$$

$$(0.40 - G) = \frac{\left(150 * \left(\frac{1000}{3600}\right)\right)^2 - (0)^2}{2 * (9.8) * 200}$$

$$G = 0.44 - 0.40 = 0.04$$

Thus the road needs to be a 4 percent uphill grade if the vehicles are going that speed on that surface and can stop that quickly.

Example 4: Accident Reconstruction



Problem:

You are shown an accident scene with a vehicle and a light pole. The vehicle was estimated to hit the light pole at 50 km/hr. The skid marks are measured to be 210, 205, 190, and 195 meters. A trial run that is conducted to help measure

the coefficient of friction reveals that a car traveling at 60 km/hr can stop in 100 meters under conditions present at the time of the accident. How fast was the vehicle traveling to

begin with?



Solution:

First, Average the Skid Marks.

$$(210 + 205 + 190 + 195)/4 = 200$$

Estimate the coefficient of friction.

$$d_b = \frac{\left(60 * \left(\frac{1000}{3600}\right)\right)^2 - (0)^2}{2 * (9.8) * (f - 0)} = 100m$$

$$f = \frac{\left(60 * \left(\frac{1000}{3600}\right)\right)^2 - (0)^2}{2 * (9.8) * 100} = 0.14$$

Third, estimate the unknown velocity

$$d_b = \frac{\left(v * \left(\frac{1000}{3600}\right)\right)^2 - \left(50 * \left(\frac{1000}{3600}\right)\right)^2}{2 * (9.8) * (0.14 - 0)} = 200m$$

$$\left(v * \left(\frac{1000}{3600}\right)\right)^2 - \left(50 * \left(\frac{1000}{3600}\right)\right)^2 = 200m * (2 * (9.8) * (0.14))$$

$$548.8 + 192.9 = v^2 \left(\frac{1000}{3600}\right)^2$$

$$v^2 = \frac{741.7}{0.077} = 9612.43$$

$$v = 98km/h$$

Example 5: Compute Stopping Sight Distance



Problem:

Determine the Stopping Sight Distance from Example 4, assuming an AASHTO recommended perception-reaction time of 2.5 seconds.



Solution:

$$d_s = ((1000/3600) * 98 * 2.5) + (98 * 0.278)^2 / (2 * 9.8 * 0.14) = 338$$

Thought Question

Problem

If the coefficient of friction is 0 (zero) and the grade is 0, how long does it take a moving vehicle to stop?

Solution

Forever

Note, the design conditions for roads are wet, i.e. a lower coefficient of friction

Sample Problem

Problem (Solution)

Variables

- d_s - stopping (sight) distance (m)
- d_r - perception reaction distance (m)
- d_b - braking distance (m)
- d_p - passing distance (m)
- v_i - initial speed (km/h)
- t_r - perception/reaction time (seconds)
- f - AASHTO stopping friction coefficient (dimensionless)
- G - roadway grade (dimensionless)

Key Terms

- SSD : Stopping Sight Distance
- PSD : Passing Sight Distance
- PRT : Perception-Reaction Time

Standards and Practices

AASHTO Recommended Friction Coefficients

| (km/hr) | Coefficient of Skidding Friction (f) |
|---------|--------------------------------------|
| 30 | 0.40 |
| 40 | 0.38 |
| 50 | 0.35 |
| 60 | 0.33 |
| 70 | 0.31 |
| 80 | 0.30 |
| 90 | 0.30 |
| 100 | 0.29 |
| 110 | 0.28 |
| 120 | 0.28 |

Fundamentals of Transportation/ Sight Distance/Problem

**Problem:**

You see a body lying across the road and need to stop. If your vehicle was initially traveling at 100 km/h and skids to a stop on a 2.5% upgrade, taking 75 m to do so, what was the coefficient of friction on this surface?

- Solution

Fundamentals of Transportation/ Sight Distance/Solution

**Problem:**

You see a body lying across the road and need to stop. If your vehicle was initially traveling at 100 km/h and skids to a stop on a 2.5% upgrade, taking 75 m to do so, what was the coefficient of friction on this surface?

Solution:

$$d_b = \frac{\left(100 * \left(\frac{1000}{3600}\right)\right)^2 - (0)^2}{2 * (9.8) * (f + 0.025)} = 75m$$

$$(f + 0.025) = \frac{(27.78)^2}{2 * (9.8) * 75}$$

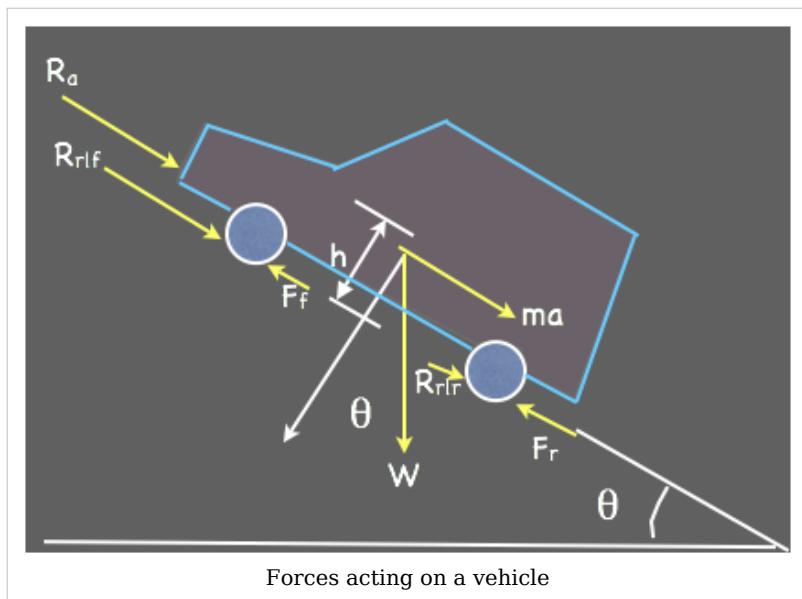
$$f = 0.50$$

Fundamentals of Transportation/Grade

Road Vehicle Performance forms the basics that make up highway design guidelines and traffic analysis. Roads need to be designed so that the vehicles traveling upon them are accommodated. For example, the Interstate Highway system in the United States has a maximum allowable grade that can be used, so that the semitrucks that frequently use these roads would be able to travel them without encountering grade problems. While this can pose quite a design challenge in mountainous regions like Colorado Rockies, without it those trucks would be forced to find better alternative routes, which costs time and money. The ability of a certain vehicle type to use a road is dependent on the power produced by its motor, as well as other road-based and environmental characteristics. When choosing a road design, these elements need to be considered.

Fundamental Characteristics

Tractive effort and resistance are the two main forces that oppose one another and determine the performance of roadway vehicles. Tractive effort is the force exerted against the roadway surface to allow a vehicle to move forward. Resistance encompasses all forces that push back and impede motion. Both of these are in units of force. The general formula for this is outlined below:



$$F_t = ma + R_a + R_{rl} + R_g$$

Where:

- F_t = Tractive Effort
- m = Vehicle Mass
- a = Acceleration
- R_a = Aerodynamic Resistance
- R_{rl} = Rolling Resistance
- R_g = Grade Resistance

These components are discussed in greater detail in the following sections.

Aerodynamic Resistance

Aerodynamic resistance is a force that is produced by turbulent air flow around the vehicle body. This turbulence is dependent on the shape of the vehicle, as well as the friction of air passing over the vehicle's surface. A small portion of this resistance comes from air flow through vehicle components, such as interior ventilation. This resistance can be estimated through the following formula:

$$R_a = \frac{\rho}{2} A C_D V^2$$

Where:

- ρ = Air Density
- A = Frontal area of the vehicle
- C_D = Coefficient of Drag
- V = Speed of the vehicle

Air density is a function of elevation and temperature. Frontal area and coefficient of drag are generally unique to each vehicle or type of vehicle.

Rolling Resistance

Rolling resistance is caused by a vehicle's internal mechanical friction and the interaction of tires with the roadway surface. Three main causes exist that create this resistance. The first is the rigidity of the tire and the roadway surface. The second is tire pressure and temperature. The third is vehicular operating speed. This value of rolling friction can be calculated from a very simplified formula, given here in metric. V is in meters per second.

$$f_{rl} = 0.01(1 + \frac{V}{44.73})$$

The resistance caused by this friction will increase as weight is added to the vehicle. Therefore, rolling resistance can be calculated.

$$R_{rl} = f_{rl} W$$

Where:

- W = Vehicle Weight
- f_{rl} = Rolling Friction

Grade Resistance

Grade resistance is the simplest form of resistance. It is the gravitational force acting on the vehicle. This force may not be exactly perpendicular to the roadway surface, especially in situations when a grade is present. Thus, grade resistance can be calculated in the following formula:

$$R_g = WG$$

Where:

- W = Vehicle Weight
- G = Grade (length/length)

Tractive Effort

Tractive effort is the force that allows the vehicle to move forward, subject to the resistances of the previous three forces. The derivation of the formula comes from understanding the forces and moments that act around the various tires. It can be summarized into a simple concept, illustrated here.

For a rear-wheel drive car:

$$F_{max} = \frac{\mu W(l_f - f_{rl}h)/L}{1 + \mu h/L}$$

For a front-wheel drive car:

$$F_{max} = \frac{\mu W(l_r + f_{rl}h)/L}{1 + \mu h/L}$$

Where:

- F_{max} = Maximum Tractive Effort
- μ = Coefficient of road adhesion
- W = Vehicle Weight
- l_r = Distance from rear axle to vehicle's center of gravity
- l_f = Distance from front axle to vehicle's center of gravity
- f_{rl} = Coefficient of rolling friction
- h = Height of the center of gravity above the roadway surface
- L = Length of wheelbase

Grade Computation

Most of the work surrounding tractive effort is geared toward determining the allowable grade of a given roadway. With a certain known vehicle type using this road, the grade can be easily calculated. Using the force balance equation for tractive effort, a value for grade can be separating, producing the formula below:

$$G = \frac{F_t - F_a - F_{rl}}{W}$$

This calculation produces the maximum grade allowed for a given vehicle type. It assumes that the vehicle is operating at optimal engine capacity and, thus, no acceleration can occur, dropping that element from the overall equation.

Examples

Example 1: Racecar Acceleration



Problem:

A racecar is speeding down a level straightaway at 100 km/hr. The car has a coefficient of drag of 0.3, a frontal area of 1.5 m^2 , a weight of 10 kN, a wheelbase of 3 meters, and a center of gravity 0.5 meters above the roadway surface, which is 1 meter behind the front axle. The air density is 1.054 kg/m^3 and the coefficient of road adhesion is 0.6. What is the rate of acceleration for the vehicle?

**Solution:**

Use the force balancing equation to solve for a .

$$F_t = ma + R_a + R_{rl} + R_g$$

Since the straightaway is a level one, the grade is zero, thus removing grade resistance from the general problem.

$$R_g = 0$$

Aerodynamic resistance is computed:

$$R_a = \frac{\rho}{2} A C_D V^2 = \frac{1.054}{2} (1.5)(0.3)(100 * 1000/3600)^2 = 183 \text{ N}$$

Rolling resistance is computed:

$$f_{rl} = 0.01 \left(1 + \frac{V}{44.73}\right) = 0.01 \left(1 + \frac{100 * 1000/3600}{44.73}\right) = 0.016$$

$$R_{rl} = f_{rl} W = 0.016(10000) = 160 \text{ N}$$

Tractive Effort is computed:

$$F_{max} = \frac{\mu W(l_f - f_{rl}h)/L}{1 + \mu h/L} = \frac{0.6(10000)(1 - 0.016(0.5))/3}{1 + 0.6(0.5)/3} = 1804 \text{ N}$$

Looking back to the force balancing equation:

$$ma = F_t - R_a - R_{rl} - R_g = 1804 - 183 - 160 - 0 = 1461 \text{ N}$$

Divide out mass, which can be computed from weight by dividing out gravity.

$$W = mg$$

$$m = \frac{W}{g} = \frac{10000}{9.8} = 1020 \text{ kg}$$

Thus, divide mass from the force and acceleration can be found.

$$a = \frac{F}{m} = \frac{1461}{1020} = 1.43 \text{ m/s}^2$$

Thus, the vehicle is accelerating at a rate of 1.43 meters per second squared.

Example 2: Going Up a Hill

**Problem:**

Using the same case from Example 1, assume that instead the racecar encounters a steep hill that it must travel up. It is desired that the driver maintain the 100 km/hr velocity at a very minimum. With that being said, what would be the maximum grade that the hill could be?

**Solution:**

At the steepest eligible hill, the racecar would be able to maintain 100 km/hr without any room for acceleration or deceleration. Therefore, acceleration goes to zero. All other values would stay the same from Example 1. Using the grade formula, the maximum grade can be calculated.

$$G = \frac{F_t - F_a - F_{rl}}{W} = \frac{1804 - 183 - 160}{10000} = 0.1461$$

The maximum allowable grade is 14.61%.

Thought Question

Problem

Why is it that, in mountainous country, trucks and cars have different speed limits?

Answer

Tractive effort is one of the leading reasons, as trucks have a harder time going up steep hills than typical passenger cars, but it is not the only one. Safety is another leading reason, surprisingly, as big rig trucks are obviously more difficult to control in a harsh environment, such as a mountain pass.

Sample Problem

Problem (Solution)

Variables

- F_t - Tractive Effort Force
- F_a - Aerodynamic resistance Force
- F_{rl} - Rolling Resistance Force
- F_g - Grade Resistance Force
- W - Vehicle Weight
- m - Vehicle Mass
- a - Acceleration
- ρ - Air Density
- A - Frontal area of the vehicle
- C_D - Coefficient of Drag
- V - Speed of the vehicle
- f_{rl} - Rolling Friction
- G - Grade
- μ - Coefficient of road adhesion
- l_r - Distance from rear axle to vehicle's center of gravity
- l_f - Distance from front axle to vehicle's center of gravity
- h - Height of the center of gravity above the roadway surface
- L - Length of wheelbase

Key Terms

- Tractive Effort
- Aerodynamic Resistance
- Rolling Resistance
- Grade Resistance

Fundamentals of Transportation/ Grade/Problem



Problem:

It has been estimated that a Tour-de-France champion could generate a sustained 510 Watts of power while a healthy young human male (HYHM) can generate about 310 Watts of power. The bicycling champion and HYHM are going to race (on bicycles) up a hill with a 6% upgrade, that is five miles long, and the elevation at the top of the hill is 5000 feet. Both rider/bicycle combinations weigh 170 lbs, with frontal area 0.4m^2 and coefficient of drag 0.9 (values being typical of bicyclists in crouched racing positions). The coefficient of rolling resistance for both bicycles is 0.01. Assume ρ is $1.0567 \text{ kg/cubic-m}$. Remember, power equals the product of force and velocity. (1) Who gets to the top first? (2) How much longer does it take the loser to make it to the top?

- Solution

Fundamentals of Transportation/ Grade/Solution


Problem:

It has been estimated that a Tour-de-France champion could generate a sustained 510 Watts of power while a healthy young human male (HYHM) can generate about 310 Watts of power. The bicycling champion and HYHM are going to race (on bicycles) up a hill with a 6% upgrade, that is five miles long, and the elevation at the top of the hill is 5000 feet. Both rider/bicycle combinations weigh 170 lbs, with frontal area 0.4 m^2 and coefficient of drag 0.9 (values being typical of bicyclists in crouched racing positions). The coefficient of rolling resistance for both bicycles is 0.01. Assume ρ is 1.0567 kg/cubic-m. Remember, power equals the product of force and velocity. (1) Who gets to the top first? (2) How much longer does it take the loser to make it to the top?


Solution:

The winner is obvious the Tour-de-France champion, as everything between the two is equal with the exception of power. Since the champion produces more power, he, by default, would win.

We need to find the maximum steady state speed for each racer to compute the differences in arrival time. At the steady state speed we must have:

$$F_t = R_a + R_{rl} + R_g$$

Where:

- F_t = Tractive Effort
- m = Vehicle Mass
- a = Acceleration
- R_a = Aerodynamic Resistance
- R_{rl} = Rolling Resistance
- R_g = Grade Resistance

We also know:

- $\rho = 1.0567 \text{ kg/cubic-m}$
- $W = 756 \text{ N}$
- $f_{rl} = 0.01$
- $G = 0.06$
- $A = 0.4 \text{ m}^2$
- $C_D = 0.9$

To estimate available tractive effort we can use the definition of power as time rate of work, $P = FV$, to get:

$$F = \frac{P}{v}$$

Substituting in the components of the individual formulas to the general one, we get:

$$\frac{P}{v} = \frac{\rho}{2} A C_D V^2 + f_{rl} W + W G$$

If we compute rolling and grade resistance, we get:

$$R_{rl} = f_{rl}W = 0.01(756) = 7.56 \text{ N}$$

$$R_g = WG = 756(0.06) = 45.36 \text{ N}$$

Together, they add up to 52.92 N.

Aerodynamic Resistance can be found to be:

$$R_a = \frac{1.0567}{2}(0.4)(0.9)v^2 = 0.19v^2$$

With everything substituted into the general formula, the end result is the following formulation:

$$v = \frac{P}{[0.19v^2 + 52.92]}$$

This problem can be solved iteratively (setting a default value for v and then computing through iterations) or graphically. Either way, when plugging in 510 watts for the Tour-de-France champion, the resulting velocity is 7.88 meters/second. Similarly, when plugging in 310 watts for the HYHM, the resulting velocity is 5.32 meters/second.

The hill is five miles in length, which translates to 8.123 kilometers, 8,123 meters. It will take the champion 1030 seconds (or 17.1 minutes) to complete this link, whereas the HYHM will take 1527 seconds (or 25.4 minutes). The resulting difference is 8.3 minutes.

Fundamentals of Transportation/ Earthwork

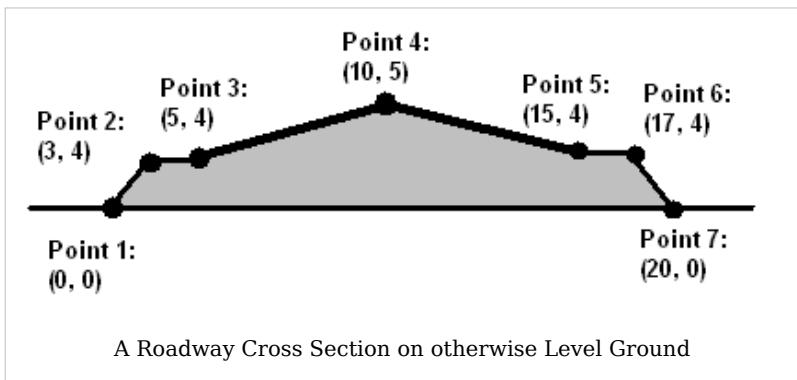
Earthwork is something that transportation projects seldom avoid. In order to establish a properly functional road, the terrain must often be adjusted. In many situations, geometric design will often involve minimizing the cost of earthwork movement. Earthwork is expressed in units of volumes (cubic meters in metric). Increases in such volumes require additional trucks (or more runs of the same truck), which cost money. Thus, it is important for designers to engineer roads that require very little earthwork.



The equipment necessary for earthwork to occur

Cross Sections and Volume Computation

To determine the amount of earthwork to occur on a given site, one must calculate the volume. For linear facilities, which include highways, railways, runways, etc., volumes can easily be calculated by integrating the areas of the cross sections (slices that go perpendicular to the centerline) for the entire length of the corridor. More simply, several cross sections can be selected along the corridor and an average can be taken for the entire length. Several different procedures exist for calculating areas of earthwork cross sections. In the past, the popular method was to draw cross sections by hand and use a planimeter to measure area. In modern times, computers use a coordinate method to assess earthwork calculations. To perform this task, points with known elevations need to be identified around the cross section. These points are considered in the (X, Y) coordinate plane, where X represents the horizontal axis paralleling the ground and Y represents the vertical axis that is elevation. Area can be computed with the following formula:



$A = \left| \frac{1}{2} \sum_{i=1}^n X_i(Y_{i+1} - Y_{i-1}) \right|$

Where:

- A = Area of Cross-Section
- n = Number of Points on Cross Section (Note: $n+1 = 1$ and $1-1=n$, for indexing)
- X = X-Coordinate
- Y = Y-Coordinate

With this, earthwork volumes can be calculated. The easiest means to do so would be using the average end area method, where the two end areas are averaged over the entire length between them.

$$V = \frac{A_1 + A_2}{2} L$$

Where:

- V = Volume
- A_1 = Cross section area of first side
- A_2 = Cross section area of second side
- L = Length between the two areas

If one end area has a value of zero, the earthwork volume can be considered a pyramid and the correct formula would be:

$$V = \frac{AL}{3}$$

A more accurate formula would be the prismoidal formula, which takes out most of the error accrued by the average end area method.

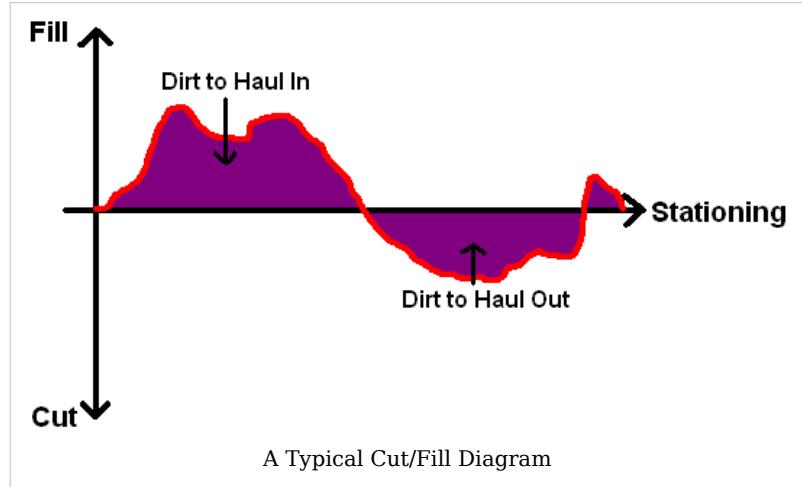
$$V_p = \frac{L(A_1 + 4A_m + A_2)}{6}$$

Where:

- V_p = Volume given by the prismoidal formula
- A_m = Area of a plane surface midway between the two cross sections

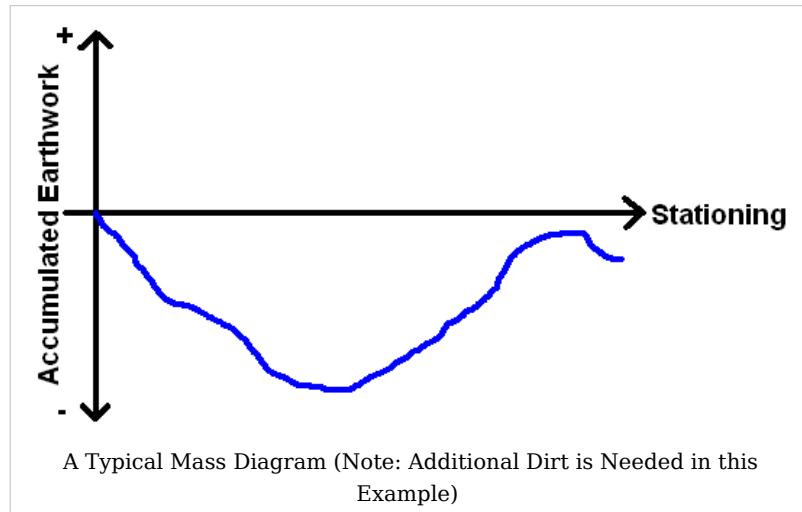
Cut and Fill

Various sections of a roadway design will require bringing in earth. Other sections will require earth to be removed. Earth that is brought in is considered **Fill** while earth that is removed is considered **Cut**. Generally, designers generate drawings called Cut and Fill Diagrams, which illustrate the cut or fill present at any given site. This drawing is quite standard, being no more than a graph with site location on the X-axis and fill being the positive range of the Y-axis while cut is the negative range of the Y-axis.



Mass Balance

Using the data for cut and fill, an overall mass balance can be computed. The mass balance represents the total amount of leftover (if positive) or needed (if negative) earth at a given site based on the design up until that point. It is a useful piece of information because it can identify how much remaining or needed earth will be present at the completion of a project, thus allowing designers to calculate how much expense will be incurred to haul out excess dirt or haul in needed additional. Additionally, a mass balance diagram, represented graphically, can aid designers in moving dirt internally to save money.



Similar to the cut and fill diagram, the mass balance diagram is illustrated on two axes. The X-axis represents site location along the roadway corridor and the Y-axis represents the amount of earth, either in excess (positive) or needed (negative).

Examples

Example 1: Computing Volume

**Problem:**

A roadway is to be designed on a level terrain. This roadway is 150 meters in length. Four cross sections have been selected, one at 0 meters, one at 50 meters, one at 100 meters, and one at 150 meters. The cross sections, respectively, have areas of 40 square meters, 42 square meters, 19 square meters, and 34 square meters. What is the volume of earthwork needed along this road?

**Solution:**

Three sections exist between all of these cross sections. Since none of the sections end with an area of zero, the average end area method can be used.

The volumes can be computed for respective sections and then summed together.

Section between 0 and 50 meters:

$$V = \frac{A_1 + A_2}{2} L = \frac{40 + 42}{2} 50 = 2050 \text{ cubic - meters}$$

Section between 50 and 100 meters:

$$V = \frac{A_1 + A_2}{2} L = \frac{42 + 19}{2} 50 = 1525 \text{ cubic - meters}$$

Section between 100 and 150 meters:

$$V = \frac{A_1 + A_2}{2} L = \frac{19 + 34}{2} 50 = 1325 \text{ cubic - meters}$$

Total Volume is found to be:

$$2050 + 1525 + 1325 = 4900 \text{ cubic - meters}$$

Example 2: Mass Balance

**Problem:**

Given the following cut/fill profile for each meter along a 10-meter strip of road built on very, very hilly terrain, estimate the amount of dirt left over or needed for the project.

- 0 Meters: 3 meters of fill
- 1 Meter: 1 meter of fill
- 2 Meters: 2 meters of cut
- 3 Meters: 5 meters of cut
- 4 Meters: 7 meters of cut
- 5 Meters: 8 meters of cut
- 6 Meters: 2 meters of cut
- 7 Meters: 1 meter of fill
- 8 Meters: 3 meters of fill
- 9 Meters: 6 meters of fill
- 10 Meters: 7 meters of fill



Solution:

If 'cut' is considered an excess of available earth and 'fill' is considered a reduction of available earth, the problem becomes one of simple addition and subtraction.

$$[(-3)+(-1)+2+5+7+8+2+(-1)+(-3)+(-6)+(-7)]*1m^2 = 3 \text{ cubic-meters}$$

3 cubic-meters of dirt remain in excess.

Thought Question

Problem

If it is found that the mass balance is indeed balanced (end value of zero), does that automatically mean that no dirt transport, either out of or into the site, is needed?

Solution

No. Any soil scientist will eagerly state that dirt type can change with location quite quickly, depending on the region. So, if half a highway cuts from the earth and the other half needs fill, the dirt pulled from the first half cannot be simply dumped into the second half, even if mathematically it balances. If the soil types are different, the exact numbers of volume needed may be different, as different soil types have different properties (settling, water storage, etc.). In the worse case, not consulting a soil scientist could result in your road being washed out!

Sample Problem

Problem (Solution)

Variables

- A - Area of Cross-Section
- n - Number of Points on Cross Section (Note: $n+1 = 1$ and $1-1=n$, for indexing)
- X - X-Coordinate
- Y - Y-Coordinate
- V - Volume
- A_1 - Cross section area of first side
- A_2 - Cross section area of second side
- L - Length between the two areas
- V_p - Volume given by the prismoidal formula
- A_m - Area of a plane surface midway between the two cross sections

Key Terms

- Cut
- Fill
- Mass Balance
- Area
- Volume
- Earthwork
- Prismoidal Volume

Fundamentals of Transportation/ Earthwork/Problem



Problem:

Given the end areas below, calculate the volumes of cut (in cubic meters) and fill between stations 0+00 and 2+50. Determine the true amount of excess cut or fill to be removed.

- 0+00: Fill = 60
- 0+50: Fill = 50
- 0+75: Cut = 0, Fill = 25
- 1+00: Cut = 10, Fill = 5
- 1+15: Cut = 15, Fill = 0
- 1+50: Cut = 30
- Solution

Fundamentals of Transportation/ Earthwork/Solution


Problem:

Given the end areas below, calculate the volumes of cut (in cubic meters) and fill between stations 0+00 and 2+50. Determine the true amount of excess cut or fill to be removed.

- 0+00: Fill = 60
- 0+50: Fill = 50
- 0+75: Cut = 0, Fill = 25
- 1+00: Cut = 10, Fill = 5
- 1+15: Cut = 15, Fill = 0
- 1+50: Cut = 30


Solution:

Two different methods need to be used here to compute earthwork volumes along the five strips. The average end area method can be used for non-zero sections. The pyramid method needs to be used for areas with zero ends.

For 0+00 to 0+50, use average end area:

$$\text{Fill} = \frac{60 + 50}{2}(50) = 2750$$

For 0+50 to 0+75, use average end area:

$$\text{Fill} = \frac{50 + 25}{2}(25) = 937.5$$

For 0+75 to 1+00, use the average end area method for the fill section and the pyramid method for the cut section:

$$\text{Fill} = \frac{25 + 5}{2}(25) = 375$$

$$\text{Cut} = \frac{10(25)}{3} = 83.3$$

For 1+00 to 1+15, use the pyramid method for the fill section and the average end area method for the cut section:

$$\text{Fill} = \frac{5(15)}{3} = 25$$

$$\text{Cut} = \frac{10 + 15}{2}(15) = 187.5$$

For 1+15 to 1+50, use the average end area method:

$$\text{Cut} = \frac{15 + 30}{2}(35) = 787.5$$

The sums of both cut and fill can be found:

- Fill = 4087.5 cubic-meters
- Cut = 1058.3 cubic-meters

Thus, 3029.2 cubic-meters of dirt are needed to meet the earthwork requirement for this project.

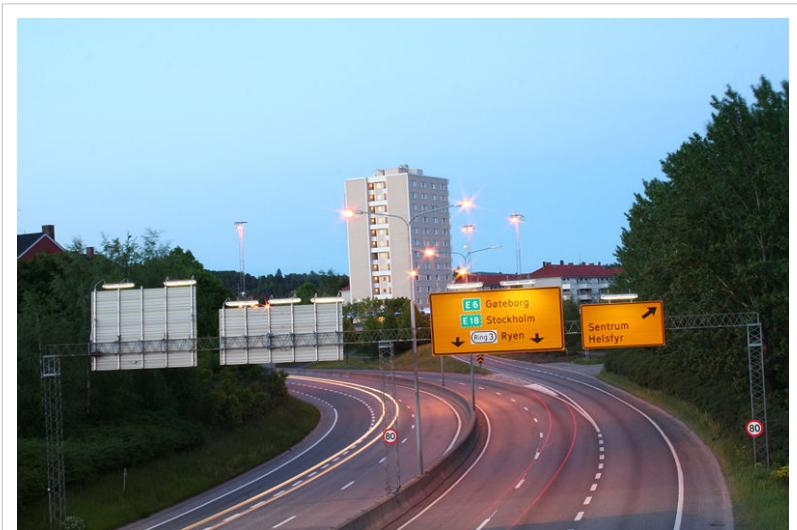
Fundamentals of Transportation/ Horizontal Curves

Horizontal Curves are one of the two important transition elements in geometric design for highways. A horizontal curve provides a transition between two tangent strips of roadway, allowing a vehicle to negotiate a turn at a gradual rate rather than a sharp cut. The design of the curve is dependent on the intended design speed for the roadway, as well as other factors including drainage and friction. These curves are semicircles as to provide the driver with a constant turning rate with radii determined by the laws of physics surrounding centripetal force.

Fundamental Horizontal Curve Properties

Physics Properties

When a vehicle makes a turn, two forces are acting upon it. The first is gravity, which pulls the vehicle toward the ground. The second is centripetal force, which is an external force required to keep the vehicle on a curved path. For any given velocity, the centripetal force would need to be greater for a tighter turn (one with a smaller radius) than a broader one (one with a larger radius). On a level surface, side friction f_s could serve as a countering force, but it generally would provide very little resistance. Thus, the vehicle would have to make a very wide circle in order to negotiate a turn. Given that road designs usually encounter very narrow design areas, such wide turns are generally discouraged.



A Horizontal Curve in Oslo, Norway

To deal with this issue, horizontal curves have roads that are tilted at a slight angle. This tilt is defined as superelevation, or ϵ , which is the amount of rise seen on an angled cross-section of a road given a certain run, otherwise known as slope. The presence of superelevation on a curve allows some of the centripetal force to be countered by the ground, thus allowing the turn to be executed at a faster rate than would be allowed on a flat surface. Superelevation also plays another important role by aiding in drainage during precipitation events, as water runs off the road rather than collecting on it. Generally, superelevation is limited to being less than 14 percent, as engineers need to account for stopped vehicles on the curve, where centripetal force is not present.

The allowable radius R for a horizontal curve can then be determined by knowing the intended design velocity V , the coefficient of friction, and the allowed superelevation on

the curve.

$$R = \frac{v^2}{g(e + f_s)}$$

With this radius, practitioners can determine the degree of curve to see if it falls within acceptable standards. Degree of curve, D_a , can be computed through the following formula, which is given in Metric.

$$\%R = \frac{1746}{D_a}\%$$

Where:

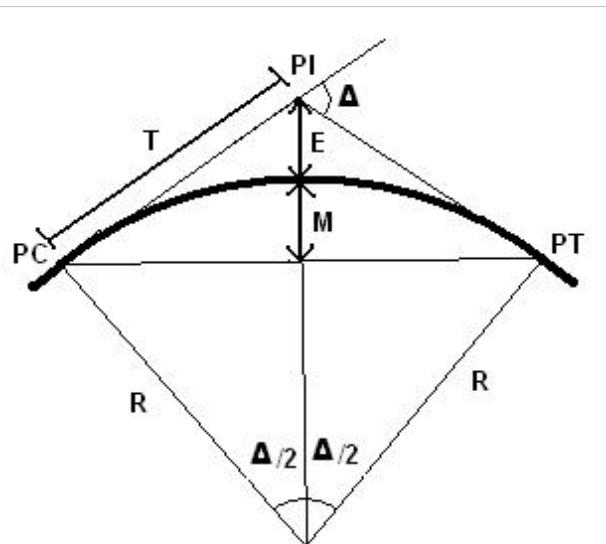
- D_a = Degree of curve [angle subtended by a 30.5-m (100 ft) arc along the horizontal curve]

Geometric Properties

Horizontal curves occur at locations where two roadways intersect, providing a gradual transition between the two. The intersection point of the two roads is defined as the **Point of Tangent**

Intersection (PI). The location of the curve's start point is defined as the **Point of Curve (PC)** while the location of the curve's end point is defined as the **Point of Tangent (PT)**.

Both the PC and PT are a distance T from the PI, where T is defined as Tangent Length. Tangent Length can be calculated by finding the central angle of the curve, in degrees. This angle is equal to the supplement of the interior angle between the two road tangents.



A Typical Horizontal Curve (Plan View)

$$T = R \tan\left(\frac{\Delta}{2}\right)$$

Where:

- T = tangent length (in length units)
- Δ = central angle of the curve, in degrees
- R = curve radius (in length units)

The distance between the PI and the vertex of the curve can be easily calculated by using the property of right triangles with T and R . Taking this distance and subtracting off the curve radius R , the external distance E , which is the smallest distance between the curve and PI, can be found.

$$E = R \left(\frac{1}{\cos\left(\frac{\Delta}{2}\right)} - 1 \right)$$

Where:

- E = external distance (in length units)

Similarly, the middle ordinate M can be found. The middle ordinate is the maximum distance between a line drawn between PC and PT and the curve. It falls along the line between the curve's vertex and the PI.

$$M = R \left(1 - \cos \left(\frac{\Delta}{2} \right) \right)$$

Where:

- M = middle ordinate (in length units)

Curve length L can be determined using the formula for semicircle length.

$$L = \frac{R\Delta\pi}{180}$$

Similarly, the geometric formula for cord length can find C , which represents the cord length for this curve.

$$C = 2R \sin \left(\frac{\Delta}{2} \right)$$

Sight Distance Properties

Unlike straight, level roads that would have a clear line of sight for a great distance, horizontal curves pose a unique challenge. Natural terrain within the inside of the curve, such as trees, cliffs, or buildings, can potentially block a driver's view of the upcoming road if placed too close to the road. As a result, the acceptable design speed is often reduced to account for sight distance restrictions.

Two scenarios exist when computing the acceptable sight distance for a given curve. The first is where the sight distance is determined to be less than the curve length. The second is where the sight distance exceeds the curve length. Each scenario has a respective formula that produces sight distance based on geometric properties. Determining which scenario is the correct one often requires testing both to find out which is true.

Given a certain sight distance S and a known curve length L and inner lane centerline radius R_v , the distance a sight obstruction can be from the interior edge of the road, M_s can be computed in the following formulas.

$$S < L : M_s = R_v \left(1 - \cos \frac{28.65S}{R_v} \right)$$

$$S > L : M_s = R_v \left(1 - \cos \frac{28.65L}{R_v} \right) + \left(\frac{S - L}{2} \right) \sin \left(\frac{28.65L}{R_v} \right)$$



Limited Curve Sight Distance Ahead

Examples

Example 1: Curve Radius


Problem:

A curving roadway has a design speed of 110 km/hr. At one horizontal curve, the superelevation has been set at 6.0% and the coefficient of side friction is found to be 0.10. Determine the minimum radius of the curve that will provide safe vehicle operation.


Solution:

$$R = \frac{v^2}{g(e + f_s)} = \frac{110^2}{9.8(.06 + 0.10)} = 595 \text{ meters}$$

Example 2: Determining Stationing


Problem:

A horizontal curve is designed with a 600 m radius and is known to have a tangent length of 52 m. The PI is at station 200+00. Determine the stationing of the PT.


Solution:

$$T = R \tan\left(\frac{\Delta}{2}\right)$$

$$52 = 600 \tan\left(\frac{\Delta}{2}\right)$$

$$\Delta = 9.9 \text{ deg}$$

$$L = \frac{R\pi\Delta}{180} = \frac{600\pi9.9}{180} = 104$$

$$PC = PI - T = 200 + 00 - 0 + 52 = 199 + 48$$

$$PT = PC + L = 199 + 48 + 1 + 04 = 200 + 52$$

Example 3: Stopping Sight Distance


Problem:

A very long horizontal curve on a one-directional racetrack has 1750-meter centerline radius, two 4-meter lanes, and a 200 km/hr design speed. Determine the closest distance from the inside edge of the track that spectators can park without impeding the necessary sight distance of the drivers. Assume that the sight distance is less than the length of the curve, a coefficient of friction of 0.3, and a perception-reaction time of 2.5 seconds.

**Solution:**

With a centerline radius of 1750 meters, the centerline of the interior lane is 1748 meters from the vertex ($1750 - (4/2)$). Using the stopping sight distance formula (See Sight Distance), SSD is computed to be 664 meters. With this, the distance from the track that spectators can be parked can easily be found.

$$S < L : M_s = R_v \left(1 - \cos \frac{28.65S}{R_v} \right) = 1748 \left(1 - \cos \frac{28.65(664)}{1748} \right) = 31.43 \text{ meters}$$

Thought Question**Problem**

Horizontal curves are semicircles, as to allow the driver to negotiate the curve without having to change the tilt of the wheel while passing through. However, what happens at the PC or PT, where tangent road transitions to a curve or vice versa. Is the driver expected to abruptly turn the wheel to match the curve or tangent?

Solution

Generally not. Engineers employ several techniques to allow drivers to gradually turn the wheel as they enter or leave the curve. One of the more popular ways is to use Splines. Splines are polynomial curves that have changing rates of curvature. Employed at the PC and PT, Splines take away any sharp changes for the drivers and make the drive a more pleasant one.

Sample Problem

Problem (Solution)

Variables

- R - Centerline Curve Radius
- D_a - Degree of curve [angle subtended by a 30.5-m (100 ft) arc along the horizontal curve]
- T - tangent length (in length units)
- Δ - central angle of the curve, in degrees
- E - Smallest distance between the curve and PI
- M - Middle ordinate
- L - Curve Length
- C - Cord Length
- S - Sight Distance
- M_s - Acceptable distance from inner edge of road for a sight obstruction to be placed without impeding sight distance
- R_v - Radius of innermost lane centerline

Key Terms

- PC: Point of Curve
- PI: Point of Tangent Intersect
- PT: Point of Tangent

Fundamentals of Transportation/ Horizontal Curves/Problem



Problem:

A given curve was very poorly designed. The two-lane road used has a lower-than-average coefficient of friction (0.05), no superelevation to speak of, and 4-meter lanes. 900 kg vehicles tend to go around this curve and are stylistically top heavy. County engineers have warned that this curve cannot be traversed as safely as other curves in the area, but politicians want to keep the speed up to boost tourism in the area. The curves have a radius of 500 feet and a design speed of 80 km/hr. Because the vehicles using the curve are top heavy, they have a tendency to roll over if too much side force is exerted on them (the local kids often race around the curve at night to get the thrill of "two-wheeling"). As an engineer, you need to prove that this curve is infeasible before an accident occurs. How can you show this?

- Solution

Fundamentals of Transportation/ Horizontal Curves/Solution



Problem:

A given curve was very poorly designed. The two-lane road used has a lower-than-average coefficient of friction (0.05), no superelevation to speak of, and 4-meter lanes. 900 kg vehicles tend to go around this curve and are stylistically top heavy. County engineers have warned that this curve cannot be traversed as safely as other curves in the area, but politicians want to keep the speed up to boost tourism in the area. The curves have a radius of 500 feet and a design speed of 80 km/hr. Because the vehicles using the curve are top heavy, they have a tendency to roll over if too much side force is exerted on them (the local kids often race around the curve at night to get the thrill of "two-wheeling"). As an engineer, you need to prove that this curve is infeasible before an accident occurs. How can you show this?



Solution:

The stated speed is 80 km/hr. The easiest way would be to prove that this is too high. We will look at the innermost lane, since forces will be greater there. Using the general curve radius formula and solving for v, we find:

$$R = \frac{v^2}{g(e + f_s)} = \frac{v^2}{9.8(0.05 + 0)} = 500 - (4/2)$$

$$v = 15.62 \text{ m/s} = 56.23 \text{ km/hr}$$

80 km/hr is much greater than 56.23 km/hr, which by default means that more force is being exerted on the vehicle than the road can counter. Thus, the curve's speed limit is dangerous and needs to be changed.

Fundamentals of Transportation/ Vertical Curves

Vertical Curves are the second of the two important transition elements in geometric design for highways, the first being horizontal curves. A vertical curve provides a transition between two sloped roadways, allowing a vehicle to negotiate the elevation rate change at a gradual rate rather than a sharp cut. The design of the curve is dependent on the intended design speed for the roadway, as well as other factors including drainage, slope, acceptable rate of change, and friction. These curves are parabolic and are assigned stationing based on a horizontal axis.

Fundamental Curve Properties

Parabolic Formulation

Two types of vertical curves exist: (1) Sag Curves and (2) Crest Curves. Sag curves are used where the change in grade is positive, such as valleys, while crest curves are used when the change in grade is negative, such as hills. Both types of curves have three defined points: PVC (Point of Vertical Curve), PVI (Point of Vertical Intersection), and PVT (Point of Vertical Tangency). PVC is the start point of the curve while the PVT is the end point. The elevation at either of these points can be computed as e_{PVC} and e_{PVT} for PVC and PVT respectively. The roadway grade that approaches the PVC is defined as g_1 and the roadway grade that leaves the PVT is defined as g_2 . These



A Road Through Hilly Terrain with Vertical Curves in New Hampshire

grades are generally described as being in units of (m/m) or (ft/ft), depending on unit type chosen.

Both types of curves are in parabolic form. Parabolic functions have been found suitable for this case because they provide a constant rate of change of slope and imply equal curve tangents, which will be discussed shortly. The general form of the parabolic equation is defined below, where y is the elevation for the parabola.

$$y = ax^2 + bx + c$$

At $x = 0$, which refers to the position along the curve that corresponds to the PVC, the elevation equals the elevation of the PVC. Thus, the value of c equals e_{PVC} . Similarly, the slope of the curve at $x = 0$ equals the incoming slope at the PVC, or g_1 . Thus, the value of b equals g_1 . When looking at the second derivative, which equals the rate of slope change, a value for a can be determined.

$$a = \frac{g_2 - g_1}{2L}$$

Thus, the parabolic formula for a vertical curve can be illustrated.

$$y = e_{PVC} + g_1x + \frac{(g_2 - g_1)x^2}{2L}$$

Where:

- e_{PVC} : elevation of the PVC
- g_1 : Initial Roadway Grade (m/m)
- g_2 : Final Roadway Grade (m/m)
- L : Length of Curve (m)

Most vertical curves are designed to be Equal Tangent Curves. For an Equal Tangent Curve, the horizontal length between the PVC and PVI equals the horizontal length between the PVI and the PVT. These curves are generally easier to design.

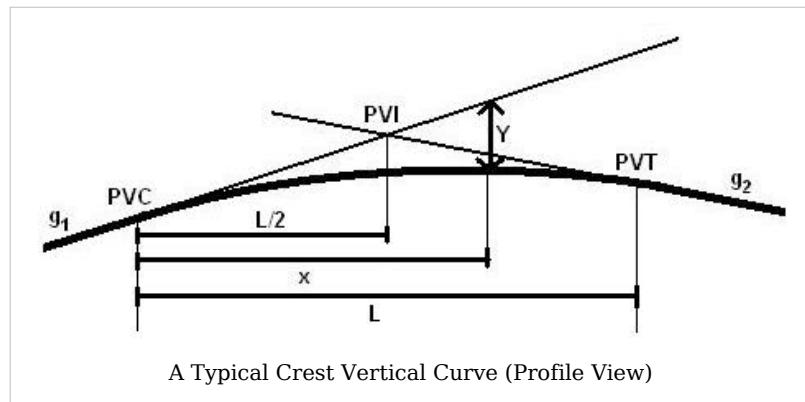
Offset

Some additional properties of vertical curves exist. Offsets, which are vertical distances from the initial tangent to the curve, play a significant role in vertical curve design. The formula for determining offset is listed below.

$$Y = \frac{Ax^2}{200L}$$

Where:

- A : The absolute difference between g_2 and g_1 , multiplied by 100 to translate to a percentage
- L : Curve Length
- x : Horizontal distance from PVC along curve



Stopping Sight Distance

Sight distance is dependent on the type of curve used and the design speed. For crest curves, sight distance is limited by the curve itself, as the curve is the obstruction. For sag curves, sight distance is generally only limited by headlight range. AASHTO has several tables for sag and crest curves that recommend rates of curvature, K , given a design speed or stopping sight distance. These rates of curvature can then be multiplied by the absolute slope change percentage, A to find the recommended curve length, L_m .

$$L_m = KA$$

Without the aid of tables, curve length can still be calculated. Formulas have been derived to determine the minimum curve length for required sight distance for an equal tangent curve, depending on whether the curve is a sag or a crest. Sight distance can be computed from formulas in other sections (See Sight Distance).

Crest Vertical Curves

The correct equation is dependent on the design speed. If the sight distance is found to be less than the curve length, the first formula below is used, whereas the second is used for sight distances that are greater than the curve length. Generally, this requires computation of both to see which is true if curve length cannot be estimated beforehand.

$$S < L : L_m = \frac{AS^2}{200(\sqrt{h_1} + \sqrt{h_2})^2}$$

$$S > L : L_m = 2S - \frac{200(\sqrt{h_1} + \sqrt{h_2})^2}{A}$$

Where:

- L_m : Minimum Curve Length (m)
- A : The absolute difference between g_2 and g_1 , multiplied by 100 to translate to a percentage
- S : Sight Distance (m)
- h_1 : Height of driver's eye above roadway surface (m)
- h_2 : Height of objective above roadway surface (m)

Sag Vertical Curves

Just like with crest curves, the correct equation is dependent on the design speed. If the sight distance is found to be less than the curve length, the first formula below is used, whereas the second is used for sight distances that are greater than the curve length. Generally, this requires computation of both to see which is true if curve length cannot be estimated beforehand.

$$S < L : L_m = \frac{AS^2}{200(H + S \tan \beta)}$$

$$S > L : L_m = 2S - \frac{200(H + S \tan \beta)}{A}$$

Where:

- A : The absolute difference between g_2 and g_1 , multiplied by 100 to translate to a percentage

- S : Sight Distance (m)
- H : Height of headlight (m)
- β : Inclined angle of headlight beam, in degrees

Passing Sight Distance

In addition to stopping sight distance, there may be instances where passing may be allowed on vertical curves. For sag curves, this is not an issue, as even at night, a vehicle in the opposing can be seen from quite a distance (with the aid of the vehicle's headlights). For crest curves, however, it is still necessary to take into account. Like with the stopping sight distance, two formulas are available to answer the minimum length question, depending on whether the passing sight distance is greater than or less than the curve length. These formulas use units that are in metric.

$$S < L : L_m = \frac{A(PSD^2)}{864}$$

$$S > L : L_m = 2PSD - \frac{864}{A}$$

Where:

- A : The absolute difference between g_2 and g_1 , multiplied by 100 to translate to a percentage
- PSD : Passing Sight Distance (m)
- L_m : Minimum curve length (m)

Examples

Example 1: Basic Curve Information



Problem:

A 500-meter equal-tangent sag vertical curve has the PVC at station 100+00 with an elevation of 1000 m. The initial grade is -4% and the final grade is +2%. Determine the stationing and elevation of the PVI, the PVT, and the lowest point on the curve.



Solution:

The curve length is stated to be 500 meters. Therefore, the PVT is at station 105+00 ($100+00 + 5+00$) and the PVI is in the very middle at 102+50, since it is an equal tangent curve. For the parabolic formulation, c equals the elevation at the PVC, which is stated as 1000 m. The value of b equals the initial grade, which in decimal is -0.04. The value of a can then be found as 0.00006.

Using the general parabolic formula, the elevation of the PVT can be found:

$$y = 0.00006x^2 + -0.04x + 1000 = 0.00006(500)^2 + -0.04(500) + 1000 = 995 \text{ m}$$

Since the PVI is the intersect of the two tangents, the slope of either tangent and the elevation of the PVC or PVT, depending, can be used as reference. The elevation of the PVI can then be found:

$$y = -0.04x + 1000 = -0.04(250) + 1000 = 990 \text{ m}$$

To find the lowest part of the curve, the first derivative of the parabolic formula can be found. The lowest point has a slope of zero, and thus the low point location can be found:

$$dy/dx = 0.00012x + -0.04 = 0.00012x + -0.04 = 0$$

$$x = 333.33 \text{ m}$$

Using the parabolic formula, the elevation can be computed for that location. It turns out to be at an elevation of 993.33 m, which is the lowest point along the curve.

Example 2: Adjustment for Obstacles

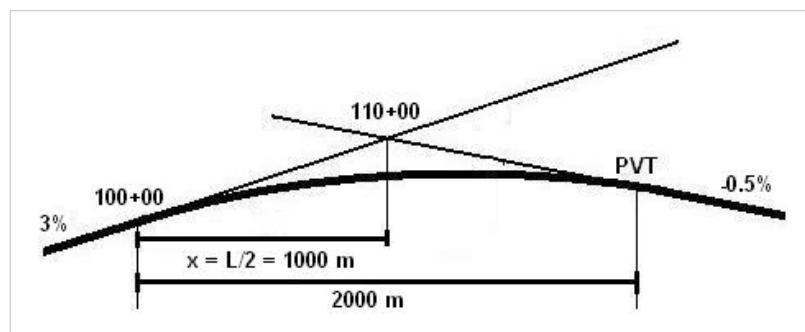


Problem:

A current roadway is climbing a hill at an angle of +3.0%. The roadway starts at station 100+00 and elevation of 1000 m. At station 110+00, there is an at-grade railroad crossing that goes over the sloped road. Since designers are concerned for the safety of drivers crossing the tracks, it has been proposed to cut a level tunnel through the hill to pass beneath the railroad tracks and come out on the opposite side. A vertical crest curve would connect the existing roadway to the proposed tunnel with a grade of (-0.5%). The prospective curve would start at station 100+00 and have a length of 2000 meters. Engineers have stated that there must be at least 10 meters of separation between the railroad tracks and the road to build a safe tunnel. Assume an equal tangent curve. With the current design, is this criteria met?



Solution:



One way to solve this problem would be to compute the elevation of the curve at station 110+00 and then see if it is at least 10 meters from the tangent. Another way would be to use the Offset Formula. Since A, L, and x are all known, this problem can be easily solved. Set x to 1000 meters to represent station 110+00.

$$Y = \frac{Ax^2}{200L} = \frac{(3.5)(1000)^2}{200(2000)} = 8.75 \text{ m}$$

The design DOES NOT meet the criteria.

Example 3: Stopping Sight Distance



Problem:

A current roadway has a design speed of 100 km/hr, a coefficient of friction of 0.1, and carries drivers with perception-reaction times of 2.5 seconds. The drivers use cars that allows their eyes to be 1 meter above the road. Because of ample roadkill in the area, the road has been designed for carcasses that are 0.5 meters in height. All curves along that road have been designed accordingly.

The local government, seeing the potential of tourism in the area and the boost to the local economy, wants to increase the speed limit to 110 km/hr to attract summer drivers. Residents along the route claim that this is a horrible idea, as a particular curve called "Dead Man's Hill" would earn its name because of sight distance problems. "Dead Man's Hill" is a crest curve that is roughly 600 meters in length. It starts with a grade of +1.0% and ends with (-1.0)%. There has never been an accident on "Dead Man's Hill" as of yet, but residents truly believe one will come about in the near future.

A local politician who knows little to nothing about engineering (but thinks he does) states that the 600-meter length is a long distance and more than sufficient to handle the transition of eager big-city drivers. Still, the residents push back, saying that 600 meters is not nearly the distance required for the speed. The politician begins a lengthy campaign to "Bring Tourism to Town", saying that the residents are trying to stop "progress". As an engineer, determine if these residents are indeed making a valid point or if they are simply trying to stop progress?



Solution:

Using sight distance formulas from other sections, it is found that 100 km/hr has an SSD of 465 meters and 110 km/hr has an SSD of 555 meters, given the criteria stated above. Since both 465 meters and 555 meters are less than the 600-meter curve length, the correct formula to use would be:

$$S < L : L_m = \frac{AS^2}{200(\sqrt{h_1} + \sqrt{h_2})^2} = \frac{2(555)^2}{200(\sqrt{1.0} + \sqrt{0.5})^2} = 1055 \text{ m}$$

Since the 1055-meter minimum curve length is greater than the current 600-meter length on "Dead Man's Hill", this curve would not meet the sight distance requirements for 110 km/hr.

This seems like a very large gap. The question becomes, was the curve even good enough at 100 km/hr? Using the same formula, the result is:

$$S < L : L_m = \frac{AS^2}{200(\sqrt{h_1} + \sqrt{h_2})^2} = \frac{2(465)^2}{200(\sqrt{1.0} + \sqrt{0.5})^2} = 740 \text{ m}$$

740 meters for a minimum curve length is far greater than the existing 600-meter curve. Therefore, the residents are correct in saying that "Dead Man's Hill" is a disaster waiting to happen. As a result, the politician, unable to hold public confidence by his "progress" comment, was forced to resign.

Thought Question

Problem

Sag curves have sight distance requirements because of nighttime sight distance constraints. The headlights on cars have a limited angle at which they can shine with bright enough intensity to see objects far off in the distance. If the government were to allow a wider angle of light to be cast out on standard car headlights, would this successfully provide more stopping sight distance?

Solution

Yes, of course. For a single car traveling on a road with many sag curves, the design speed could be increased since more road could be seen. However, when additional cars were

added to that same road, problems would begin to appear. With a greater angle of light being cast from headlights, drivers in opposing lanes would be severely blinded, forcing them to slow down to avoid causing an accident. Just think of the last time somebody drove by with their 'brights' on and blinded you. This problem could cause more accidents and force people to slow down, thus producing a net loss overall.

Sample Problem

Problem (Solution)

Variables

- L - Curve Length
- e - Elevation of designated point, such as PVC, PVT, etc.
- g - Grade
- A - Absolute difference of grade percentages for a certain curve, in percent
- y - Elevation of curve
- Y - Offset between grade tangent from PVC and curve elevation for a specific station
- h_1 - Height of driver's eye above roadway surface
- h_2 - Height of object above roadway surface
- H - Height of headlight
- β - Inclined angle of headlight beam, in degrees
- S - Sight Distance in question
- K - Rate of curvature
- L_m - Minimum Curve Length

Key Terms

- PVC: Point of Vertical Curve
- PVI: Point of Vertical Intersection
- PVT: Point of Vertical Tangent
- Crest Curve: A curve with a negative grade change (like on a hill)
- Sag Curve: A curve with a positive grade change (like in a valley)

Fundamentals of Transportation/ Vertical Curves/Problem



Problem:

To help prevent future collisions between cars and trains, an at-grade crossing of a rail road by a country road is being redesigned so that the county road will pass underneath the tracks. Currently the vertical alignment of the county road consists of an equal tangents crest vertical curve joining a 4% upgrade to a 3% downgrade. The existing vertical curve is 450 feet long, the PVC of this curve is at station 48+24.00, and the elevation of the PVC is 1591.00 feet. The centerline of the train tracks is at station 51+50.00. Your job is to find the shortest vertical curve that provides 20 feet of clearance between the new county road and the train tracks, and to make a preliminary estimate of the cut that will be needed to construct the new curve.

- Solution

Fundamentals of Transportation/ Vertical Curves/Solution



Problem:

To help prevent future collisions between cars and trains, an at-grade crossing of a rail road by a country road is being redesigned so that the county road will pass underneath the tracks. Currently the vertical alignment of the county road consists of an equal tangents crest vertical curve joining a 4% upgrade to a 3% downgrade. The existing vertical curve is 450 feet long, the PVC of this curve is at station 48+24.00, and the elevation of the PVC is 1591.00 feet. The centerline of the train tracks is at station 51+50.00. Your job is to find the shortest vertical curve that provides 20 feet of clearance between the new county road and the train tracks, and to make a preliminary estimate of the cut at the PVI that will be needed to construct the new curve.



Solution:

A.) What is the curve length?

The curves are equal tangent, so we know that: $PVI - PVC = PVT - PVI = L/2$

From this, we know:

- PVC of First Curve: 48+24
- PVI of Both Curves (this is a constant): $(48+24.00 + (450/2) = 50+49)$
- x_{rr} : Distance between PVI and railroad tracks: $(51+50 - 50+49 = 101 \text{ feet})$
- PVI Elevation: $1591 + (0.04*225) = 1600 \text{ feet}$

Use the offset formula for the first curve to find the vertical distance between the tangent and the curve:

$$Y = \frac{Ax^2}{200L}$$

Where:

- $A = 7$ (Change in Grade)

- $L = 450$ feet
- $x = 51+50 - 48+24 = 326$ feet

The offset, Y, at the railroad tracks's station is computed to be 8.27 feet.

The road is to be lowered an additional 20 feet. Therefore, the new offset at that sight would become 28.27 feet. The equation for the second curve becomes:

$$Y = 28.27 = \frac{7x^2}{200L}$$

Neither L (length of curve) or x (distance of offset from PVC) are known. However, we know this is an equal tangent curve, meaning the distance from PVC to PVI is $L/2$ for the curve in question. Also, the distance between PVI and the railroad tracks is x_{rr} , which is 101 feet. Therefore, $(L/2) + x_{rr}$ equals the distance from PVC to the railroad tracks, which is what we want for x. Thus, we are left with one unknown and one equation.

$$Y = 28.27 = \frac{7(L + 101)^2}{200L}$$

The new curve length is found to be 2812 feet.

B.) How deep is the cut at the PVI?

To find this, the offset formula can be used again, using the length $L/2$ as the distance from the PVC on any curve.

For the old curve, the offset is:

$$Y = \frac{A(L/2)^2}{200L} = \frac{7(450/2)^2}{200(450)} = 3.935 \text{ feet}$$

For the new curve, using data found from before, the offset is:

$$Y = \frac{A(L/2)^2}{200L} = \frac{7(2812/2)^2}{200(2812)} = 24.605 \text{ feet}$$

The difference between the two offsets is 20.67 feet. This is the depth of the cut at the PVI.

Transportation Economics/Pricing

Pricing

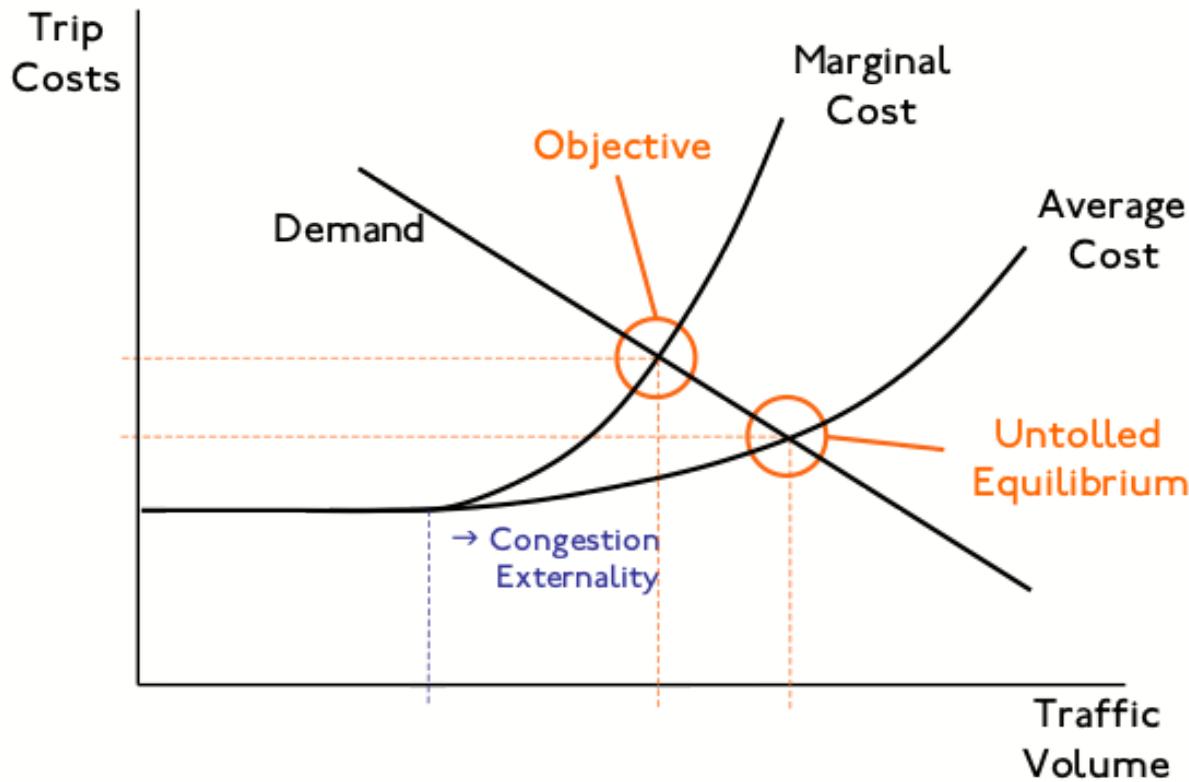
Rationales for Pricing

- Revenue
- Congestion management
- Off-loading costs or reallocating costs (changing who bears burden)
- Changing energy supply indicates declining gas tax revenue
- Encourage alternatives to driving

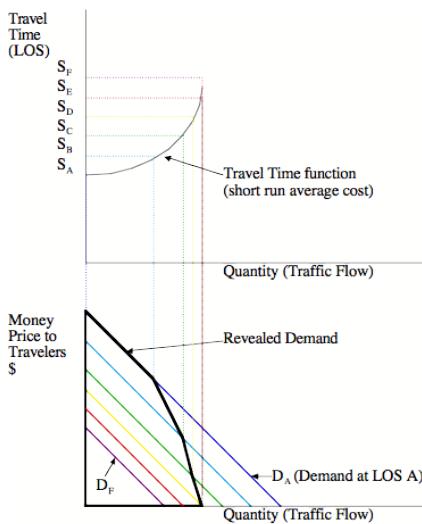
Theory

(based on Levinson, David (2002) *Financing Transportation Networks*^[1] Cheltenham (UK) and Northampton (US): Edward Elgar. (Chapter 10).

Theory: Congestion Pricing Brings About Efficient Equilibrium



Unpacking



A. (Top) Travel Time as Function of Traffic Flow
 B. (Bottom) Traffic Flow (Implicit and Revealed Demand) as a function of Money Price to Travelers

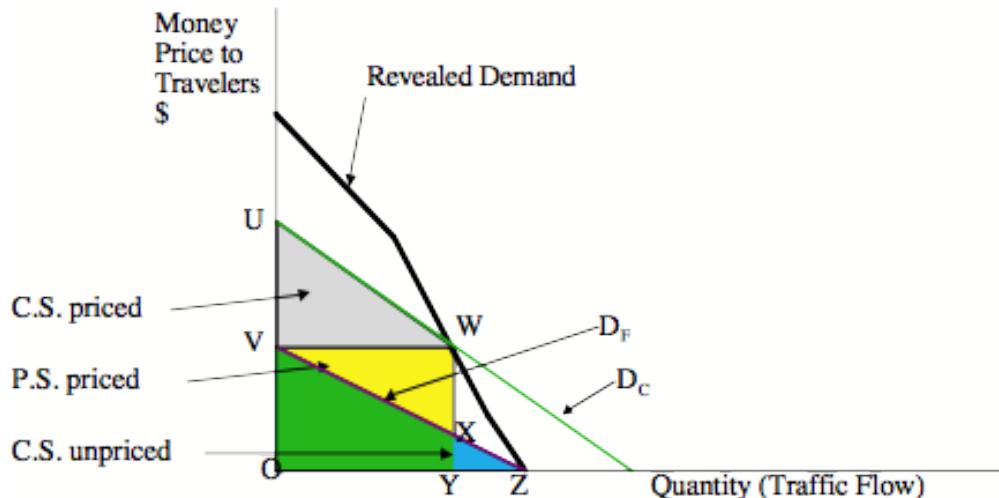
The top part of Figure 2 shows schematically the travel time to a driver (short run average cost) at a bottleneck or on a capacitated link resulting from various levels of approach flow. The travel time function relates travel time (or delay) and approach traffic flow. The greater the approach flow, the higher the travel time. At flows below capacity (level of service A (SA) or B (SB)), traffic flows smoothly, while at high approach flows, those above capacity, traffic is stop and start and is classified as level of service E (SE) or F (SF).

The bottom part of Figure 2 shows schematically the implicit demand for travel on a link as a function of the travel time. All else equal (for instance the price charged to users), demand to travel on a link at level of service A (DA) is higher than demand at level of service F (DF). However the demand and the travel time on a link are not independent, as shown in Figure 2(A).

So the implicit demand and revealed demand are not identical, rather the revealed demand is formed by projecting the travel time at a given flow onto the implicit demand curves. So for instance, when the price charged users is high, the revealed demand coincides with the implicit demand at level of service A (DA). As the prices are lowered, the revealed demand crosses the implicit demand curve at level of service B (DB), then DC, DD, DE and finally at a zero money price it crosses DF. While the actual prices that generate specific demand levels vary from place to place with local circumstances, demand preferences, and market conditions, the general trend (higher prices gives lower approach flow gives better level of service) is simply an application of the law of demand from economics along with traffic flow theory.

In other words, the change in welfare with congestion pricing depends not only on both the change in price and quantity, but also on the change in reservation price. The reservation price is the amount travelers would be willing to pay at a given level of service. And at better levels of service, travelers (and potential travelers) have a higher reservation price.

Welfare Analysis



Legend

Consumer's Surplus Priced = UVW

Producer's Surplus Priced = OVWY

Consumer's Surplus Unpriced = VOZ

Producer's Surplus Unpriced = 0

The movement along the revealed demand curve follows the shape of the curve shown above because of the relationship between traffic flow (quantity demanded) and travel time. Assume for instance that each level of service category represents a one-minute increase in travel time from the immediately better travel time. So in the graph, let the level of service for a one minute trip be denoted SA , and for a six minute trip, SF. The amount of traffic necessary to move from 1 minute to 2 minutes exceeds the amount to move from 2 to 3 minutes. In other words, there is a rising average (and thus marginal) cost in terms of time.

The concepts in Figure 2 can be used to develop the welfare analysis shown in Figure 3. There are several areas of interest in Figure 3. The first is defined by the lower left triangle (the blue + green) (triangle VOZ) which is the consumer surplus when the road is unpriced. The second is the producer surplus (profit) to the road authority when the road is priced, illustrated by the rectangle formed in the lower left (yellow + green) (rectangle OVWY). The third is the consumer surplus when the road is priced, shown in gray (triangle UVW). This consumer surplus represents a higher reservation price than the other because the level of service is better when flow is lower.

That first area needs to be compared to the sum of the second and third areas. If the sum of the second and third areas (OUWY) is larger than the first (OVZ), then pricing has higher welfare than remaining unpriced. Similarly, two price levels can be compared. In other words, the welfare gain from pricing is equal to the yellow + gray area (VUWX) minus the blue area (XYZ). In this particular figure, consumer's surplus is maximized when the good is free, but overall welfare (including producer's surplus) is not. Whether consumer's surplus is in fact higher in a given situation depends on the slopes of the various demand curves.

The greatest welfare is achieved by maximizing the sum of the producer's surplus rectangle and the consumer's surplus "triangle" (it may not be a true triangle). This must recognize that the consumers surplus triangle's hypotenuse must follow an underlying demand curve,

not the revealed demand curve. Differentiating the level of service (for instance, providing two different levels of service at two different prices) may result in higher overall welfare (though not necessarily higher welfare for each individual).

Use of the Revenue

How welfare is measured and how it is perceived are two different things. If the producer's surplus is not returned to the users of the system somehow the users will perceive an overall welfare gain as a personal loss because it would be acting as an additional tax. The money can be returned through rebates of other taxes or reinvestment in transportation. It should be noted that the entire argument can be made in reverse, where consumer and producers surplus are measured in time rather than money, and the level of service is the monetary cost of travel. This however has less practical application.

Pricing and Cost Recovery

In low volume situations, those that are uncongested, it is unlikely that the revenue from marginal cost congestion pricing will recover long term fixed costs. This is because the marginal impacts of an additional car when volume is low is almost zero, so that additional revenue which can be raised with marginal cost pricing is also zero. Imagine a road with one car - the car's marginal impact is zero, a marginal cost price would also be zero, its revenue would thus be zero, which is less than the fixed costs.

Add a second car, and marginal impacts are still nearly zero - a phenomenon which remains true until capacity is approached.

Vickery's Types of Congestion

- Simple interaction - light traffic, one car blocked by another, delay is proportional to Q^2
- Multiple interaction - $0.5 < V/C < 0.9$

$$Z = t - t_o = 1/s - 1/s_o = ax^k$$

t actual time, to freeflow time, K ~ 3-5

- Bottleneck see below
- Triggerneck - overflow affects other traffic
- Network and Control - Traffic control devices transfer delay
- General Density - high traffic level in general

Marginal Cost Pricing

Transportation is a broad field, attracting individuals with backgrounds in engineering, economics, and planning, among others who don't share a common model or worldview about traffic congestion. Economists look for received technological functions that can be analyzed, but risk misinterpreting them. Engineers seek basic economic concepts to manage traffic, which they view as their own purview. These two fields intersect in the domain of congestion pricing. However many engineers view pricing with suspicion, believing that many economists are overstating its efficacy, while the economists are frustrated with engineering intransigence, and consider engineers as lacking understanding of basic market principles.

This section applies the microscopic model of traffic congestion that to congestion pricing, and allows us to critique the plausibility of several economic models of congestion that have appeared in the literature.

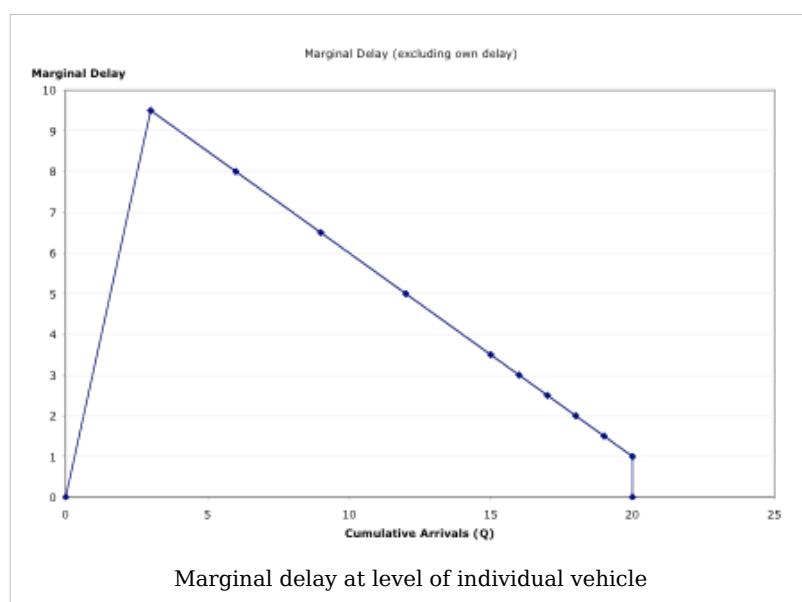
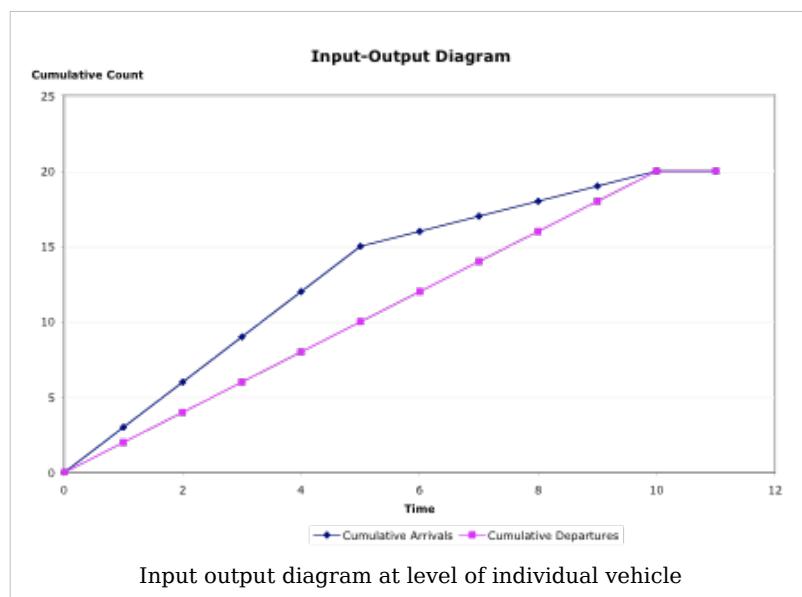
This section uses the idea of queueing and bottlenecks to explain congestion. If there were no bottlenecks (which can be physical and permanent such as lane drops or steep grades, or variable such as a traffic control device, or temporary due to a crash), there would be very little congestion. Vehicles interacting on an uncongested road lead to relatively minimal delays and are not further considered (Vickery 1969, Daganzo 1995). We define congestion, or the congested period, to be the time when there is queueing. This exceeds the time when arrivals exceed departures, as every vehicle has to wait for all previous vehicles to depart the front of the queue before it can.

A previous section developed the queueing model of congestion.

What is the implication of our queueing models for marginal cost pricing?

First, the use of hourly average time vs. flow functions such as the Bureau of Public Roads function, (which we introduced in the discussion of route choice (which approximates the hourly average delay from a queueing model) ignores that different vehicles within that hour have different travel times. They are at best useful for coarse macroscopic analyses, but should never be applied to the level of individual vehicles.

Second, travel time for a vehicle through a bottleneck depends on the number of vehicles that have come before, but not on the number of vehicles coming after. Similarly, the marginal delay that a vehicle imposes is only imposed on vehicles that come after. This implies that the first vehicle in the queue imposes the highest marginal delay, and the last vehicles in the queue have the lowest marginal delay. A marginal delay function that looks like figure on the right (bottom) is generated from the typical input-output diagram shown on the figure on the right (top). If marginal cost were equal to



the marginal delay, then our pricing function would be unusual, and perhaps unstable. The instability might be tempered by making demand respond to price, rather than assuming it fixed (Rafferty and Levinson 2003), and by recognizing the stochastic nature of arrival and departure patterns, which would flatten the arrival curve to more closely resemble the departure curve, and thereby flatten the marginal delay curve.

However, the idea that the first vehicle "causes" the delay is a controversial point. Economists will sometimes argue Coase's position (1992) - that it takes two to have a negative externality, that there would be no congestion externality but for the arrival of the following vehicles. Coase is, of course, correct. Moreover, they would note that charging a toll to the following vehicle will discourage that vehicle from coming and might also eliminate the congestion externality. This may also be true. This would however be charging the sufferer of congestion twice (once in terms of time, a second in terms of toll), while the person with the faster trip (earlier in the queue) wouldn't pay at all. Further, it is the followers who have already internalized the congestion externality in their decision making, so tolling them is charging them twice, in contrast to charging the leaders. Given that charging either party could eliminate the externality, it would be more reasonable to charge the delayer than the delayed, which is much like the "polluter pays principle" advocated by environmentalists. It would also be more equitable, in that total costs (congestion delay + toll) would now be equalized across travelers. The disadvantage of this is that the amount of the delay caused is unknown at the time that the first traveler passes; at best it can be approximated.

Implicitly, this privileges the "right to uncongested travel" over the "right to unpriced travel". That is a philosophical question, but given that there is to be some mechanism to finance highways, we can eliminate the notion of unpriced travel altogether, the remaining issue is how to implement financing: with insensitive prices like gas taxes or flat tolls or with time-dependent (or flow-dependent) congestion prices.

The marginal cost equals marginal delay formulation does ignore the question of schedule delay. There are practical reasons for not including schedule delay in a marginal cost price. Unlike delay, schedule delay is not easily measured. While a road administrator can tell you from traffic counts how much delay a traveler caused, the administrator has no clue as to how much schedule delay was caused. Second, if the late (early) penalty is large, then it dominates scheduling. Travelers can decide whether they would rather endure arriving early (without delay) or arriving on time, with some delay (or some combination of the two), presumably minimizing their associated costs. If they choose the delay, it is the lower cost alternative. That lower cost is the one they suffer, and thus that serves as a lower bound on the marginal cost to attribute to other travelers. If they choose schedule delay (which then becomes the cost they face) and avoid the delay, they are affected as well by other travelers, but in a way that is unknown to pricing authorities. They are "priced out of the market", which happens all the time. In short, endogenizing schedule delay would be nice, but requires more information than is actually available.

References

- Coase, Ronald (1992), "The Problem of Social Cost, and Notes on the Problem of Social Cost", reprinted in *The Firm, The Market and the Law*, Chicago: University of Chicago Press.
- Daganzo, Carlos. (1995). *An Informal Introduction To Traffic Flow Theory*. University of California at Berkeley, Institute of Transportation Studies. Berkeley, CA.
- Levinson, David and Peter Rafferty (2004) Delayer Pays Principle: Examining Congestion Pricing with Compensation. *International Journal of Transport Economics* 31:3 295-311
- Vickery, William (1969) Congestion Theory and Transport Investment. *American Economic Review* 59, pp.251-260.
- Levinson, David and Peter Rafferty (2004) Delayer Pays Principle: Examining Congestion Pricing with Compensation [2]. *International Journal of Transport Economics* 31:3 295-311

References

- [1] <http://nexus.umn.edu/Books/FTN.html>
[2] <http://nexus.umn.edu/Papers/DelayerPays.pdf>

Fundamentals of Transportation/ Conclusions

Transportation is multi-modal, multi-disciplinary field that requires the efforts of people from a spectrum of backgrounds. Planners, engineers and policy makers are the primary groups, but architects, economists, and operations researchers serve in roles that benefit the industry. In order for an idea to become a road, it must go through several planning and design stages. This section serves as a conclusion for this wikibook as well as a summary of the building process for a roadway.

Connecting Places

- Planning
- Assessing Needs: Testing Alternatives (Hypotheses)
- Predicting Demand
- Trip Generation
- Trip Distribution
- Mode Choice
- Route Assignment

Determining Road Capacity (Width)

- Level of service
- Design to satisfy standard
- Understand implications of traffic (queues, shockwaves, congestion)

Determining Road Alignment (Length)

- Properties of drivers, vehicles, roadway
- Maximum grade associated with speed, vehicle type
- Stopping sight distance
- Vertical alignment (sags & crests)
- Horizontal alignment

Determining Road Strength (Depth)

- Given load on the road, how long can the road hold up.
- Design pavement depth to satisfy predicted traffic.

Putting it together

Planning determines origin/destination demand by vehicle type ... feeds into Level of Service

- LOS feeds into in-depth traffic analysis and determines number of lanes
- Truck demand determines maximum grade
- Stopping sight distance determines horizontal and vertical curvatures.
- Truck demand determines pavement thickness

Implications

Fraction of demand (trucks) is determining design characteristics (grade, depth)

Fraction of demand (peak hour traffic) is determining width.

1. Financing should consider this
2. Design should consider this. Maybe cars and trucks should have separate routes. (car-only highways would be cheaper, separation would be safer).
3. We don't ask one building to serve all needs, why should one road?

Thought Question

How does an idea become a road? Explain how an idea becomes a road. Illustrate your explanation with a consistent set of examples for all of the major steps of analysis that are required (i.e. the output of one example should be the input to the next example).

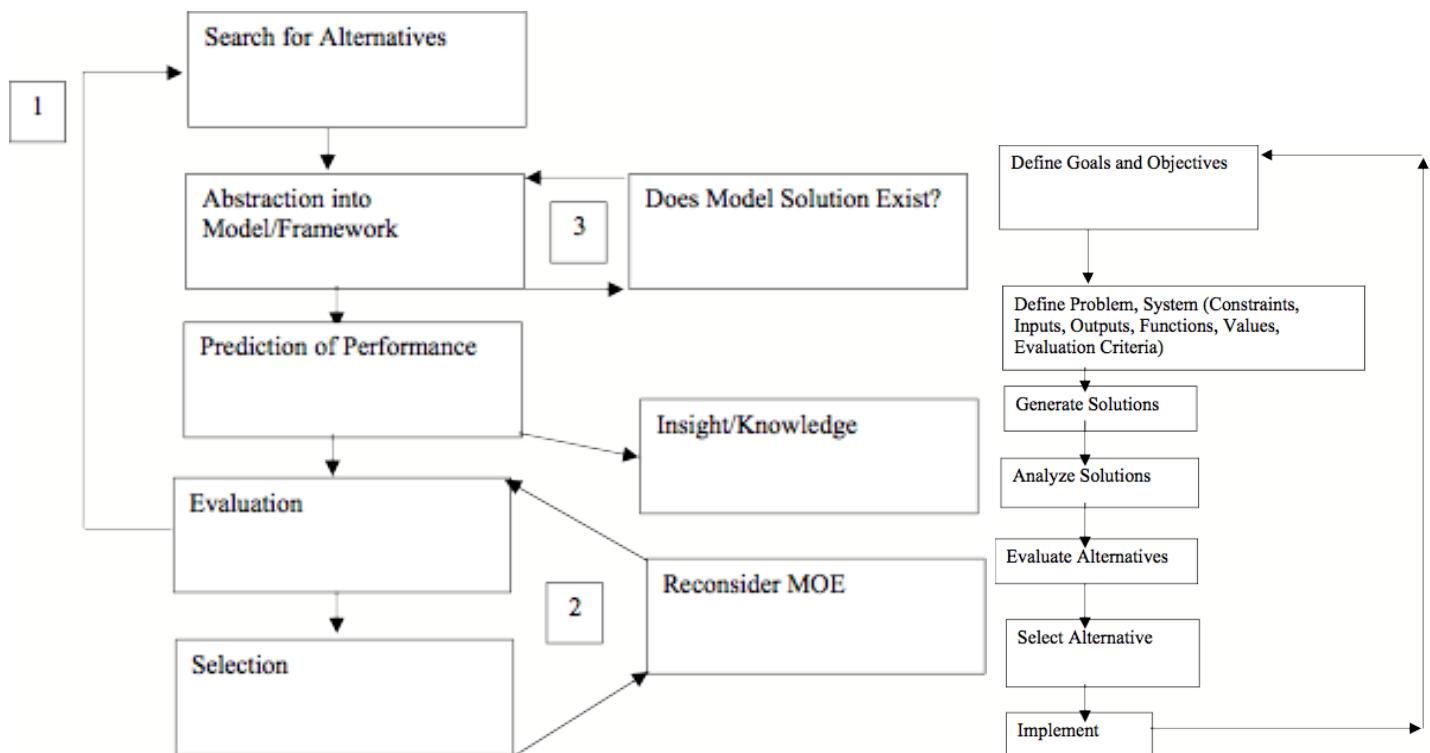
Transportation Economics/Decision making

Decision Making

Decision Making is the process by which one alternative is selected over another. Decision making generally occurs in the planning phases of a transportation project, but last minute decision making has been shown to occur, sometimes successfully. Several procedures for making decisions have been outlined in effort to minimize inefficiencies or redundancies.

Rational Planning

(see *Rational planning model article in wikipedia* [1])



How does one (rationally) decide what to do?

1. Identify Needs
2. Set Objectives
3. Develop Options
4. Evaluate Options
5. Select Best Option
6. Implement Best Option
7. Evaluate Outcome

We need a tool to "Identify Needs" and "Evaluate Options". This may be the transportation forecasting model.

Step A. Define the System

1. objectives - measures the effectiveness or performance
2. environment - things which affect the system but are not affected by it
3. resources - factor inputs to do the work
4. components - set of activities or tasks of the system
5. management - sets goals, allocates resources and exercises control over components
6. model of how variables in 1-5 relate to each other

Step B Generate and assess alternatives available to management

1. . algorithms-systematic search over available alternatives
 1. analytical
 2. exact numerical
 3. heuristic numerical
2. generate alternatives selectively, evaluate subjectively
 1. fatal flaw analysis
 2. simple rating schemes
 3. Delphi exercises
3. generate alternatives judgmentally, evaluate scientifically using system model
4. need for "stopping rule"

Step C Choose alternatives

The analyst is generally not the decision maker. The actual influence of the results of the analysis in actual decisions will depend on:

1. determinacy of evaluation
2. confidence in the results on the part of the decision maker
3. consistency of rating among alternatives

Step D Implementation

Just Do It.

Step E Evaluation

1. definition: output from a later step in systems analysis used as input to a later step.
1. Examples.

analysis leads to revisions in systems definition implementation experience leads to a revision of output system definition or values that underlay that definition.

Thought Question

Is the Rational Planning process Rational?

Some Issues with Rational Planning

Nevertheless, some issues remain with the rational planning model:

- Limited Computational or Solution Generating Capacity
- Incomplete Information
- Cost of Analysis
- Conflicting Goals/Evaluation Criteria
- Reliance on Experts (What about the People?)

Alternative Planning Decision Making Paradigms. Are They Irrational?

No one really believes the rational planning process is a good description of most decision making, as it is highly idealized. Alternatives paradigms to the rational planning process include:

- Satisficing
- Incrementalist
- Organizational Process
- Political Bargaining
- Decomposition/Hierarchical Strategy/Tactics/Operations

References

[1] http://en.wikipedia.org/wiki/Rational_planning_model

Article Sources and Contributors

Fundamentals of Transportation *Source:* <http://en.wikibooks.org/w/index.php?oldid=1523822> *Contributors:* AdamD, Adrignola, Dallas1278, Darklama, DavidLevinson, Mike.lifeguard, 1 anonymous edits

Fundamentals of Transportation/Accidents *Source:* <http://en.wikibooks.org/w/index.php?oldid=1419499> *Contributors:* AdamD, DavidLevinson, Jomegat, 2 anonymous edits

Fundamentals of Transportation/Introduction *Source:* <http://en.wikibooks.org/w/index.php?oldid=1243078> *Contributors:* AdamD, DavidLevinson

Transportation Economics/Introduction *Source:* <http://en.wikibooks.org/w/index.php?oldid=1329327> *Contributors:* DavidLevinson

Fundamentals of Transportation/Geography and Networks *Source:* <http://en.wikibooks.org/w/index.php?oldid=1255572> *Contributors:* AdamD, DavidLevinson, 1 anonymous edits

Fundamentals of Transportation/Trip Generation *Source:* <http://en.wikibooks.org/w/index.php?oldid=1271845> *Contributors:* AdamD, DavidLevinson, Reddi, 1 anonymous edits

Fundamentals of Transportation/Trip Generation/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238333> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Trip Generation/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1242935> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Destination Choice *Source:* <http://en.wikibooks.org/w/index.php?oldid=1243600> *Contributors:* AdamD, DavidLevinson, 1 anonymous edits

Fundamentals of Transportation/Destination Choice/Background *Source:* <http://en.wikibooks.org/w/index.php?oldid=1505583> *Contributors:* Adrignola, DavidLevinson

Fundamentals of Transportation/Mode Choice *Source:* <http://en.wikibooks.org/w/index.php?oldid=1329322> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Mode Choice/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238343> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Mode Choice/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238346> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Route Choice *Source:* <http://en.wikibooks.org/w/index.php?oldid=1505578> *Contributors:* AdamD, Adrignola, DavidLevinson

Fundamentals of Transportation/Route Choice/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238354> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Route Choice/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238361> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Evaluation *Source:* <http://en.wikibooks.org/w/index.php?oldid=1294547> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Planning *Source:* <http://en.wikibooks.org/w/index.php?oldid=1243100> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Operations *Source:* <http://en.wikibooks.org/w/index.php?oldid=1242946> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Queueing *Source:* <http://en.wikibooks.org/w/index.php?oldid=1266267> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Queueing/Problem1 *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238371> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Queueing/Solution1 *Source:* <http://en.wikibooks.org/w/index.php?oldid=1242949> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Queueing/Problem2 *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238375> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Queueing/Solution2 *Source:* <http://en.wikibooks.org/w/index.php?oldid=1242950> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Queueing/Problem3 *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238385> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Queueing/Solution3 *Source:* <http://en.wikibooks.org/w/index.php?oldid=1242952> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Traffic Flow *Source:* <http://en.wikibooks.org/w/index.php?oldid=1307128> *Contributors:* AdamD, DavidLevinson, 1 anonymous edits

Fundamentals of Transportation/Traffic Flow/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1241928> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Traffic Flow/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1242955> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Queueing and Traffic Flow *Source:* <http://en.wikibooks.org/w/index.php?oldid=1242956> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Shockwaves *Source:* <http://en.wikibooks.org/w/index.php?oldid=1385890> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Shockwaves/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238396> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Shockwaves/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238399> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Traffic Signals *Source:* <http://en.wikibooks.org/w/index.php?oldid=1242983> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Traffic Signals/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1241902> *Contributors:* AdamD

Fundamentals of Transportation/Traffic Signals/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1241903> *Contributors:* AdamD

Fundamentals of Transportation/Design *Source:* <http://en.wikibooks.org/w/index.php?oldid=1236498> *Contributors:* AdamD, DavidLevinson, 1 anonymous edits

Fundamentals of Transportation/Sight Distance *Source:* <http://en.wikibooks.org/w/index.php?oldid=1241905> *Contributors:* AdamD, DavidLevinson, 2 anonymous edits

Fundamentals of Transportation/Sight Distance/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238406> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Sight Distance/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238407> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Grade *Source:* <http://en.wikibooks.org/w/index.php?oldid=1436014> *Contributors:* AdamD, DavidLevinson, Van der Hoorn

Fundamentals of Transportation/Grade/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238412> *Contributors:* AdamD

Fundamentals of Transportation/Grade/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1241907> *Contributors:* AdamD, Ramac

Fundamentals of Transportation/Earthwork *Source:* <http://en.wikibooks.org/w/index.php?oldid=1242994> *Contributors:* AdamD

Fundamentals of Transportation/Earthwork/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1241911> *Contributors:* AdamD, Ramac

Fundamentals of Transportation/Earthwork/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1241912> *Contributors:* AdamD

Fundamentals of Transportation/Horizontal Curves *Source:* <http://en.wikibooks.org/w/index.php?oldid=1243010> *Contributors:* AdamD, DavidLevinson

Fundamentals of Transportation/Horizontal Curves/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238525> *Contributors:* AdamD

Fundamentals of Transportation/Horizontal Curves/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1242971> *Contributors:* AdamD, Ramac

Fundamentals of Transportation/Vertical Curves *Source:* <http://en.wikibooks.org/w/index.php?oldid=1385897> *Contributors:* AdamD, DavidLevinson, 2 anonymous edits

Fundamentals of Transportation/Vertical Curves/Problem *Source:* <http://en.wikibooks.org/w/index.php?oldid=1238532> *Contributors:* AdamD

Fundamentals of Transportation/Vertical Curves/Solution *Source:* <http://en.wikibooks.org/w/index.php?oldid=1241917> *Contributors:* AdamD, Ramac

Transportation Economics/Pricing *Source:* <http://en.wikibooks.org/w/index.php?oldid=1329317> *Contributors:* DavidLevinson, 1 anonymous edits

Fundamentals of Transportation/Conclusions *Source:* <http://en.wikibooks.org/w/index.php?oldid=1235516> *Contributors:* AdamD, DavidLevinson

Transportation Economics/Decision making *Source:* <http://en.wikibooks.org/w/index.php?oldid=1505581> *Contributors:* Adrignola, DavidLevinson, 2 anonymous edits

Image Sources, Licenses and Contributors

Image:TransportationInputsOutputs.png Source: <http://en.wikibooks.org/w/index.php?title=File:TransportationInputsOutputs.png> License: unknown Contributors: User:DavidLevinson

Image:DemandForTravel.png Source: <http://en.wikibooks.org/w/index.php?title=File:DemandForTravel.png> License: unknown Contributors: User:DavidLevinson

Image:NegativeFeedback.png Source: <http://en.wikibooks.org/w/index.php?title=File:NegativeFeedback.png> License: unknown Contributors: User:DavidLevinson

Image:TransportationSupplyAndDemand.png Source: <http://en.wikibooks.org/w/index.php?title=File:TransportationSupplyAndDemand.png> License: unknown Contributors: User:DavidLevinson

Image:PositiveFeedbackVirtuous.png Source: <http://en.wikibooks.org/w/index.php?title=File:PositiveFeedbackVirtuous.png> License: unknown Contributors: User:DavidLevinson

Image:PositiveFeedbackVicious.png Source: <http://en.wikibooks.org/w/index.php?title=File:PositiveFeedbackVicious.png> License: unknown Contributors: User:DavidLevinson

Image:Routemaster.JPG Source: <http://en.wikibooks.org/w/index.php?title=File:Routemaster.JPG> License: unknown Contributors: Dapete, KF, Kneiphof, Kotasik, LoopZilla

Image:FOT-Planning-Hierarchy.png Source: <http://en.wikibooks.org/w/index.php?title=File:FOT-Planning-Hierarchy.png> License: Public Domain Contributors: David Levinson

Image:HomeWorkShopEat.png Source: <http://en.wikibooks.org/w/index.php?title=File:HomeWorkShopEat.png> License: Public Domain Contributors: User:DavidLevinson

Image:TripGeneration01.png Source: <http://en.wikibooks.org/w/index.php?title=File:TripGeneration01.png> License: Public Domain Contributors: User:DavidLevinson

Image:TripGeneration02.png Source: <http://en.wikibooks.org/w/index.php?title=File:TripGeneration02.png> License: Public Domain Contributors: User:DavidLevinson

Image:Evolution-tasks.png Source: <http://en.wikibooks.org/w/index.php?title=File:Evolution-tasks.png> License: GNU General Public License Contributors: Brian0918, Gryffindor, TOR, The wub, Thuresson, Zscout370, 3 anonymous edits

Image:DiversionCurveMC.png Source: <http://en.wikibooks.org/w/index.php?title=File:DiversionCurveMC.png> License: Public Domain Contributors: Original uploader was DavidLevinson at en.wikipedia Later version(s) were uploaded by Andrew pmk at en.wikipedia.

Image:ThurstonMC.png Source: <http://en.wikibooks.org/w/index.php?title=File:ThurstonMC.png> License: Public Domain Contributors: Original uploader was DavidLevinson at en.wikipedia Later version(s) were uploaded by Andrew pmk at en.wikipedia.

Image:LikelihoodMC.png Source: <http://en.wikibooks.org/w/index.php?title=File:LikelihoodMC.png> License: Public Domain Contributors: David Levinson

Image:TwoRouteNetwork.png Source: <http://en.wikibooks.org/w/index.php?title=File:TwoRouteNetwork.png> License: Public Domain Contributors: Original uploader was DavidLevinson at en.wikipedia Later version(s) were uploaded by Andrew pmk at en.wikipedia.

Image:EquilibriumAssignment2.png Source: <http://en.wikibooks.org/w/index.php?title=File:EquilibriumAssignment2.png> License: Public Domain Contributors: Original uploader was DavidLevinson at en.wikipedia

Image:EquilibriumAssignment3.png Source: <http://en.wikibooks.org/w/index.php?title=File:EquilibriumAssignment3.png> License: Public Domain Contributors: Original uploader was DavidLevinson at en.wikipedia Later version(s) were uploaded by Andrew pmk at en.wikipedia.

Image:PDFNPV.PNG Source: <http://en.wikibooks.org/w/index.php?title=File:PDFNPV.PNG> License: Public Domain Contributors: User:AdamD

Image:PDF2NPV.PNG Source: <http://en.wikibooks.org/w/index.php?title=File:PDF2NPV.PNG> License: Public Domain Contributors: User:AdamD

Image:PDF3NPV.PNG Source: <http://en.wikibooks.org/w/index.php?title=File:PDF3NPV.PNG> License: Public Domain Contributors: User:AdamD

Image:ChangeInConsumersSurplus.png Source: <http://en.wikibooks.org/w/index.php?title=File:ChangeInConsumersSurplus.png> License: Public Domain Contributors: User:DavidLevinson

Image:Queue_System.PNG Source: http://en.wikibooks.org/w/index.php?title=File:Queue_System.PNG License: Public Domain Contributors: User:AdamD

Image:QueueingIO.png Source: <http://en.wikibooks.org/w/index.php?title=File:QueueingIO.png> License: unknown Contributors: User:DavidLevinson

Image:Traffic Congestion Brasilia.jpg Source: http://en.wikibooks.org/w/index.php?title=File:Traffic_Congestion_Brasilia.jpg License: unknown Contributors: User:Mariordo

Image:TimeSpaceDiagram.png Source: <http://en.wikibooks.org/w/index.php?title=File:TimeSpaceDiagram.png> License: unknown Contributors: User:DavidLevinson

Image:TraditionalTrafficFlowModel.png Source: <http://en.wikibooks.org/w/index.php?title=File:TraditionalTrafficFlowModel.png> License: unknown Contributors: User:DavidLevinson

Image:TriangularTrafficFlowModel.png Source: <http://en.wikibooks.org/w/index.php?title=File:TriangularTrafficFlowModel.png> License: unknown Contributors: User:DavidLevinson

Image:TrafficPhases.png Source: <http://en.wikibooks.org/w/index.php?title=File:TrafficPhases.png> License: unknown Contributors: User:DavidLevinson

Image:QueueingIOPhasess.png Source: <http://en.wikibooks.org/w/index.php?title=File:QueueingIOPhasess.png> License: unknown Contributors: User:DavidLevinson

Image:ShockwaveGraphic.png Source: <http://en.wikibooks.org/w/index.php?title=File:ShockwaveGraphic.png> License: Public Domain Contributors: David Levinson

Image:ShockwaveGraphic02.png Source: <http://en.wikibooks.org/w/index.php?title=File:ShockwaveGraphic02.png> License: Public Domain Contributors: User:DavidLevinson

Image:Feuvert2.jpg Source: <http://en.wikibooks.org/w/index.php?title=File:Feuvert2.jpg> License: Public Domain Contributors: User:Wazouille

Image:AASHTOSSD.png Source: <http://en.wikibooks.org/w/index.php?title=File:AASHTOSSD.png> License: unknown Contributors: User:DavidLevinson

Image:I-90_floating_bridges_looking_east.JPG Source: http://en.wikibooks.org/w/index.php?title=File:I-90_floating_bridges_looking_east.JPG License: GNU Free Documentation License Contributors: Infrogmation, Jmabel, Ronaldino, Strangnet, Tano4595, Xnatedawgx

Image:Triaxle dump truck 2005-10-06.km.jpg Source: http://en.wikibooks.org/w/index.php?title=File:Triaxle_dump_truck_2005-10-06.km.jpg License: GNU Free Documentation License Contributors: AdamD, G@iffen, Kymacpherson, MB-one, Stefan Knauf, Wesha, 1 anonymous edits

Image:Cross-section.PNG Source: <http://en.wikibooks.org/w/index.php?title=File:Cross-section.PNG> License: Public Domain Contributors: User:AdamD

Image:CutFillDiagram.PNG Source: <http://en.wikibooks.org/w/index.php?title=File:CutFillDiagram.PNG> License: Public Domain Contributors: User:AdamD

Image:MassDiagram.PNG Source: <http://en.wikibooks.org/w/index.php?title=File:MassDiagram.PNG> License: Public Domain Contributors: User:AdamD

Image:Ring_3_and_E6_Oslo_Quevaal.jpg Source: http://en.wikibooks.org/w/index.php?title=File:Ring_3_and_E6_Oslo_Quevaal.jpg License: unknown Contributors: User:Quevaal

Image:Horizontal_Curve.JPG *Source:* http://en.wikibooks.org/w/index.php?title=File:Horizontal_Curve.JPG *License:* Public Domain *Contributors:* User:AdamD

Image:Costal_Redwood.jpg *Source:* http://en.wikibooks.org/w/index.php?title=File:Costal_Redwood.jpg *License:* GNU Free Documentation License *Contributors:* Ayacop, H-stt, Quadell, Urban

Image:Centerline_Rumble_Strip.jpg *Source:* http://en.wikibooks.org/w/index.php?title=File:Centerline_Rumble_Strip.jpg *License:* Public Domain *Contributors:* User:SayCheeeeeese

Image:Vertical_Curve.PNG *Source:* http://en.wikibooks.org/w/index.php?title=File:Vertical_Curve.PNG *License:* Public Domain *Contributors:* User:AdamD

Image:Vertical_Curve_Example_2.JPG *Source:* http://en.wikibooks.org/w/index.php?title=File:Vertical_Curve_Example_2.JPG *License:* Public Domain *Contributors:* User:AdamD

Image:TE-Pricing-EquilibriumCongestion.png *Source:* <http://en.wikibooks.org/w/index.php?title=File:TE-Pricing-EquilibriumCongestion.png> *License:* unknown *Contributors:* User:DavidLevinson

Image:TE-Pricing-EquilibriumUnpacked.png *Source:* <http://en.wikibooks.org/w/index.php?title=File:TE-Pricing-EquilibriumUnpacked.png> *License:* unknown *Contributors:* User:DavidLevinson

Image:TE-Pricing-WelfareAnalysis.png *Source:* <http://en.wikibooks.org/w/index.php?title=File:TE-Pricing-WelfareAnalysis.png> *License:* unknown *Contributors:* User:DavidLevinson

Image:CongestionPricingA.png *Source:* <http://en.wikibooks.org/w/index.php?title=File:CongestionPricingA.png> *License:* unknown *Contributors:* User:DavidLevinson

Image:CongestionPricingB.png *Source:* <http://en.wikibooks.org/w/index.php?title=File:CongestionPricingB.png> *License:* unknown *Contributors:* User:DavidLevinson

Image:TE-Systems-RationalPlanning.png *Source:* <http://en.wikibooks.org/w/index.php?title=File:TE-Systems-RationalPlanning.png> *License:* unknown *Contributors:* User:DavidLevinson

Image:TE-Systems-Rational2.png *Source:* <http://en.wikibooks.org/w/index.php?title=File:TE-Systems-Rational2.png> *License:* unknown *Contributors:* User:DavidLevinson

License

Version 1.2, November 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.
51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others. This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

1. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
2. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
3. State on the Title page the name of the publisher of the Modified Version, as the publisher.
4. Preserve all the copyright notices of the Document.
5. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
6. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
7. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

8. Include an unaltered copy of this License.
9. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
10. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
11. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
12. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
13. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
14. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
15. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>. Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (c) YEAR YOUR NAME.

Permission is granted to copy, distribute and/or modify this document

under the terms of the GNU Free Documentation License, Version 1.2

or any later version published by the Free Software Foundation;

with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU

Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the

Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.