UC Berkeley
Research Reports

Title

Definition and Measurement of Transportation System Performance

Permalink

https://escholarship.org/uc/item/3cg3r4wr

Author

Dahlgren, Joy

Publication Date

1998

eScholarship.org Powered by the California Digital Library
University of California

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Definition and Measurement of Transportation System Performance

Joy Dahlgren

California PATH Research Report
UCB-ITS-PRR-98-24

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department of Transportation, Federal Highway Administration.

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.

Report for MOU 267, 268

June 1998

ISSN 1055-1425

CALIFORNIA PARTNERS FOR ADVANCED TRANSIT AND HIGHWAYS

DEFINITION AND MEASUREMENT OF TRANSPORTATION SYSTEM PERFORMANCE

Joy Dahlgren
California PATH
Institute of Transportation Studies
University of California
Berkeley

June 8, 1998

Abstract

Performance measures are needed to inform decisions regarding the overall level of resources to devote to transportation, where to allocate these resources, and how best to use them. The first two types of decision require regular monitoring of the system to reveal problems, which present opportunities for improvement. A few, easily measured indicators of the major benefits and costs of the system are appropriate for this task. For the third type of decision, how to best address a specific problem, a more comprehensive set of benefits and costs must be considered. Indicators must be found for those benefits and costs that would be impacted by alternative means of addressing the problem, so that the overall impacts of the alternatives can be compared. This report discusses the benefits and costs of transportation and recommends specific measures to use as indicators of the benefits and costs, as well as data sources and methods of measurement. It discusses the role of intelligent transportation systems as both an object of measurement and a means of measurement

Executive Summary

This study was sponsored by Caltrans New Technology and Research Program. However, this report is intended for a larger audience, including Caltrans Planning and District staff, regional transportation agency staff, and people currently developing the transportation system performance module of the 1998 California Transportation Plan. It is hoped that transportation professionals, citizen activists, and decision makers at all levels of government will find it useful in clarifying thinking about the goals of transportation and the linkage of measurable indicators of performance to these goals.

Performance measures should inform decisions. They should:

- reveal problems, which can be thought of as opportunities for improvement
- facilitate judging and choosing among strategies to utilize these opportunities
- measure the actual performance of the chosen strategy

Thus they inform decisions regarding the overall level of resources to devote to transportation, where to allocate these resources, and how best to use resources.

Transportation provides benefits, but it also imposes costs on the traveler and others, as well. The primary benefits are access to activities and markets. Costs are primarily time, money, property damage and injury, environmental degradation, and discomfort, such as stress due to difficult driving conditions or uncertainty regarding travel time. The overarching goal of the transportation system is to maximize the excess of benefits over costs. Therefore, performance measures should relate to these benefits and costs.

To reveal opportunities for improvement, and inform decisions regarding the level and allocation of resources to utilize these opportunities, regular assessment of system performance is needed. For these purposes, performance should be measured in terms of the amount of transportation provided, which indicates the amount of access provided (person-trips, person-miles, and $-miles of freight), travel time (both the average and the variation), the amount of property damage and injury, and public monetary costs. These are the primary determinants of overall benefits and costs, they are measurable, and they can be influenced by actions of transportation agencies.

A more comprehensive list of benefits and costs should be used in deciding which of these opportunities for improvement to exploit and how best to exploit them. Here, the costs of the improvements must be weighed against the benefits of the improvements. In many cases these benefits will be simply other costs that will be reduced by the improvement. For example the benefits of better traffic signal control will be reduced travel time.

Intelligent transportation systems (ITS) are both an object of performance measurement and a means of measurement. They will allow much more extensive and robust measurements of vehicle-miles and travel time, providing a better basis for decision.

They can also reduce travel time and accidents. However, their greatest value may be to reduce travel time variation and uncertainty.

In order to provide useful data for routine performance assessment, the following is recommended:

- limiting the number of routine measures, to concentrate resources on the most useful measures
- continuing research to improve methods for measuring volumes and travel times
- developing a prototype travel time and volume measurement system, including data collection, processing, and storage, and characterization of performance
- developing schedules for measuring performance that are appropriate for different traffic situations

Table of Contents

List of Tables

List of Figures

Acknowledgments

This study was funded by Caltrans New Technology and Research Program under MOU 267, "Definition and Measurement of the Performance of the Transportation System," and MOU 268, "Data for Measuring Performance." The author would like to acknowledge the helpful ideas of Joe Palen, Pravin Varaiya, Marty Wachs, David Gillen, Jianling Li, and Stein Weissenberger. Also helpful were the comments of reviewers in Caltrans New Technology and Research Program. The material in Appendix C, which is also contained in the LEAP website (www.path.berkeley.edu/~leap), was prepared by Jason Black. The work also benefited from the deliberations of the Caltrans Transportation Assessment Steering Committee, chaired by John Wolf and staffed by Bruce deTerra.

Definition and Measurement of Transportation System Performance

Joy Dahlgren

June 8, 1998

Introduction

This study was sponsored by Caltrans New Technology and Research Program. However, this report is intended for a larger audience, including Caltrans Planning and District staff, regional transportation agency staff, and people involved in developing transportation system performance measures for the 1998 California Transportation Plan and future updates. It is hoped that transportation professionals, citizen activists, and decision makers at all levels of government will find it useful in clarifying thinking about the goals of transportation and the linkage of measurable indicators of performance to these goals.

Purpose of Performance Measures

Transportation performance measures should inform decisions. They should reveal problems, which can be thought of as opportunities for improvement. They should be used to judge and choose among various strategies to utilize these opportunities. And they should measure the actual effects of implemented strategies so that future predictions of effects will be more accurate and future choices better informed. Thus they inform decisions regarding the overall level of resources to devote to transportation, where to allocate these resources, and how best to use resources.

Subject of Performance Measures

This report addresses the performance of the transportation system itself, not the performance of the organizations managing the system. Many factors affecting system performance can not be controlled by the system managers, and therefore are not suitable measures of management performance. Transportation agencies and service providers should have their own internal performance measures to assess management performance and activities that affect system performance. These would include such measures as the speed it takes to execute a contract or the number of people required to provide a certain level of transit service.

Organization of This Report

A clear understanding of the goals of the transportation system is essential to developing useful and cost-effective performance measures. Therefore Section 1 contains a discussion of goals from a variety of perspectives. These can all be tied to the basic benefits and costs of the transportation system and to a fundamental goal of maximizing the excess of these benefits over costs. In Section 2 performance measures for assessing the state of the transportation system and identifying opportunities for improvements are discussed. Section 3 describes a performance measurement process to inform decisions regarding which improvements to make. Section 4 briefly discusses performance measurement in evaluating implemented programs. The effects of intelligent transportation systems and measures needed to assess and predict their performance are discussed in Section 5. Finally, Section 6 contains recommendations for 1) a California performance measure program to assess the state of the transportation system 2) effects to be considered in allocating funds for competing uses, and 3) development of data sources for measuring performance. There are four appendices. Appendix A compares the benefit/cost framework developed in this report with the outcome framework proposed for the California Transportation Plan. Appendix B contains a discussion of the economic concept of consumer surplus, which some readers may find useful in understanding the interplay between benefits and costs. Appendix C describes surveillance methods currently in use or in development. A bibliography is also included.

1 Goals of the Transportation System

1.1 Travelers' Goals

1.1.1 Access to Activities

Most trips that people make are not made primarily for the fun of moving about, but in order to engage in some activity, such as work or visiting with friends. The primary goal for individual travelers is to gain access to activities they value[1].

1.1.2 Entertainment

However, sometimes people *do* travel just for fun. It is common to go out for walk or a bicycle ride solely for purposes of entertainment, and sometimes people go out for a drive, just for fun. People do not often go for pleasure drives during peak commute periods—it is more common on weekends

Entertainment is often a component goal of a trip taken for another purpose. People want to get somewhere *and* be entertained along the way. They enjoy the solitude and comfort of their car after a busy day, the view from an airplane or automobile, the fun of a ferry ride, or the thrill of riding a motorcycle. This is why people buy high-performance cars and car stereos and why Caltrans plants trees and flowering shrubs along freeways and pays attention to the visual design of overpasses and sound walls.

1.1.3 Minimizing Travel Costs

People also want to minimize their cost when they travel. These costs are:

- **Time--**Transportation takes time. There is not only the time spent actually traveling but also the time spent waiting to travel and arranging to travel, as well as the extra

[1] It is sometimes argued that access should be increased by more compact development and that this will increase welfare more than increased travel. However, people and communities choose the level of compactness that they prefer. People do this by their choice of housing location, employment, and other activities. Communities do this through their general plans, zoning, and planning decisions. If people choose to live in communities, take jobs, and engage in other activities that require a certain level of travel over the long run, this must reflect a preference for this activity pattern over one that would require less travel. This would seem to indicate that for them, more travel provides greater benefits than less travel. Of course, choices that require a high level of travel may impose higher costs on others than choices that require less travel.

time allowed as a buffer in case of delay. Another time-related cost is so-called *schedule* delay, that is, arriving earlier than desired in order to avoid travel delay.

- **Money—**Travelers spend money to purchase, maintain, and operate vehicles (including the cost of insurance and garage space) and to purchase transportation services such as transit and air travel.
- **Property loss and injury—**People sometimes have accidents or are robbed or assaulted by other people as they travel. The result can be damage to their car or other vehicle, loss of possessions, and injury, sometimes fatal.
- **Discomfort**—This is a class of costs that are hard to define and measure, but that are just as real as the other costs outlined above and may be just as significant. They include cognitive or psychic costs, such as the stress due to driving itself—which can be exacerbated by difficult driving conditions caused by weather, heavy traffic, mechanical problems, or other factors—as well as by worry about taking the wrong route, being late, missing the bus, having an accident, or having an unpleasant or dangerous encounter with another person. They include physical discomfort caused by being cold or wet, being exposed to loud noises, being pushed, sitting in one position too long, or having a bumpy ride.

1.2 Shippers' Goals

1.2.1 Access to Markets

Goods movement increases the value of goods by moving them from one location to another where they will have more value. Transportation allows producers to access more markets for raw materials, thus increasing the variety and complexity of products they can efficiently produce in a particular location. Transportation of finished goods expands producers' markets and consumers' choice. In both cases this leads to a more efficient allocation of resources[2].

1.2.2 Minimizing Shipping Costs

The costs of goods movement to the shipper are what he pays the carrier[3] to cover carrier costs and profit, as well as the inventory cost of the goods during shipment, which is proportional to the time elapsed during shipment.

1.3 Private Carriers' Goals

1.3.1 Profit

The goal of private carriers, whether carrying people or freight, is long term profit. At different stages of their development or in different situations they may focus on different aspects of profit, such as increasing market share or maximizing revenues. They will

[2] The provision of services is analogous to goods movement. Transportation of service providers, such as gardeners, plumbers, policemen, and teachers expands the area they can serve, and benefits consumers by providing them with a wider range of services and a larger selection of service providers.

[3] The shipper and carrier are the same in cases where firms supply their own transportation.

select markets, set fares and tariffs, and control costs to try to meet these goals. Travel time and safety affect both their costs and revenues.

1.4 Community Goals

1.4.1 Development

The goal of transportation system improvements is often to promote or enable development. In fact, even when the original goal is simply to reduce travel time, additional development often results. Over the long term, economic and social development has been the major benefit of transportation. Transportation has enabled spatial arrangements of activities that would not otherwise have been possible. When the only transportation was by foot, people's economic and social life was limited to the other people and the goods within walking distance. With transportation, larger numbers of people and a greater variety of goods can come together, enabling economies of scale and agglomeration, which make production more efficient. The resulting increased wealth makes more resources available for additional social and cultural activities. Today, because the goods necessary for life can be transported, large numbers of people can live in metropolitan areas, thus enabling development of a richer mix of economic and social activities. Furthermore, because of transportation they can come together for these activities without having to live in close proximity.

Of course, transportation only *enables* development, it does not by itself *cause* development. Here technology and population play the key role. Transportation facilities are built and services instituted because people see an opportunity for doing something new or something old in a different manner or in a different place.

The transportation improvements that open up new routes and modes of access, such as the railroads that linked California with the rest of the United States, have the greatest developmental effect, changing the direction and scale of development. These are followed by transportation improvements that expand access within the new area or add new modes of access that enable new types of economic activity. Finally, there are improvements that are not designed to affect development, but rather to reduce costs, generally time. But of course, by reducing costs they can also influence development. In fact, each of these types of transportation improvements exerts its effects through reducing transportation costs—it is only the degree of cost reduction that differs.

At this time in California, there are few opportunities for improvements of the first type. Most areas with development potential have been opened up and the only new mode on the horizon is the automated vehicle, which is still some years in the future. A few transportation improvements are of the second type, allowing expansion of development. However, many of the economic and social benefits of location and the spatial organization of activities have already been achieved. Most transportation improvements that will be made in California in the future are of the third type, designed to reduce the cost of transportation or to allow increased use without increased cost. Consequently their primary benefits will not be development. Rather, they will be access for more people and goods and/or reduced transportation costs.

1.4.2 Minimizing the External Costs of Transportation

Transportation often imposes costs on people other than the individual traveler or freight carrier. It can damage the environment, delay other travelers, or facilitate development that damages the environment or community character. Reducing these costs has been an important goal of transportation policy ever since environmental concerns gained attention and highway construction slowed in the early 1970s.

1.4.2.1 Environmental Costs

The use of transportation vehicles imposes environmental costs on travelers and non-travelers alike. These costs vary with location and mode. For example, the effects of emissions of pollutants are likely to be minimal in a sparsely populated or windy area. But in a heavily populated area or where the topography and climate favor the formation of ozone, effects can be significant. The effects of spilled fuel or oil leaks depend on how street runoff is treated. If runoff from a heavily used road or gas station drains directly into a small creek, the effects can be devastating, but if it drains into a sewer system where it is treated, the effects will be less serious. However, in all locations, carbon dioxide emissions from the burning of carbon-containing fuel contribute to greenhouse gases in the upper atmosphere.

Worn out or obsolete transportation vehicles and infrastructure, such as old tires, car wrecks, rotting piers, and abandoned railroad tracks can degrade the environment by becoming eyesores or adding to the solid waste burden. The extent to which these are removed from public view and their materials are recycled, determines the extent of degradation.

Other environmental costs include depletion of natural resources and damage to the land and water resources that are caused by the extraction, production and shipment of fuels, manufacture of vehicles, and construction of the infrastructure. These costs depend on the fuel efficiency of vehicles and the design, material intensity, and types of materials used for vehicles and infrastructure. Social costs, such as the splitting of a neighborhood by a freeway or railroad, can also be considered environmental costs.

This definition of environmental costs encompasses the concept of sustainability—that is, meeting the needs of the present without compromising the ability of future generations to meet their own needs. The primary ways in which the current transportation system might compromise this ability is by contributing to global warming or depleting resources, including land available for transportation infrastructure and other uses.

1.4.2.2 Costs Imposed on Other Travelers

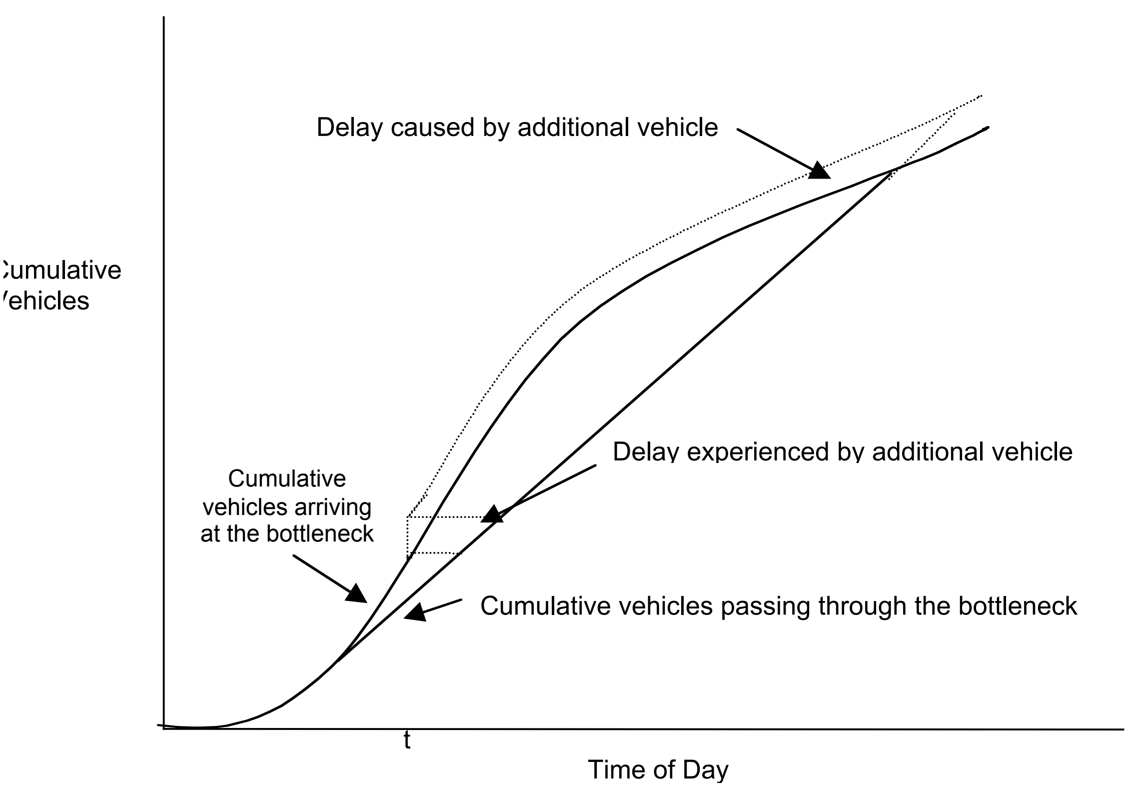
An additional user of an uncrowded road or street will not cause delay to other users. But if the road is crowded, each additional user imposes additional delay on all of the other users. The more crowded the road, the greater this cost. The nature of this cost can be seen in Figure 1. The vertical axis shows the cumulative number of trips during the peak period on a particular transportation link. The horizontal axis shows the time of day. The curved line represents the number of people wanting to pass through the link[4]. At first

[4] Actually this would be a stepped line, one step for each vehicle. However, for simplicity, it is shown here as a smooth line.

the demand exceeds capacity, so a queue develops, then demand decreases below capacity and the queue lessens until it is completely dissipated. The vertical distance between the cumulative vehicle curve and the capacity line represents the number of vehicles in the queue. The horizontal distance represents the delay. The area between represents the total delay experienced by all travelers. The second, dashed curved line represents the cumulative number of trips if an additional traveler joins the queue at time, t. The delay to that traveler is D, but the delay that he imposes on other travelers is the area between the two curves. Clearly, this delay is greater than the delay he experiences. This phenomenon leads to inefficient use of the transportation infrastructure—the delay experienced by the traveler is less than the delay imposed on others. This additional delay not only increases the time costs for other travelers, it increases vehicle-hours, thus increasing carbon monoxide and hydrocarbon emissions. But the trip does not necessarily lead to a net social cost, because the benefit of the trip may exceed the total time costs resulting from the trip. Similarly, a person who uses a crowded bus or train imposes additional crowding on the other passengers. The more other passengers there are, the greater the total discomfort he imposes. As with the automobile traveler, his cost is always less that the total cost he or she imposes on others.

Figure 1

Effect of an Additional Vehicle on A Congested Road



From the graph, one can see that the problem is not really too many travelers, but too many trying to travel at the same time. If there were some kind of reservation system or pricing system, the number of cars arriving at the bottleneck could be made to match the capacity.

1.4.2.3 Costs of Development

To the extent that transportation contributes to development, it imposes costs associated with development. These costs are the destruction of open space or agricultural land, the delay caused by sharing transportation facilities with too many other users, and the separation of economic groups and the alienation that can result. However, the costs of transportation that are considered in this paper will not include inner city decay, the isolation of the poor, or social alienation. Although transportation enables large metropolitan areas, which tend to result in spatial separation by economic level and concentrations of poverty, it does not cause population growth, unequal opportunity, or the unequal distribution of wealth. On the contrary, transportation provides the means to overcome isolation and inequality—poor people suffer not from too much transportation but from too little.

Suburban sprawl will not be considered as a cost. First, because one person's "sprawl" is another's "compact development." Second, because suburban living rather than higher density urban living has been the choice of the majority of Californians, so it is not clear that they consider suburbs a bad thing. The question should not be "Will this transportation improvement increase suburban sprawl?" Rather, the question should be "Will providing better access to this location result in a land use that has higher costs than benefits?" This will never be an easy question to answer because the interests of current residents are often in direct conflict with those of developers, who represent the interests of potential future residents. Data are useful in such cases, but they often do not provide a conclusive answer. Frequently the resolution depends on which group can wield the most political power.

1.4.3 Other Community Goals

Transportation improvements are sometimes made because of other perceived benefits, such as community pride or learning about new technologies. A city may feel that a light rail system adds prestige to the city. Or the goal may be to test or demonstrate a new technology. In the latter case the primary goal would be to provide information regarding the effects of the technology in order to inform decisions regarding future development of the technology and the circumstances in which it would be effectively implemented. If the technology proves useful or shows promise of being useful, this increases the likelihood that is will be used elsewhere. If not, additional ineffective experimentation can be avoided. In this case, the community served would be national or perhaps even worldwide.

1.5 Goals of Public Transportation Agencies

Public agencies are charged with the goal of serving the public. In doing so they face the dilemma of serving all of the above transportation goals, many of which are in conflict, such as providing access while minimizing vehicle emissions. The dilemma is resolved, however, if the agencies take maximizing the net benefits[5] of the system as their goal, given that these are equitably distributed across the population. Weighing benefits and costs provides a mechanism for prioritizing transportation investments and determining when progress toward a certain goal is worth the cost. These benefits and costs would include not only those that can be given a monetary value, but also those that can not be monetized or even quantified.

1.6 Overall Benefits and Costs of the Transportation System

The benefits and costs described earlier in this section are listed below. The overarching goal of the system is to undertake improvements that will most increase net benefits, thus maximizing the net benefits of the system.

Table 1 Overall Benefits and Costs of the Transportation System

Benefits	**Costs**
Activities enabled by transportation	Time
Markets enabled by transportation	Money (public and private monetary costs)
Economic and social development	Property loss and injury
Entertainment	Discomfort
Other	Environmental degradation

1.6.1 Double Counting

Monetary costs of property loss and injury must be included in only one of these categories, not both. Fares, tariffs, tolls, and transportation taxes that are used to offset transportation service provider costs should not be counted as individuals' or shippers' monetary costs unless they are subtracted from the service providers' monetary costs. If double counting is thus eliminated, these categories are mutually exclusive. However, they are not necessarily independent. For example, congestion affects the type and number of accidents and also increases travel time, emissions, and discomfort.

1.6.2 Other Groupings of Benefits and Costs

Benefits and costs could be grouped into different categories than those in Table 1. The classification scheme is not important. What matters is that 1) all significant benefits and costs are considered, 2) neither benefits nor costs are double counted, and 3) benefits can be assessed in terms of their costs. Appendix A provides a comparison of the categories used in this paper and the categories currently proposed for the performance measure module of the 1998 California Transportation Plan.

1.6.3 Effects Not Considered to be Benefits

Transportation jobs and investment costs, although sometimes called economic benefits, are *not* benefits, but rather costs, because the people and other resources used in transportation would otherwise be providing other services to society. Society is giving up those services in order to have the transportation facilities. Even if the people would

[5] Here net benefits are defined as the excess of benefits over costs.

be otherwise unemployed, and the government is paying them to work in transportation, society is giving up the other services that the government might have paid them to provide. The alleged benefits are actually costs that are included in the money costs of transportation that are borne by the providers of the transportation infrastructure and services and by travelers and shippers. The true economic benefits of transportation are the benefits from reduced costs and the development benefits.

1.6.4 The Relationship Between Individual and Overall Net Benefits

An additional trip, while beneficial to the traveler, can reduce overall net benefits when the system is congested. The additional traveler does not make a trip unless his or her perceived benefits exceed perceived costs. However, this trip imposes costs on others. These will be low if the system is not congested, so that the additional trip will result in a net benefit to the population as a whole. However, if the system is very congested, the costs the trip imposes on other travelers may be higher than the benefit of the trip, thus *reducing* net benefits to the population as a whole. For example, if a person in a car with a good emissions control system makes a trip during an uncongested time, the costs to others will be low and the trip will increase overall net benefit. But if traffic is very congested, the trip will impose additional delay on a large number of people, and this cost may exceed the value of the trip to the traveler, reducing overall net benefit. Similarly, an additional person on a crowded train imposes additional crowding on a large number of people, perhaps causing them additional discomfort that outweighs his benefit from the trip. The interplay between individual and overall benefits and costs is shown in Table 2.

Table 2 Individual versus Overall Benefits

	Total Benefit	Costs	Net Benefit
Individual trip Effect on traveler	• Activity enabled by transportation • Entertainment	• Time • Monetary operating costs • Property loss and injury • Discomfort	Positive
Individual trip Effect on other travelers	None	• Time • Property loss and injury • Discomfort	Negative
Individual trip Effect on the community	None	• Environmental degradation	Negative
Individual trip Total effect	• Same as individual benefit	• Sum of cost to the traveler, other travelers, and the community	Depends
Overall Total effects of all trips	• Sum of benefits of individual trips	• Sum of costs of individual trips • Capital and operating costs of the transportation system	Depends

This table explains why people want to get other people out of their cars but do not want to get out of their cars themselves. They recognize the costs these other people impose on them and the benefit they themselves gain from each trip. It also explains why residents of areas with congested traffic oppose development—it imposes net costs on them. Of course, the reverse is true for the people who will ultimately inhabit or work in the new development. The value of the development to the developer derives from its ultimate value to these people. It is possible that added development may impose costs on current residents that exceed the benefits to the new residents. Total net benefits and equity are both issues with new development.

1.6.4.1 Consumer Surplus

Some readers may find the economic concept of consumer surplus useful is seeing the interplay between overall and individual benefits and between benefits and net benefits. Appendix B contains a discussion of the effects on consumer surplus of reducing transportation cost and increasing transportation system use.

1.6.5 Equity

Equity should be a constraint, not a goal. The benefits and costs of transportation should be fairly distributed across the population over the long run. Equity has many dimensions, and there are many long-standing policies designed to achieve equity. One example is the use of a gas tax for funding road improvements—those that benefit, pay. Compensating people whose land is taken for transportation facilities is another example. Sound walls are an attempt to avoid imposing unfair noise burdens on people living adjacent to freeways. Measures taken to reduce noise near airports are designed to keep air travelers from imposing an unfair burden of nearby residents. But inequities do occur. In some areas transit funding has been diverted from relatively cost-effective buses serving lower income urban areas to less efficient rail transit serving more affluent suburban areas. As a result, urban users paid a higher proportion of the cost of their transit service than the suburban users and suffered a loss in service as well. Furthermore, costs were shifted to people less able to pay.

Subsidizing transportation for the disadvantaged will not be considered here as an equity issue because it is not related to the equitable distribution of benefits and costs of transportation. It is a social policy issue, and the relevant question is how to maximize the benefits that can be obtained with a given transportation subsidy. This does not necessarily mean increased transit service, as is often assumed. Greater net benefits might be achieved by providing paratransit or shared-ride taxi service, transportation vouchers that could be used for any type of transportation, by facilitating a car-sharing program, or by providing bicycles.

In some cases transit services to facilitate the movement from welfare to work may not require any subsidy. For example, in a situation where entry level jobs have moved to the suburbs and welfare recipients live in the central city, existing deadhead transit trips back to the suburbs to pick up a second load of suburb to city commuters can be converted into revenue service for these people to commute from the city to the suburbs. This is simply a matter of providing a fair and effective distribution of transit service.

2 Measures for Monitoring System Performance

In monitoring system performance, opportunities for increasing the net benefits of the transportation system are identified.

In a highly developed transportation system like California's, most of the benefits of increased economic and social development due to economies of scale and agglomeration have already been realized, as was noted in Section 1.4.1. And people can usually get where they want to go. Consequently, opportunities to increase benefits to current users of the system are limited, and transportation investments will focus on reducing costs or preventing increased costs in the face of increased use. Performance measures can show where transportation costs are high, that is, where there are problems, such as roads with heavy congestion, sites where there are frequent accidents, and neighborhoods with transit dependent people and infrequent transit service. These problems provide the opportunities for reducing costs and thereby increasing net benefits.

Although street and highway system performance measures provide an excellent means for *identifying potential opportunities to improve* performance, they are not suitable for *judging* performance. This is because the managers of the streets and roads can not limit the level of use. They may use intelligent transportation systems to maximize capacity. They may expand capacity with additional lanes or roads. They may make the most cost-effective improvements possible. But except for a few toll bridges and toll roads, they can not moderate or influence the time of use. Therefore, they can not control the level of delay, because delay is related to the level of use. So performance can not be judged by measuring delay alone. It can not even be judged by delay *and* use. Which road is performing better, the one with more use and more delay or the one with less use and no delay? A congested freeway could be considered a great success because so many people use it and are able to participate in so many activities as a result. Or, it could be considered a failure because of the delay people experience in using it. Conversely, a less used, uncongested road could be considered a success because it is not congested or a failure because it is not well utilized. These performance measures are not appropriate as "report cards."

The issue of performance is simpler in private, for-profit transportation. Here the ultimate indicator of performance is consumer surplus and profit. The market measures performance and sends signals regarding how to improve performance and how much service to provide. Interfering with the market by setting performance standards can actually *reduce* net benefits. For example, before passenger air travel was deregulated, carriers were required to maintain certain levels of service. When the airlines were deregulated, fares dropped and service deteriorated. But the number of passengers increased, showing that they preferred a different mix of service and fares than they had received when the airlines were regulated.

Regardless of how performance measures are used, they draw attention to what is being measured and tend to distort incentives. To reduce the danger of focusing attention on the wrong things, performance measures should be closely aligned with the primary goals of the organizational unit. This suggests having a very limited number of performance measures[6].

Because the collection and management of data for performance measures is expensive, performance measures should not be instituted in situations where they will not be used as the basis for some decision. Furthermore, a performance measure system should not be expected to answer all questions that might arise. Ad-hoc studies may be more appropriate in many situations.

This section has discussed the issues related to monitoring performance. Sections 2.1 to 2.5 describe what should be monitored by various transportation agencies. Section 2.5 lists specific measures and describes how measurements can be made.

2.1 Performance Monitoring for State Departments of Transportation

Measures used by a state department of transportation (DOT) for monitoring performance should be related to its primary goal of providing access to activities and markets and minimizing the cost of such access. It is not possible to directly estimate the benefits of this access, but it is possible to estimate indicators of the number and range of activities accessed. These indicators are the number of person-trips and the distance traveled—person-miles or freight-miles—in other words, the quantity of transportation provided. The more trips made, the more activities accessed. The greater the distance traveled the greater the choice of activities.

Some land use strategies attempt to put activities closer to where people live and work or closer to transit service. However, state departments of transportation typically have no direct influence over land use. They can exert influence only indirectly, through expansion of existing facilities and construction of new facilities. This can facilitate new travel patterns and new development. Departments of transportation generally do not close down or reduce the capacity of facilities, so their actions by their nature tend to increase travel, providing increased access for current travelers and access for new residents and workers.

The primary indicators of the traveler costs over which the state has influence are travel time and accidents. With these performance measures the state DOT can see where additional resources are needed to reduce accidents and to reduce travel time or delay, both the maximum experienced by particular individuals and the total experienced by all travelers. Each of the latter is an important consideration in allocating resources. For example, given two problem locations with equal total person-delay and equal resources required to mitigate delay, the resources would be used where the delay per traveler was highest. A one-minute delay experienced by 6000 people is not as bad as a 60 minute delay experienced by 100 people. The decision is more complicated in cases where the resource requirements are not equal—there a three-way trade-off is required. Individual traveler costs, costs to all travelers, and agency costs (which are borne by the public at large) must be weighed against each other. The state DOTs also have some control over how much public money they spend. This, too is an important measure of performance.

---

[6] As noted earlier, sub-units can establish their own performance measures related to their own particular responsibilities, but these measures should belong to the sub-unit and not be the concern of the larger organization.

In summary, the important performance measures for state DOTs are measures of the amount of transportation provided, travel time/delay, accidents, and public cost. Of course, measures of the amount of transportation provided are needed to put these costs in perspective.

Concentration on these costs for monitoring performance does not mean that a state DOT is not concerned about the environment, equity, or economic development, only that the latter are not its primary reason for being and that it has limited power to affect them. Concern for these goals comes into play when the DOT is determining what actions to take in the pursuit of its primary goals, as will be discussed in Chapter 3. For example, when a section of freeway that collapsed in the Loma Prieta earthquake in 1989 was reconstructed through Oakland, the alignment was changed to address both environmental and equity concerns. Furthermore, economic well being, equality of income distribution, and air quality are all routinely monitored by other agencies, whose primary missions are more closely aligned with these measures.

2.2 Performance Monitoring for Regional and Local Transportation Agencies

These agencies have the same core goals as the state transportation agency. But they may have a different emphasis. For example, in a large metropolitan area, the primary problem is likely to be congestion, whereas in a sparsely populated region, the main problem is more likely to be safety. Not only are the problems they face different, the costs of addressing the problems are different. It costs much less to add a lane to a road in a rural area than in a city. Therefore, the level of congestion that is acceptable in a city is higher, because it costs so much more to reduce it. Similarly, because there are more miles of road per person in a rural area, and therefore fewer resources to devote to maintaining each mile, the level of maintenance that is acceptable to residents may be lower because their maintenance costs per capita are higher. Performance measures should not take the form of standards for all regions or cities because standards do not provide for weighing the benefits against the costs, and there can be no one standard that is appropriate for all situations.

As in the case of the state DOT, there should be a limited number of performance measures, which are closely aligned with agencies' primary goals. Although the local agencies' primary concerns will be the same as the state's, the measures may be different. Because the local street network is denser than the highway network, it will not be possible to measure person- or freight-miles. Simple vehicle counts at key locations will be used. Delay is strongly influenced by the mode of operation of the traffic lights, so spot speeds can not be used to estimate travel times. Only accident and cost measures would be similar to those monitored by state DOTs.

2.3 Performance Monitoring for Public Transit and Passenger Transport Agencies

The Federal Transit Administration requires extensive performance monitoring of transit agencies that receive federal funds[7]. Measures used include costs by category, passenger

[7] This includes most transit agencies.

trips, passenger-miles, vehicle-revenue-miles, and other measures for the transit system as a whole. Transit agencies need the same data on a route basis in order to determine where resources can best be deployed. The same data is needed for publicly operated inter-city passenger rail.

2.4 Performance Monitoring by Ports and Airports

These facilities are interested in the same core values as other state and local transportation agencies: access, time, safety, and cost. Therefore, they should monitor the number of passengers and the tons and value of goods passing through the facility. They should also monitor the person-delay and the $ value of freight-delay that is port-related, as opposed to carrier related. Accidents and costs for capital, maintenance, and operations should also be monitored.

2.5 Private Carriers

The cost-effectiveness of private carriers is ultimately enforced by the market. Those carriers whose costs exceed the benefits they provide will soon be out of business. So there is no need for routine public monitoring to gauge their efficiency. However, these carriers all use or affect public transportation facilities directly or indirectly, so in order to make informed decisions regarding transportation investments public agencies need some idea of the quantity of transportation they provide. The public also has an interest in the costs carriers impose on others and costs that might not be fully captured in their profit calculations. Accidents that damage others or that injure their passengers are one such cost. Vehicle emissions are another, but these are already regulated by environmental agencies.

2.6 Summary List of Performance Measures

The previous sections described the indicators that should be monitored in order to assess performance relative to the primary benefits and costs of the transportation system. This section lists specific measures of these indicators and discusses how they can be measured. As noted earlier, other costs, such as discomfort and environmental degradation will not be routinely monitored but will come into play in decisions regarding actions to take to improve the transportation system. These costs tend to move in the same direction as the primary costs. For example, discomfort is often the result of anxiety about the possibility of delays and accidents, the time spent sitting in a car, or the time spent waiting for a bus. Fuel consumption and emissions of carbon monoxide, hydrocarbons, and carbon dioxide increase with increases in vehicle hours and variation in speed, given no change in the quantity of transportation or vehicle fleet. Private monetary costs tend to move in the same direction as delay.

There are no universal performance measures for all modes. Different modes serve different purposes and operate in different ways. Therefore different measures are appropriate. For example, traveler surveys show that people using transit find the waiting time more onerous than the in-vehicle time. So waiting time is an important measure of performance for transit and other purchased transportation, but there is no waiting time associated with personal transportation such as walking, bicycling, or traveling alone by automobile.

2.6.1 Quantity of Transportation

As noted in Section 2.1, the benefits of access to activities or markets are impossible to measure. Fortunately, these benefits can be inferred from the quantity of transportation consumed. The greater the number of trips and distance traveled, the greater the number and range of activities. With freight, the more ton-miles (or $value-miles) the greater the benefits of trade. The transportation measures shown in the table below are facility-based and do not include non-motorized modes or most local streets. Traveler-based information on travel patterns is needed to provide this information, as well as to explain the observed patterns of facility use.

Table 3 Quantity of Transportation – Indicator of Access

Mode	Measures	Geographic aggregation
• Autos and light trucks	• Person-miles – highway	• Link • Corridor • Regional • Statewide • National
	• Person-volume – streets	• Arterial link
• Transit	• Passenger-trips • Passenger-miles	• Route • System • Regional • Statewide • National
• Inter-city bus and rail • Commercial airlines	• Passenger-trips • Passenger-miles	• Corridor • Statewide • National
• Private airplane	• Passenger- trips	• Airport • Regional • Statewide • National
• Truck freight • Rail freight	• $ value-miles or ton-miles	• Corridor • Statewide • National
• Air freight	• $ value-miles or ton-miles	• Corridor • Statewide • National
	• $ value or tons	• Airport
• Marine freight	• $ value-miles or ton-miles	• Trade route
	• $ value or tons	• Port

Volumes measured at the link or route levels can be aggregated up to corridor, region, state, and national levels.

2.6.1.1 Measuring Transportation via Roads and Streets

2.6.1.1.1 Measurement Locations

Measurements should be made where they might inform some action. For example, there is little need to measure traffic volume on a residential street on a routine basis. However, volumes should be measured on arterials and on state highways. In congested areas, volumes should be measured on every link. In uncongested areas volumes could be measured on fewer, representative links.

2.6.1.1.2 Measurement Schedule

Measurements should be more frequent in locations that are growing rapidly than in those that are growing slowly. If the area is growing slowly or not at all, counts can be made every few years, but they should be made on the same days of the week at the same time of year, when conditions are normal, so that seasonal variations do not confound year to year comparisons. In congested locations, counts are needed for several days at different times of the year to adequately capture the range of travel times and volumes.

2.6.1.1.3 Measurement Methods

The choice of measurement methods for vehicle volumes depends on the frequency of measurement and the existing traffic surveillance infrastructure. For infrequent measurements, a less capital intensive, more labor intensive, method is appropriate, such as use of pneumatic counting devices. The reverse is true for very congested areas with changing traffic patterns. Here, a more automated system with communication links to a transportation management center would be more appropriate. Such a system would also require facilities for pre-processing and transmitting the data to the site where it will be further processed and stored.

Estimates of vehicle occupancy are needed to convert vehicle-miles into person-miles. These usually require field observations, but in some locations, survey data may be available to provide a basis for these estimates. Highway person-miles can be derived from highway link volumes multiplied by link lengths and average vehicle occupancy.

2.6.1.1.4 Data Storage and Processing

In congested areas software that can store and process volume and travel time measurements and calculate distributions for various groupings of links is required. There will be a large number of links and the software must account for the dependence of travel time on one link on travel times on adjacent links. There must also be checks on the accuracy of the reported data and methods for initiating repairs for inaccurate surveillance equipment and for estimating volumes and times until repairs are made. In many Caltrans districts, proprietary systems for data storage and processing are being developed. To date, none is completely tested and operational.

2.6.1.1.5 Organizational Responsibility

Measurements can be made by state agencies or by regional or local agencies. If they are made locally, there should be standard measurement methodologies used by all agencies and state assistance in selecting and maintaining accuracy.

Completeness and accuracy depend on the unit responsible for performance measurement having some control over the means of measurement. The unit should have staff available to install and maintain equipment for data collection, transmission, and storage, as well as to oversee development and maintenance of data processing software. If the performance measurement unit is a secondary user of the data, it should have adequate authority to obtain the data when needed and to insure that the data is accurate. The performance measurement unit should be closely connected with the users of the measures, that is, those who use the measures to identify opportunities for transportation improvements.

2.6.1.2 Statistics on Transportation via Transit

As noted earlier, public transit agencies that receive federal assistance are required to report passengers and passenger-miles each year. The Federal Transit Administration publishes these annually in the National Transit Database. However, because there is some delay between when statistics are reported and published, the state might wish to obtain data directly from the transit agencies.

2.6.1.3 Statistics on Passenger and Freight Transport via Private Carrier

Passenger and Freight Access Rail passenger data can be obtained from Amtrak, rail freight from the Association of American Railroads. Air passenger and freight data can be obtained from the USDOT Bureau of Transportation Statistics, Office of Airline Information or from individual airports. Bus data and private air travel are estimated by the Eno Foundation, Inc. (703) 729-7200. Waterborne freight data is published by the Pacific Maritime Association (415)576-3200 and can also be obtained from the US Army Corps of Engineers. Periodic Commodity Flow Surveys conducted by the Bureau of the Census provide detailed information on the value and origins and destination of freight. The last survey was in 1993.

2.6.2 Time

Travel time is used as the indicator of time cost. Ideally, data on road travel time would be collected by link. Such data allows graphical representation of variations in volume and travel time across time and space. Figures 2 shows travel time contours developed by Caltrans for westbound Highway 8 in San Diego during the AM peak period. Such a representation gives decision-makers and planners a good idea of the geographical and temporal distribution of delay. Figure 3 shows variations in travel time and volume by time of day for a particular highway segment. This could be any link or group of adjacent links. Information could be stored in a database and retrieved for whichever links or groups of links were desired. The three lines on the top half of the graph show the distribution of travel times. The center line is the median travel time. On 95% of the days travel times were less than the top line. On 5% of days travel times were less than the bottom line. Such a representation is useful not only to planners and decision-makers, but also to people planning trips. To see trends in congestion, graphs for the same area covering the same times of year can be compared from year to year. These might be like those shown in Figures 2 and 3 or some simplified version of Figure 3. When this type of representation is coupled with a similar representation of travel volumes, it provides an even more complete picture of delay and the number of people affected. Travel times can be aggregated to the corridor level, but further aggregations do not make sense. The travel time and volume statistics can be converted into delay statistics, which can be aggregated up to the national level.

Figure 2

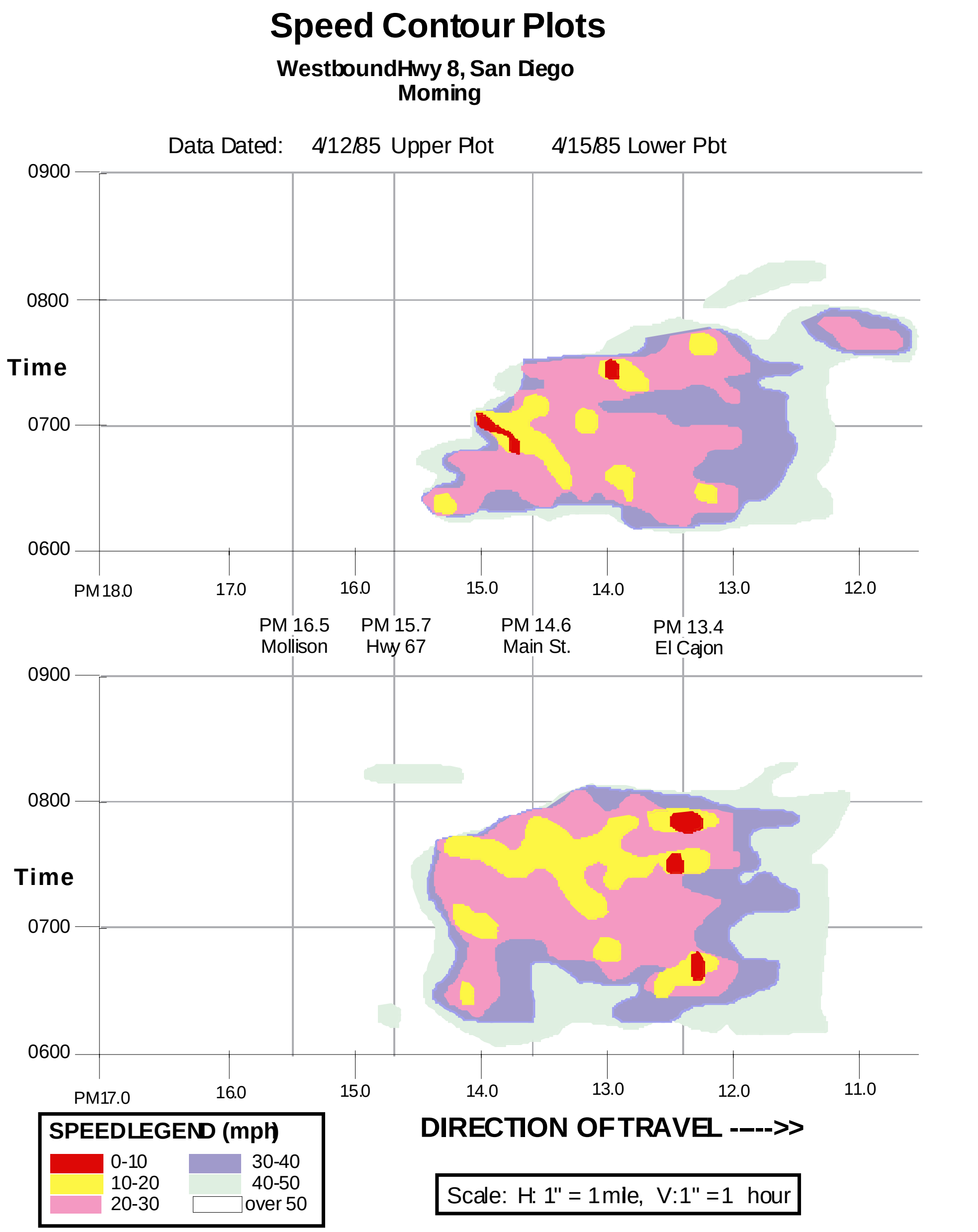
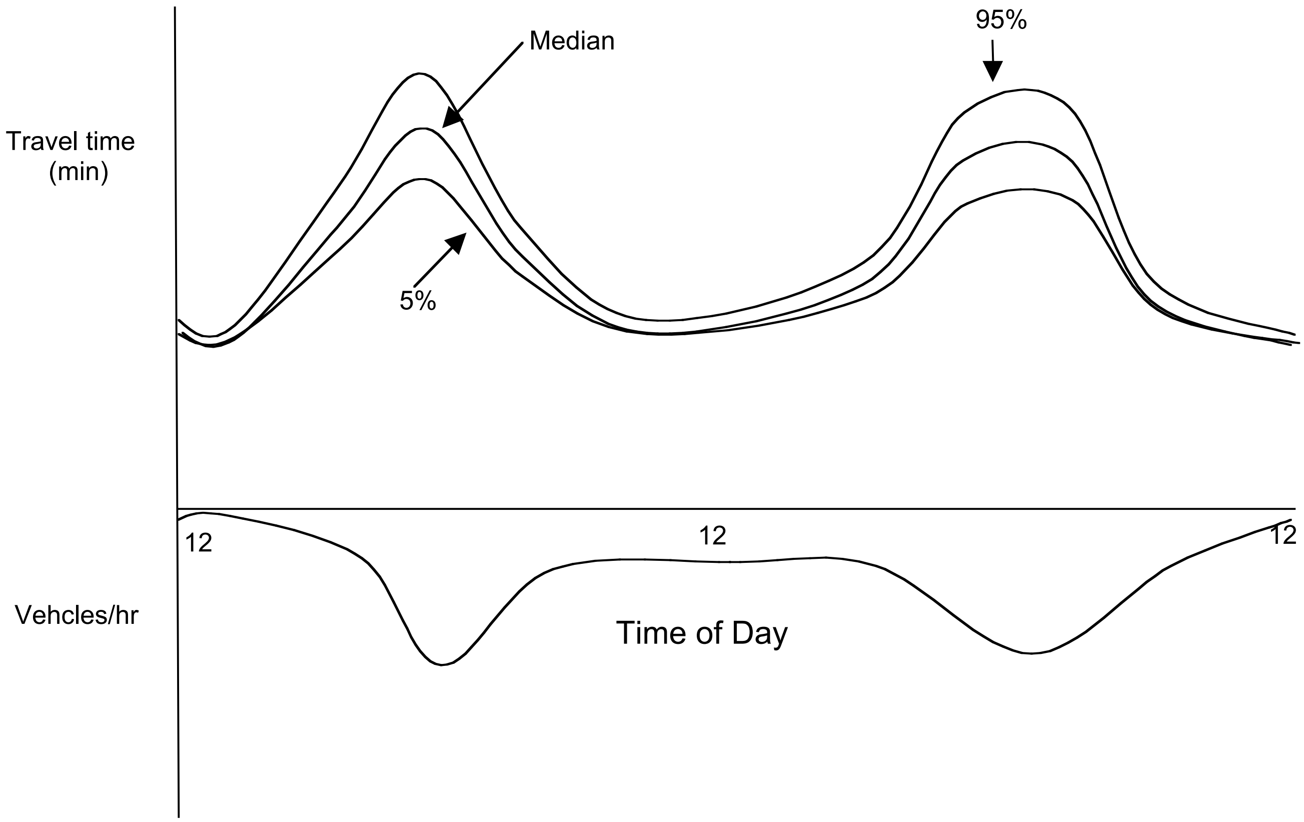

Speed Contour Plots
WestboundHwy 8, San Diego
Morning
Data Dated: 4/12/85 Upper Plot 4/15/85 Lower Plot
0900
0800
Time
0700
0600
PM 18.0
17.0
16.0
15.0
14.0
13.0
12.0
PM 16.5
Mollison
PM 15.7
Hwy 67
PM 14.6
Main St.
PM 13.4
El Cajon
0900
0800
Time
0700
0600
PM17.0
16.0
15.0
14.0
13.0
12.0
11.0
SPEEDLEGEND (mph)
0-10
10-20
20-30
30-40
40-50
over 50
DIRECTION OF TRAVEL ----->>
Scale: H: 1" = 1 mile, V:1" = 1 hour

Figure 3

Weekday Travel Time Distribution and Traffic Volume



2.6.2.1 Travel Time on Streets and Roads

Travel time measurements are needed only on routes that have significant delay at some times. Measurements should be made periodically throughout the year to capture seasonal variations. Because of these seasonal variations, year to year comparisons should be based on the same time of year.

Travel time data can be estimated from loop detector data assuming an average vehicle length, from double loop detectors that capture speed, or from more sophisticated speed estimates using 1-second loop detector data. Floating cars, often equipped with tachographic equipment, are often used to measure travel times. Several cars travel the route departing at 10 or 15 minute headways. Data from these cars provides detailed information about the speeds at different points along the route, across the various links, and for the entire route. This method is more accurate than measures based on spot speeds. However, it is quite labor-intensive. Fortunately, more cost-effective methods of measuring speeds are on the horizon. As the use of automatic vehicle identification for toll or parking payment increases, vehicles themselves can be used as probes to provide travel time information. Vehicles are identified and then re-identified downstream to get travel time. This is already being done in Houston, and there are plans to try this in

California. Additionally, PATH is developing an entire suite of inexpensive detectors specifically optimized to determine travel time for the entire fleet without requiring installation of ID tags in the vehicles. Initial tests of these systems are expected in 1998.

2.6.2.2 Travel Time via Transit

A significant proportion of transit travel time is the time spent waiting. Therefore, headway data, as well as travel time data, is needed. These data can be obtained from transit agency schedules. Average in-service vehicle speed can be computed from transit vehicle hours and vehicle miles in revenue service. This gives an indication of overall system travel time.

2.6.2.3 Airport and Port Delay

The Federal Aviation Administration has data on air carrier delay by cause. No published source for port delay data was found.

2.6.2.4 Traveler-based Travel Time Data

Travel time data based on facility or carrier performance as described above may give a misleading picture because it does not capture the shift of trips from these more crowded facilities to less crowded facilities. This effect is impossible to capture from volume and travel time data, which provide no information about the number of trips. David Jones (1995), using 1990 Census Journey-to-Work data has shown that although travel times on major facilities in the San Francisco Bay Area increased dramatically between 1980 and 1990, travelers, on average, experienced only a marginal increase in travel times. This occurred because a higher proportion of travel was taking place on the less congested roads. Therefore, it is important to see how people adjust to more crowded facilities. Traveler-centered data from the census or travel surveys is needed to supplement the facility-centered travel time data.

Table 4 Travel Time Measures

Mode	Measures	Geographic aggregation
• Autos and trucks	• Travel time distribution by time of day	• Highway link • Arterial • Corridor
	• Average work-trip travel time	• Census subdivisions • City • County • Regional • Statewide • National
• Transit	• Travel time distribution by time of day • Service headway by time of day	• Route
	• Average in-service speed	• System
	• Average work-trip travel time	• Census subdivisions • City • County • Regional • Statewide • National
• Air carriers	• Distribution of delay	• Airport
• Marine freight	• Distribution of delay	• Port

2.6.3 Safety

Although there is interest in tracking the overall safety of the system, it is more helpful in reducing accidents and assaults to be able to pinpoint their exact location so that their sources can be identified and corrected. Therefore, both types of data should be monitored. Clusters of accidents, which might indicate a dangerous condition, should be reported when they are first observed.

2.6.3.1 Highway Accident Data

Data on highway deaths and injuries on all state highways, county roads, and city streets in the state are recorded in the Statewide Integrated Traffic Records System (SWITRS) by the California Highway Patrol (CHP). Property damage only collisions are recorded for CHP jurisdictions but not all cities and counties. Caltrans maintains a freeway accident data base, TASAS, that contain additional information not included in SWITRS

2.6.3.2 Carrier, Airport, and Port Accident Data

The National Transportation Safety Board keeps statistics on air carrier and general aviation accidents. The Federal Railroad Administration keeps statistics on rail accidents. Transit statistics are available from the Federal Transit Administration, and maritime accident statistics are available from the Marine Safety Information System.

Table 5 Safety Measures

Mode	Measures	Geographic aggregation
• Autos and light trucks	• Accidents by type	• Exact location
	• Total accidents • Accidents/VMT • Total fatalities • Fatalities/VMT • Total Injuries • Injuries/VMT • Total property damage • Property damage/VMT	• Corridor • Regional • Statewide • National
• Transit	• Accidents/assaults by type	• Exact location
	• Total accidents/assaults • Accidents & assaults/VMT • Total fatalities • Fatalities/VMT • Total Injuries • Injuries/VMT • Total property damage • Property damage/VMT	• System • National
• Inter-city bus and rail • Commercial airlines	• Accidents/assaults by type	• Exact location
	• Total accidents/assaults • Accidents & assaults/VMT • Total fatalities • Fatalities/VMT • Total Injuries • Injuries/VMT • Total property damage • Property damage/VMT	• Carrier • National
• Private airplane	• Accidents by type	• Exact location
	• Total accidents • Accidents/VMT • Total fatalities • Fatalities/VMT • Total Injuries • Injuries/VMT • Total property damage • Property damage/VMT	• Airport • National
• Truck freight • Rail freight • Air freight • Marine freight (aggregated by trade route rather than nationally)	• Accidents by type	• Exact location
	• Total accidents • Accidents/VMT • Total fatalities • Fatalities/VMT • Total Injuries • Injuries/VMT • Total property damage • Property damage/VMT	• Port or airport • Carrier • National

2.6.4 Public Monetary Costs

Table 6 Public Monetary Costs

Mode	Measures	Level of aggregation
State department of transportation	• Annual capital costs • Annual maintenance costs • Annual operating costs	• District • State
City or regional transportation agency	• Annual capital costs • Annual maintenance costs • Annual operating costs	• City • County • Region
Transit agency	• Annual capital costs • Annual maintenance costs	• Each mode • District
	• Annual operating costs	• Each route • Each mode • District
Airports and marine ports	• Annual capital costs • Annual maintenance costs • Annual operating costs	• Airport or port

All of the above agencies are public agencies with publicly available budgets.

3 Performance Measures for Determining Which Opportunities to Utilize and How to Utilize Them

Once opportunities for reducing transportation costs or increasing benefits have been identified, actions to utilize them must be designed. These actions and their results must be compared to 1) the status quo, 2) other actions in the same place, and 3) other actions in other places. Elements of such comparisons are contained in project study reports, alternatives analyses, and major investment studies. As noted earlier, for such comparisons a more comprehensive set of performance measures is required than for monitoring.

In Section 2 it was argued that only a few measures are needed for monitoring to identify opportunities for increasing net benefits. In contrast, determining which opportunities to exploit and how, requires considering *all* effects. The reason for this is that although the alternatives to be evaluated will presumably address some perceived need of the agency's constituency, their other effects may differ substantially from one alternative to another. For example, to address congestion on a major arterial, an agency may be considering installing centrally controlled signals on the arterial, expanding bus service along the arterial, or widening a parallel street. Although each alternative is intended to reduce delay, each has additional effects. Expanding bus service will increase access for people without cars, centrally controlling signals may reduce congestion on the cross streets, and widening the parallel street will have significant impacts upon the people living on that street. A valid comparison of the three alternatives would include all of the effects of all of the alternatives, in this case not only delay and the public cost of the improvement, but also access for people without cars, corridor-wide congestion, and neighborhood disruption.

Access

Transportation improvements affect access by reducing the various costs of transportation so that people extend their range of activity, or businesses extend the range of markets they can afford to access. This may include labor markets. Businesses often promote transportation improvements in order to increase the size of the labor market on which they can draw. To the extent that businesses require or produce physical inputs and outputs, they can also afford to increase their market access as the costs of transportation are reduced. This increased access increases the quantity of transportation provided. However, often in transportation planning, the quantity of transportation is taken as given. Population and employment projections based on land use policies and economic forecasts are used as a basis for projecting trip patterns. These trip patterns and the projected transportation facilities are used to project modes and routes. In these cases, transportation investments are intended to minimize the travel time given a fixed level of transportation use.

Travel time

Most major road and transit investments are intended to reduce congestion or travel time[8]. Although an improvement may be intended to reduce the maximum delay on a particular section of the transportation system, for purposes of weighing benefits and costs, a broader range of effects over a broader area and time range must be compared. Change in one part of the network can affect congestion elsewhere in the network. For example, eliminating a bottleneck may increase downstream congestion. The distribution of person-delay or vehicle-delay over the course of the day and at each time of day is the most useful measure of effects, because it provides information about both average travel time and reliability of travel time over the course of the day. Furthermore, it can be aggregated into total delay over all travelers and over the entire area affected by the investment.

There may also be travel time benefits from investments intended to increase safety. Preventing the collapse of facilities due to earthquakes prevents long delays due to detours and reduced capacity. In fact, preventing any type of accident reduces delay.

Uncertainty

Intelligent transportation systems are generally intended to reduce travel time. But the greatest effects of traveler information improvements may be a reduction in uncertainty regarding travel time or upcoming road conditions. The importance of reducing uncertainty should not be underestimated. The important thing to a traveler is how he or she feels about the trip. Knowing how long a trip will take, what conditions one may encounter, or how long one will be delayed can mean the difference between a pleasant, relaxed trip and a tense, anxiety ridden trip.

Safety

Investment is sometimes made solely to increase safety. This is the case with the earthquake retrofit of bridges in California. In this case, the cost of the retrofit is weighed against the expected reduction in death, injury, property damage, and loss of access due

[8] This may not be strictly true during the period of earthquake retrofit in California.

to future earthquakes. The latter are the performance measures in this case. Warning and control devices where railroads cross roads and streets are another example of an investment intended to increase safety.

Monetary Cost

When transportation improvements are made, public monetary costs are generally incurred in order to reduce other costs, such as travel time or accidents. Public monetary costs include total capital, operating, and maintenance cost over the life of the project. Sometimes public investment reduces public costs later, as would be the case with repaving and could also be the case with replacement of old structures or equipment.

Private vehicle costs are independent of the system except for small increases in gasoline consumption and wear and tear due to congestion or poorly maintained roads. In most cases they would be insignificant.

Environmental Effects

Almost any transportation improvement disrupts human, animal and plant habitats and consumes land and other resources to some degree. Clearly, widening a road consumes land, energy, gravel, and asphalt and destroys some plant life. But by reducing congestion, it might reduce fuel consumption and emissions. Instituting rail service clearly disrupts habitats and increases emissions from the train or the electrical generating system. Increased bus service also increases emissions. The environmental disruption from implementing intelligent transportation systems is likely to be quite minor—digging holes for signs, trenching to install communications, and transmitting various types of signals to sense traffic and inform travelers. The important thing is that environmental benefits and costs are identified and if found to be significant, are included in the evaluation of alternatives.

Aesthetics

Congestion reduction projects often include other features that do not affect congestion, such as highway landscaping intended to increase the enjoyment of traveling or sound walls and drainage facilities intended to reduce environmental damage. However, sometimes these improvements are undertaken on their own, not as a part of a larger construction project.

Table 7 Possible Effects of Transportation Investment

Costs of Improvement	Benefits of Improvement
Public monetary cost (capital, maintenance, operating)	Increased access
Environmental costs (land, destruction of habitat)	Reduced person-hours of delay and travel time variance
	Reduced cost of freight delay
	Reduced uncertainty regarding travel conditions and delay
	Reduced death, injury, and property damage due to accidents or assault
	Reduced public monetary cost (operations and maintenance)
	Reduced emissions of pollutants and greenhouse gases
	Reduced noise
	Improved aesthetics

It should be no surprise that this table closely matches the table of costs and benefits in Section 1. It is not intended as an exhaustive list of benefits or costs. In each planning situation, the people affected by the investment should be given the opportunity to identify all effects, good and bad. Fortunately, only those costs and benefits that differ between alternatives need be measured. For example, if all alternatives were equally safe, there would be no need for accident estimates.

3.1 Weighing Benefits and Costs

3.1.1 A Simple Concept

The decision about which investments to make ultimately rests on weighing the benefits and costs of the various alternative investments. This is a simple concept—individuals do it many times in the course of each day. There are a number of ways to compare benefits to cost: the ratio of benefits to costs, benefits minus costs, the benefits that can be obtained for a particular cost, and the cost required to achieve a particular benefit. The most appropriate depends on the circumstances. For example, sometimes people try to get the most benefit from a given expenditure of time or money—they pick the movie that is most critically acclaimed. Other times they try to achieve a particular benefit for the minimum expenditure—they buy the lowest priced gasoline. For more complex decisions, such as which car to buy, they weigh all of the good points of each affordable car against the costs and select the one with that provides the greatest value for the money, considering comfort, prestige, performance, gas mileage and maintenance ratings. Decisions by public agencies are the same in principle, but more complex in application.

3.1.2 Choice of Alternatives

Alternatives should be the best application of the various actions to be considered. In the earlier example, the most cost-effective bus service expansion should be compared to the most cost-effective widening of a parallel arterial and to the most cost-effective method of signal control.

Unfortunately, when a particular action is strongly favored by a powerful interest, a straw-man alternative may be selected as the alternative against which to compare the action. For example, a rail system might be compared to a bus system that includes unnecessary high-occupancy vehicle lane construction. Comparisons using straw-man alternatives manipulate rather than inform the decision-making process.

One alternative to be considered is generally the "do-nothing" or "continue with current plans" alternative. This shows whether taking any new action is worthwhile. If no action has greater net benefits than the current state, then no action should be taken. A typical example might be downtown congestion in a big city. It might be nice to have less congestion, but the cost of achieving this state might be more than the congestion reduction would be worth.

3.1.3 Point of View and Equity

Public transportation agencies, whether national, state, or local, should be trying to maximize the overall public good. This requires considering *all* of the significant effects of alternatives and the distribution of these effects as well. Unfair distribution of benefits and costs not only reduces overall welfare, it often is politically fatal. This does not mean that an action that damages particular individuals can not taken—only that some way to adequately compensate them for these damages must be found and applied.

3.1.4 Analytical Approach

In considering alternative transportation improvements, a transportation agency will be trying to achieve some benefit associated with its core mission—to provide access at minimum cost. One approach to the analysis is to consider the largest effects first.

Consider the example in Section 3 in which three alternatives are being considered to reduce congestion on a major arterial:

- Install centrally controlled, coordinated signals on the arterial
- Expand bus service along the arterial
- Widen a parallel street

In this case the largest effects would be the public construction, operation and maintenance costs and the reduction in delay. The reduction in delay from signal coordination could be estimated with a signal optimization model. The reduction in delay from expanding the bus service could be estimated by using a mode choice model to estimate the number of people who would shift to buses and then a queueing analysis to determine how this shift would affect delay on the arterial. The reduction in delay from widening the parallel street could be estimated with a queueing analysis.

If these analyses showed that traffic signal control performed better than widening the parallel street in terms of delay reduction and costs of construction, operations, and maintenance, there would be no need to further consider the street widening. However, if widening the parallel street were better than the other alternatives in terms of travel time savings and cost, it would be necessary to get some idea of the cost of the neighborhood disruption. Although this could not be directly quantified, it might be possible to ask residents how much they would have to be paid in compensation in order to be indifferent between no widening and widening with payment.

If the traffic signal control performed better than the bus service in the initial comparison, there is still a chance that the additional mobility plus the congestion reduction provided by the bus might make the bus service a better choice. Here the reduced delay for transit users and the resulting new trips for transit dependent people are additional benefits that must be compared to the superior performance of the signal coordination in terms of congestion reduction and cost.

3.1.5 Estimating Effects

Benefit cost analysis is a well-established tool for evaluating public investment. There is a rich literature on the subject including basic principles, specific examples, and practical guides. Therefore, the following section will discuss general issues but will not provide detailed guidance regarding benefit cost analysis, only noting sources that provide such guidance.

The most challenging task in forecasting the performance of proposed transportation improvements is determining how travelers will respond to them. Although it is not necessarily easy to accurately forecast the effects of a new freeway lane, because travelers adjust their travel behavior in response to travel conditions, it is even more difficult to forecast how many people will use a new transit system. Forecasting the effects of information on travel behavior is even more challenging, because there has been relatively little experience in providing route information. There is a continuing need for evaluation of the effects of newly implemented systems to provide information for developing better techniques for forecasting the effects of various types of investment.

3.1.5.1 Values of Effects

In comparing alternatives it will likely be necessary to assign monetary values to travel times, lives saved, and less tangible benefits, such as reduced uncertainty and improved aesthetics. Values have been estimated for the first two, but not for the latter two. People can be asked what they would pay for these or how they value them relative to something else of known value.

3.1.5.2 Period of Analysis

The period of analysis should be long enough to capture differences in the mix of initial capital costs versus on-going operating or maintenance costs. For example, this might be the length of time before major replacement of capital facilities. However, the period of analysis should not be too long. Effects far in the future are much less certain and are less significant due to discounting.

3.1.5.3 Geographic Scope of Analysis

The geographic scope should be large enough to capture the significant effects. Often single road improvements have effects throughout the corridor, and corridor improvements have effects throughout the region

3.1.5.4 Discounting Future Effects

Both benefits and costs should be discounted to account for the ability of resources to produce wealth over time. This is an easy concept to grasp in terms of money. If one spends money now rather than in five years one loses the interest that would have been paid on that money as a result of its productive value. So a cost five years from now is less than a cost today. Similarly, a person's time has productive value. Time lost now results in the loss of the wealth that could have been generated in that time. The correct discount rate is the true interest rate without the risk and inflation components.

Effects may be expressed as the discounted net present value over the period of analysis. In this case the stream of effects is projected over the period of analysis, discounted, and summed. This is the best method if effects other than monetary costs are expected to change over the course of the period. If effects other than monetary costs are expected to be constant from year to year, the equal annual monetary cost (or annualized monetary cost) that would result in the discounted net present value over the period of analysis can be compared to the other annual costs.

3.1.5.5 Uncertainty

Rarely do transportation agencies take uncertainty into account in weighing benefits and costs. This is a major limitation on the validity of such analyses. There is always uncertainty regarding future conditions and future values, and there is additional uncertainty due to error in predicting the effects of investments. Travel behavior depends upon such unpredictable factors as the state of the economy, the price of gasoline, and regional demographics. An alternative with more certain but lower expected benefits may be preferred to an alternative with higher expected, but less certain, benefits. Software that accounts for uncertainty in benefits and costs has been developed, but uncertainty can be taken into account even without such tools.

3.1.5.6 Additional Information on Benefit Cost Analysis

The Federal Highway Administration has been developing practical materials for practitioners in transportation. A computer program for benefit cost analysis called STEAM, has just been developed. The Texas Transportation Institute has developed PC-based software, MicroBencost, for simple benefit cost analysis. Other, proprietary software, has been developed that allows consideration of uncertainty in the analysis. Caltrans has a PC-based computer program for benefit cost analysis of highway improvements that is currently being upgraded to include effects of other improvements.

PATH is developing a framework for benefit cost analysis intended for use by Caltrans district staff and regional and local transportation planning. This is scheduled for completion in 1998. It will be tested during FY1998-99.

4 Performance Measures for Evaluation of Implemented Projects

This paper has focused on performance measures for assessment and planning only. These are the primary functions of public transportation agencies—to assess the system in order to identify problems and to develop plans to mitigate the problems. Performance measures for assessment and planning directly inform decisions made by these agencies. Evaluation, on the other hand, affects their decisions only indirectly. Few transportation improvements are such that they can be discontinued if found not to be cost effective, but of course, the agency may decide not to invest in similar improvements in the future. Evaluations are of value primarily to other agencies that may learn from the experience of the implementing agency. Because of this, and because an unfavorable evaluation may harm the reputation of the implementing agency and its employees, evaluations should generally be conducted by an outside organization, particularly one whose future business prospects would not be damaged by producing a negative evaluation.

Evaluation of implemented projects is the most complex and difficult type of performance measurement. A good evaluation will show not only the outcome but also how that outcome came about. This requires additional, internal performance measures. However, if done well, such evaluations can also be the most valuable because they provide basic information about the effectiveness of various strategies. This imformation is useful in improving future implementations as well as projecting their effectiveness. As with planning, evaluation requires consideration of *all* effects, whether intended or not. Because evaluation is so complex and because it is generally undertaken by an outside evaluator, details of how it is done will not be included in this paper.

5 Performance Measures and Intelligent Transportation Systems

Intelligent transportation systems (ITS) are both a subject of performance measurement and a means of performance measurement

5.1 ITS as a Means of Performance Measurement

Advanced traffic surveillance systems will be required for the type of performance measurements recommended in this report and for measuring the outcomes proposed for the California Transportation Plan. An often overlooked benefit of ITS is the information it can provide for transportation planning—that is determining the condition of the system, forecasting the effects of actions that might be taken to improve the system, and selecting a course of action.

Loop detectors remain the dominant surveillance devices despite their high cost and maintenance requirements. Although no other technology has replaced loop detectors, ITS promises to produce surveillance methods that are superior in terms of accuracy, reliability and cost. Many are already on the market and many others are under development. Improved communications promise to lower the cost of transmitting the raw data, and more powerful computers will make it possible to quickly manipulate large volumes of data, store them for efficient access, and produce graphics that allow people to quickly visualize performance. The Internet will allow access to this information to a wide range of people. The various surveillance devices are described in Appendix C.

5.2 ITS as a Subject of Performance Measurement

One reason for the current interest in benefits and costs is the implementation of new intelligent transportation systems. The effects of many components of these systems are qualitatively different from the benefits of traditional transportation improvements. The potential benefits most often cited are conventional benefits—reduced delay and accidents. But the less conventional benefits of ITS may be equally significant. Primary among these is reduced discomfort due to increased knowledge of what to expect when traveling. This may result from the popular Internet displays of freeway travel times, more convenient means of accessing transit information, radio messages that advise of chain requirements, highway signs advising of route closures, and route guidance devices in rental cars. Discomfort may also be reduced by knowing that help will always be provided when needed. This includes the services of freeway service patrols as well as mayday systems that ask the vehicle occupants if they are alright when there is a crash and send an ambulance if there is no response. Information provided by ITS also appears to have an entertainment element, similar to that from reading the paper or watching the TV news—people just like to know what is going on. Entertainment may become an important element of ITS if automated vehicles free people from the necessity of paying attention to driving. But we can not now know what benefits an automated vehicle system might bring. Ultimately, they may allow people to do new things in new ways that we can not imagine now, leading to increased economic and social development. Table 8 shows the benefits that are likely to be realized and the costs that are likely to be reduced by ITS. A plus indicates an increase in benefits or a reduction in costs.

Table 8 Effects of ITS

	Access And Develop ment	Entertain ment	Time	Money (public agency costs)	Property damage and injury	Discomfort	Environ-mental Degrad-ation
Advanced signal control			++	+	+	+	++
Ramp metering			+		+	+	?
Incident Management			++		+	++	+
Electronic toll collection			+	++		+	
Automated fare payment			+	++		+	
Automated vehicle location			+			+	
Advanced transit management			+	+	+	+	+
Pre-trip Information		+	+			+	
En-route information		+	+			+	
Route guidance			+			++	
Emissions and environmental hazard sensing							+++
Commercial vehicle regulation			+	+			
Mayday support					+	+	
Driver safety Assistance					++	++	
Longitudinal/ lateral vehicle control					++	++	
Automated vehicles	++	++	++		++	++	?

5.3 Estimating Traditional Effects

Almost all ITS strategies save time. However, many strategies are relatively new, and methods of estimating how much time is saved are still crude. Much work is needed to determine the travel time savings resulting from ITS strategies that have been implemented and to use this to develop methods for forecasting travel times savings due to ITS. Similar work is needed with respect to property damage and injury due to accidents and assaults that might be reduced as a result of ITS and on the effects of ITS on public costs and vehicle emissions.

5.4 Estimating Non-traditional Benefits

Table 8 shows that most ITS strategies reduce discomfort in some way and the some strategies provide an element of entertainment. They make traffic smoother, they reduce anxiety about having correct change or having an accident, they satisfy curiosity, they amuse travelers, or they reduce uncertainty regarding routing, travel time, or road conditions. These are the types of benefits that categories such as "consumer satisfaction" might capture. Although it is possible to measure the number of people receiving these ITS benefits, placing a value on them is another story. The situation is somewhat analogous to the clock on a car radio, except that with radio one can determine the value of the clock by seeing the price differential between the models with and without the clock. One way to estimate the value of traveler information is to ask travelers what they would be willing to pay for such services or to ask them how this service ranks compared to various services for which they do pay. Although establishing the value of these benefits may be difficult, they should not be disregarded. After all, a lot of car radios have clocks.

6 Recommendations

6.1 Recommendations for a Statewide Multimodal Performance Monitoring Program

A statewide performance measurement program will be adopted as an element of the 1998 California Transportation Plan. The purpose of the program should be to reveal opportunities for increasing the net benefits of the transportation system.

The following criteria were used in recommending measures for routine monitoring of performance:

- Usefulness in identifying opportunities for increasing the net benefits of the transportation system through public investment in transportation improvements or changes in the management of the transportation system
- Low cost for necessary measurement accuracy
- Low potential for distorting incentives

6.1.1 Recommended Measures

The following recommended measures indicate access, travel time, property loss and injuries, and public cost. To some extent they indicate traveler stress, which is related to speed and accidents, and vehicle emissions, which are related to congestion. Please refer to Section 2 for more detail on their measurement and data sources.

- **Access -- indicated by the quantity of transportation**
  - Autos and light trucks—person-miles
  - Transit and inter-city passenger service—passenger-trips and passenger-miles
  - Freight—ton-miles or $ value-miles by mode and route
- **Travel time**
  - Streets and roads—travel time distribution by time of day
  - Transit—route travel time distribution by time of day, route headway distribution by time of day, average speed
  - Travelers—average work trip travel time
  - Air and marine freight—distribution of port-induced delay

- **Safety**
  - Streets and roads—accidents and assaults and their consequences
  - Transit—accidents and assaults and their consequences
  - Commercial transportation providers—accidents and accident consequences
  - Airports and ports—accidents and accident consequences
- **Public Monetary Costs**
  - Transportation agency—capital, operating, and maintenance costs by function

6.2 Recommendations for Effects to Be Considered in Allocating Public Funds for Competing Uses

In evaluating projects and allocating and programming public funds, local agencies, regional agencies, and the state should consider the following effects:

- **Costs**
  - Public monetary costs of the project
  - Environmental costs of the project, such as destruction of wetlands
  - Social costs of the project, such as dividing a neighborhood
- **Potential benefits**
  - Changes in the distribution of travel time and delay
  - Change in the number and types of accidents and assaults or their consequences
  - Changes in emissions of pollutants and greenhouse gases
  - Changes in aesthetics
  - Changes in noise levels
  - Changes in traveler convenience
  - Changes in freight or traveler use of the system (access)
  - Changes in uncertainty regarding travel conditions or best route
- **Other effects identified as important by the people affected by the project**

Agencies should follow the guidelines in Section 3.1 for weighing the benefits and costs of projects. Additional guidelines are currently under development by PATH and should be available for use in 1998.

6.3 Recommendations for Development of Data Sources

6.3.1 Recommendations for Measuring Volumes and Travel Times

Reasonably good sources and methods are available for measuring public monetary costs and accidents. However, the measurement of street and road volumes and travel times is expensive and incomplete. Furthermore, several surveys have shown that people are as concerned about travel time variability as about absolute travel time. Therefore, average travel time is not sufficient. There must be information about the variation in travel time as well. Systems to measure travel times and volumes over the course of the day over a large number of days are much needed for congested streets and roads. The resulting distribution of travel times and volumes can be used to advise travelers, assess the overall impact of the delay, and to understand the sources of traffic volumes that cause the delay.

The City of Los Angeles and some Caltrans districts in Southern California have extensive systems of loop detectors and communication facilities for transmitting the loop detector data to transportation management centers. However, even in these areas systems for storing and processing the data into volume distributions have not been developed. Furthermore, maintaining the loop detector controllers is expensive, and in some areas large numbers of detectors are not functioning properly, resulting in missing or inaccurate data, thus compromising the usefulness of the data. In some areas, there are double loops for measuring speeds, from which link travel times can be estimated. But speed can vary over the link, so the speed at one point may not provide a good basis for estimating link travel time. Another problem with loop detectors is the high cost of installation and maintenance.

Many new types of detectors have been developed. These are described in Appendix C. However, none have replaced loop detectors as the primary surveillance device. But development continues, and it is likely that a more reliable and cost-effective substitute will be developed. Methods for estimating link travel time based on vehicle profiles or patterns of vehicle groups are also under development. Another development on the horizon is automatic vehicle identification for use on toll facilities, and perhaps even for parking. Automated vehicle identification presents the opportunity to use identified vehicles as probes to provide travel time information. This type of system is already operating in Houston. There are 100,000 vehicles with identification systems that are used on the toll roads. A fairly dense array of readers records the times when vehicles pass the readers. This information is communicated to a system that matches vehicle identifications and calculates the travel time between readers. Unfortunately, the readers are still quite expensive. This type of system has promise if the cost of readers can be brought down.

6.3.1.1 Continued Research on Surveillance Methods

Because there is no completely satisfactory method for collecting travel time and volume data, and because there is considerable development of surveillance methods underway, continuing research and testing of surveillance devices and travel time measurement methods is recommended.

6.3.1.2 Development of a Prototype Travel Time and Volume Measurement System

This need not wait until the ideal surveillance system is has emerged. It can be developed using loop detector data in a TMC that already has wide loop detector deployment. The system would monitor detector accuracy, trigger repairs, and adjust for inaccuracy until repairs are made. It would utilize the available data to estimate travel time. It would process the data to provide distributions of travel times and volumes over the course of the day, which could be aggregated to any section of road. Finally, it would develop methods to characterize performance and communicate it to specific audiences. Varaiya (1997) has suggested such a system and described its use.

6.3.2 Recommendations for Obtaining Other Data

Data sources were identified in Section 2. However, collection costs could be reduced by obtaining as much data as possible from other data collection agencies or organizations. It is recommended that Caltrans consult with the Bureau of Transportation Statistics

(BTS) and arrange to use as much of the BTS data as possible. The expertise of the BTS should be tapped in developing a data collection program.

6.3.3 Development of Data Collection and Management Plans

In Section 2 the need for different schedules and intensity of data collection, measurement methods, and processing methods for different situations was discussed. Yet consistent data is needed for making statewide decisions. To the extent that Caltrans will be collecting data directly, it should develop guidelines for data collection schedules and sites, data collection methods, data format, data processing, and organizational placement of data activities and responsibilities. Specially trained Caltrans staff should be made available to regional and local agencies to assist in setting up appropriate data collection and processing systems that provide consistent data for statewide decision making. The purpose would be to develop the lowest cost method of providing the data to support local as well as statewide needs.

6.4 Recommendations Regarding Performance Standards and Comparisons

Statewide or regional performance standards are not recommended. There is a tendency to make the uniform standard the goal, even in circumstances where a higher standard would be more appropriate. Standards provide no mechanism for weighing benefits against costs in order to identify those improvements that provide the greatest benefit for a given cost.

Performance monitoring should not be used to produce a report card for the system or for the agencies managing the system, because key determinants of system performance are outside the managers' control, and actions required to maintain a particular performance level may not be cost-effective. For example, a key determinant of travel time on a facility is the number of people using the facility. The agency managing the facility can not control the level of use[9]. The agency might wish to expand the facility, but the cost, especially if the facility were located in a congested city, might exceed the value of the resultant travel time savings. Only the relationships between benefits and costs should be compared, to determine where and for what types of improvements the net benefits will be greatest.

[9] A toll facility would be an exception.

Bibliography

General

_____1995) *Advanced Transportation Systems Program Plan,* California Department of Transportation, New Technology and Research Program

_____(1996) *Advanced Transportation Systems Program Plan: 1996 Update: The Framework for a California Partnership,* California Department of Transportation, New Technology and Research Program

Performance Measures

_____(1997) "Minnesota Department of Transportation Families of Measures", Presentation at conference on "Performance Measures for California Transportation System Users and Investors" October 6-7, 1997, Sacramento, California

An exhaustive and exhausting list of measures of transportation system, agency, and employee performance

_____(1997) *Performance Measures for California Transportation System Users and Investors* Compendium of material presented at the conference on "Performance Measures for California Transportation System Users and Investors" October 6-7, 1997, Sacramento, California

An overview of performance measurement systems in use or under design throughout the country

_____(1997) "Working Draft, Issue Paper #1, California Transportation Plan, Transportation System Performance Measures—System Performance Measures Project Discussion", Transportation System Information Program, California Department of Transportation, Sacramento, California

Banks, J.H. and Kelly, G. (1997) *Traffic Management Systems Performance Measurement: Final Report"* California PATH Program, Institute of Transportation Studies, University of California, Berkeley, California

Bloomberg, L., Throckmorton, J., and Klim, T. "Development and Application of the Portland Traffic System Performance Evaluation (TSPE) System" Preprint of paper presented at the Transportation Research Board 76th Annual Meeting, January 12-16, 1997, Washington, D.C.

Codd, N. and Walton, C.M. (1995) *Performance Measures and a Framework for Decision-making Under the National Transportation System",* Center for Transportation Research, University of Texas, Austin, Texas

A comprehensive report on performance measurement issues. Suggests measures, perhaps an unnecessarily large number, and discusses data sources and decision making frameworks.

Czerniak, R., Gaiser, S., and Gerard, D. *The Use of Intermodal Performance Measures by State Departments of Transportation,* Prepared for Office of Environment and Planning, Federal Highway Administration and New Mexico State Highway and Transporation Department, Washington D.C.

A survey of performance measures used by various states on intermodal freight and passenger movement

Ewing, Reid (1995) "Measuring Transportation Performance," *Transportation Quarterly,* Volume 49, Number 1, Eno Transportation Foundation, Inc. Landsdowne, Virginia

Jones, D.W., (1995) *Intermodal Performance Measures for the Bay Area Transportation System,* Report for the Metropolitan Transportation Commission, Oakland, California

A strong argument for user-based performance measures

Poister, T. H. (1997) *Performance Measurement in State Departments of Transportation, Synthesis of Highway Practice 238* National Cooperative Highway Research Program, Transportation Research Board, Washington D.C.

A useful overview of current practice including many examples from various states

Rose, D.C. (1997) "Making Performance Measures Work for You" Presentation at conference on "Performance Measures for California Transportation System Users and Investors" October 6-7, 1997, Sacramento, California

Talvitie, A. (1997) "Performance Indicators for the Road Sector Summary of OECD Scientific Expert Group Report" Paper presented at the Transportation Research Board 76th Annual Meeting, January 12-16, 1997, Washington, D.C.

Varaiya, P. (1997) "How to Measure Transportation System Performance" unpublished

Benefit Cost Analysis and Corridor Planning

______(1995) *Evaluation of Transportation Alternatives, Least-Cost Planning: Principles, Applications and Issues* ECONorthwest and Parsons, Brinkerhoff, Quade, and Douglas under contract to COMSIS Corporation for FHWA, Washington DC

A thorough discussion of the history and principles behind least cost planning

______(1995) *Major Investment Studies (MIS), Case Study Examples of Corridor Planning* COMSIS Corporation and Parsons, Brinkerhoff, Quade, and Douglas for FHWA, Washington DC

Three case studies of major investment studies

Allen, B., Baumel, C.P., Forkenbrock, D.J., and Otto, D. (1993) *Guide to the Economic Evaluation of Highway Projects* Iowa Department of Transportation
De Corla-Souza, P., "Using Benefit-Cost Analysis to Evaluate the Distribution of Highway Project Benefits" Paper presented at the Transportation Research Board 76th Annual Meeting, January 12-16, 1997, Washington, D.C.
Considers geographic and traveler group distribution of benefits

De Corla-Souza, P., Everett, J., Gardner, B., and Culp, M. (1997) " Total Cost Analysis: An Alternative to Benefit-Cost Analysis in Evaluating Transportation Alternatives" *Transportation,* 24, pp 107-123
An explanation of total cost analysis with example

De Corla-Souza, P., Gardner, B., Culp, M., Everett, J., Ngo, C., and Hunt, J. (1997) " Estimating the Costs and Benefits of Transportation Corridor Alternatives" Paper presented at the Transportation Research Board 76th Annual Meeting, January 12-16, 1997, Washington, D.C.
A brief step by step guide to benefits cost analysis for corridor alternatives

Lewis, D.L. (1991) *Primer on Transportation, Productivity and Economic Development,* National Cooperative Highway Research Program Report 342, Transportation Research Board, Washington D.C.
Primer on the practice of benefit cost analysis for transportation improvements

Federal Highway Administration, Washington D.C.

Weisbrod, G. and Treyz, F. (1997) "New Methods for Measuring Highway Economic Benefits—The Elements of Business Productivity" Preprint of paper presented at Transportation Research Board 76th Annual Meeting, January 12-16, 1997, Washington, D.C.
Good discussion of economic effects of transportation

Weisbrod, G. and Weisbrod, B (1997) "Assessing the Economic Impact of Transportation Projects: How to Match the Appropriate Technique to Your Project" Paper presented at the Transportation Research Board 76th Annual Meeting, January 12-16, 1997, Washington, D.C.
A sensible guide to estimating economic impacts

Data Sources

_____(1996) *Data Integration Project Catalog of the Bay Area Partnership,* Metropolitan Transportation Commission, Oakland, CA
A comprehensive list and description of currently available performance data on the San Francisco Bay Area transportation system.

_____(1997) *National Transportation Statistics* Bureau of Transportation Statistics, US Department of Transportation, Washington, DC

Appendix A

Comparison of Benefits and Costs Measures with Outcomes Identified in the Draft Report on Transportation System Performance Measures for the California Transportation Plan

Each of the Key Outcomes contained in the Draft Report fits under one or more of the benefits or costs identified in Section 1. The advantage of the benefit and cost approach is that the measures clearly indicate one or the other. For example, an increase in activities enabled by transportation represents an increase in benefits. However, if it results in more congestion, there will be an associated time cost. The *net benefit* will depend on the relative magnitude of the increased benefits and costs. Note that it is possible for the overall net benefits to increase at the same time that individual net benefits are decreasing. This would occur when the benefits obtained by the activity enabled by an additional trip exceeded the delay and other costs that trip imposed on the traveler and others.

Table A1 Comparison of Benefits and Costs with Outcomes in the Draft Report on Transportation System Performance Measures for the California Transportation Plan

	Activities Enabled by Transportation	Trip Enjoyment	Time	Money	Accidents	Discomfort	Environmental Degradation
Mobilitiy/ Accessibility	X		X				
Reliability			X				
Cost-effectiveness	X	X	X	X	X	X	X
Customer satisfaction	X	X	X		X	X	
Economic well-being	X		X	X	X		
Sustainability							X
Environmental quality							X
Safety and security					X		
Equity	X	X	X	X	X	X	X

Although the California Transportation Plan (CTP) outcomes are included in the benefit/cost structure used in this paper, there are important differences in the two approaches. The proposed performance measure for Mobility/accessibility in the (CTP) is travel time, which is a *cost,* not a benefit. The CTP has no measure of transportation benefits other than economic-well being and customer satisfaction, neither of which can be determined with any objective measure. Moreover, the first is the result of many factors other than transportation, and the later depends on expectations rather than performance.

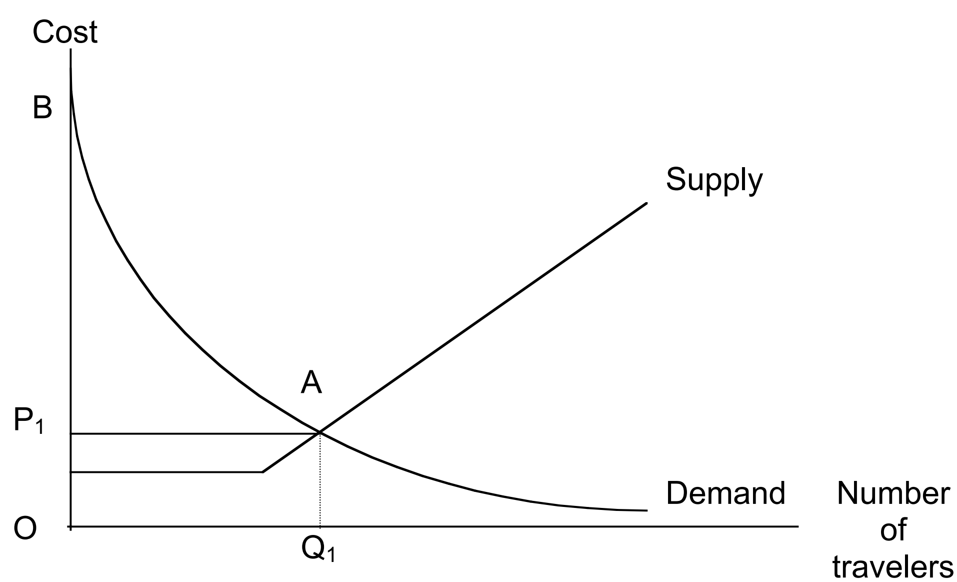
Appendix B

Consumer Surplus

A brief review of the economic concept of consumer surplus may be helpful in thinking about the benefits and costs that accrue to individuals and the society as a whole.

Figure A1 shows two curves: a demand curve and a supply curve. The vertical axis on the demand curve represents the benefits that individuals receive from using a particular transportation facility at a particular time. These benefits are primarily derived from the activity the transportation allows them to access. The benefits of all individuals are summed from left to right, with those of the individuals' receiving the greatest benefit on the left. People will use the facility if the cost does not exceed the benefit. The higher the cost (shown on the vertical axis), the fewer the users (shown on the horizontal axis).

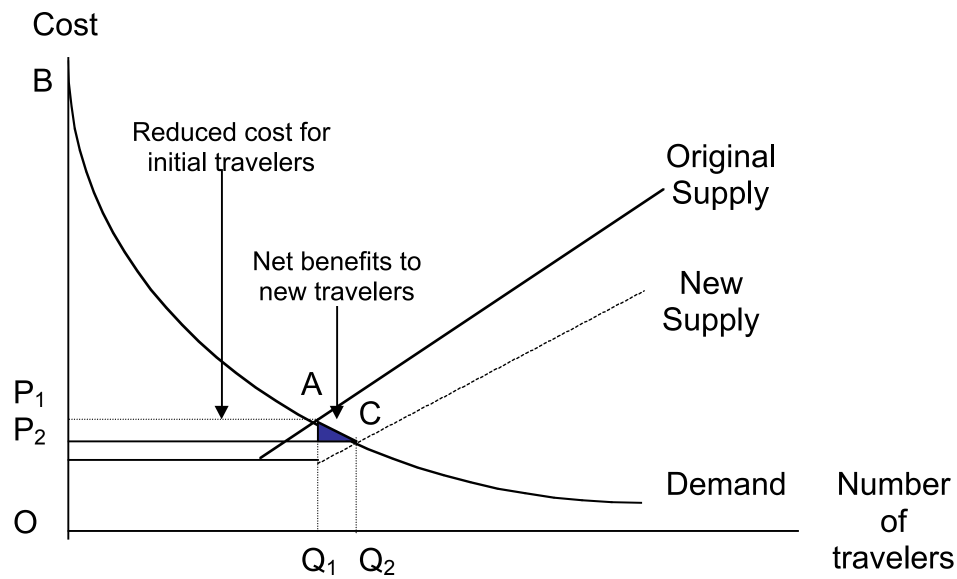
Figure B1 Supply and Demand for Trips as a Function of Time Cost



The supply curve indicates that there is a fixed cost to the traveler up to the point where the curve bends. These costs would include time to make the trip, discomfort en route, monetary cost—all of the costs borne by the traveler. These costs would be based on free flow travel time on a freeway or scheduled time on a bus. To the right of this point, additional users cause additional costs—a queue develops on the freeway, or the bus becomes unpleasantly crowded and slows due to increased loading time. Given this supply and demand shown above, there will be Q_1 users, each paying a cost of P_1. The cost to the Q_1th user exactly equals the benefit he receives. People represented by the portion of the curve to the right will not use the facility or service because its cost to them is greater than the benefit they would receive. However, the people represented by the demand curve to the left of Q_1 all receive benefits in excess of their cost. These benefits are represented by the area under the demand curve, $OBAQ_1$. The area under the demand curve but above P_1 — BP_1A — represents the excess of benefits over costs, known as consumer surplus. This is the overall net benefit to users. (For simplicity, this representation does not include the costs or benefits to society at large or to travelers on the facility at other times or travelers on other facilities.)

The situation is similar with the shipment of goods. One shipper may be willing to pay more for transportation because his production costs are lower and he can make a profit even with higher shipping costs. The excess of demand over supply becomes some combination of profit to the shipper and consumer surplus for consumers.[10]

Figure B2 How the Number of Trips Responds to a Change in Supply



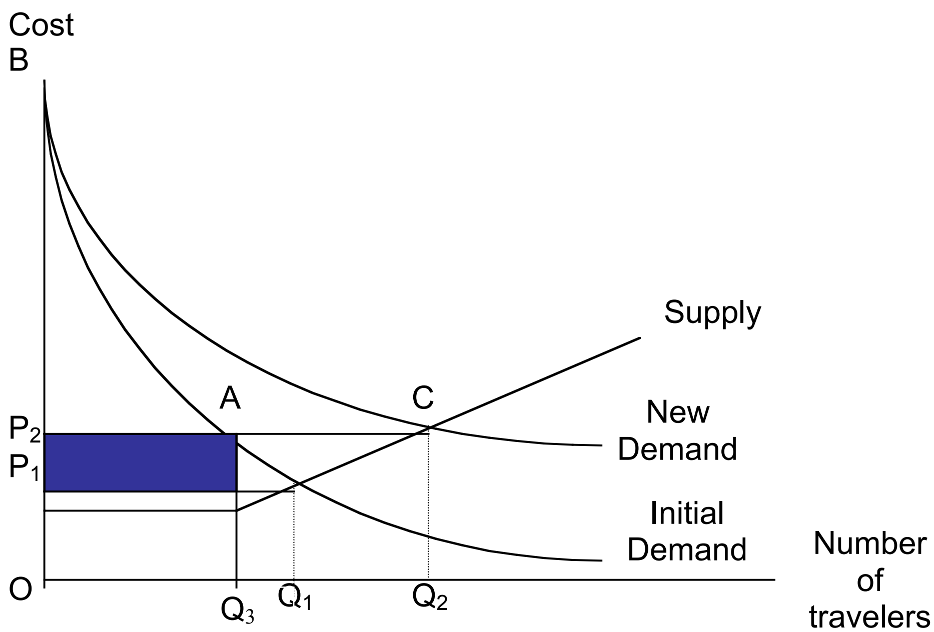
If capacity is increased so that more people can use the facility before they begin to impose costs on each other, the supply curve shifts to the right. The number of users increases to Q_2, and the cost to users falls to P_2, as shown in Figure 2. The initial users' *total* benefits do not increase, but their costs are reduced, resulting in an increase in their *net* benefits, the lighter shaded area. Overall benefits are increased by the benefits to the new users—Q_1ACQ_2. Notice that their individual net benefits are less than the difference between the current and previous costs and that their total net benefits are represented by the darkly shaded triangle between A and C with its base on the line representing P_2. Consumer surplus, overall net benefit, has increased by P_1ACP_2.

---

[10] This is an oversimplification in the case of purchased transportation of freight. Suppliers of transportation services generally charge different prices depending on the value of the shipment. For example, freight charges are generally based on the value of the goods—the shipper of computers will pay more than the shipper of shovels for a given portion of space on a truck or ship. However, this does not change the basic concept.

Figure 3 shows the effects of an increase in demand with no change in capacity, such as with increased development. In this case, there is an increase in overall consumer benefits due to the increased demand resulting from the new development. But there is also an increase in the overall cost, due to the facility becoming more crowded. The original users now have higher costs and lower net benefits, the darker rectangle. The overall net benefit, which was originally BP_iA, is now BP_2C. This may represent an increase or a decrease. The lower the sensitivity of demand to cost (a steeply sloped demand curve) and the lower the sensitivity of cost to volume (a flatly sloped supply curve), the more likely an increase in consumer surplus.

Figure B3   How Time Cost Changes with An Increase in Demand



If the costs the additional user imposes on other users exceed the benefits to that user, consumer surplus and overall net benefits will *decrease* with additional use.

A consumer surplus construct can also be used to show that even if a capacity enhancement induces so much development that travel times return to pre-enhancement levels, net benefits increase. This is because of the increased consumer surplus due to the new travelers. Any additional development once the pre-enhancement level of congestion is reached can not be attributed to the enhancement but must be the result of other factors inducing growth.

These graphs clearly show that reducing the number of trips, as with transportation demand management or congestion pricing, newer increases the *total benefits* of the transportation system; it only reduces the *costs*. However, if the system is very crowded, costs may be reduced more than benefits, thereby increasing *net benefit.*

If the overall demand curve and supply curve for a particular transportation facility at a particular time were known, the consumer surplus could be directly measured, giving a good representation of the net benefit of the facility to the different users and the overall net benefit. Unfortunately, it is not known. Even though the costs and volumes can be observed at various points in time as are Q1, Q2, P1, and P2 in Figure 3, Q3 can not be observed.

However, the overall quantity of transportation—vehicle volumes and vehicle-miles, passenger trips and passenger-miles—can be observed and used as an indication of benefits. This indicates all types of benefits to the travelers, both mobility and recreation. All trips that are made have benefits that exceed their costs, otherwise they would not be made. Similarly, the quantity or dollar value of freight—ton-miles or $-miles of freight—are indicators of the benefits of goods transportation.

Appendix C

Traffic Surveillance Methods

Surveillance involves the monitoring and data collection of prevailing traffic conditions. Loop detectors are the oldest and most commonly used surveillance devices. Closed circuit television is an newer, but also common method. A number of newer methods have been developed in recent years but technical problems remain and none is developed to the point of full market acceptance[11].

Loop Detectors

As a mature technology, inductive loop detectors (ILDs) have become indispensable tools for traffic operation and surveillance systems. For traffic management systems, loop data is typically relayed to a centralized transportation management center (TMC) for analysis. Although the loops are read many times a second, smoothed data is typically reported back to the TMC at intervals of 20 or 30 seconds. Loop detectors are capable of measuring flow, occupancy, and vehicle speed. The accuracy of such measurements is intimately related to the proper and uniform installation and calibration of loop detectors, a requirement that in practice can prove difficult to achieve.

The material below describes the technology and reported performance of ILD's. The document is divided into the following sections:

- Principles and theory of operation
- Detector configuration
- Selection criteria
- Installation and maintenance considerations
- Attainable information
- Data reliability
-

Principles and Theory of Operation

Loop detectors operate on the principle of inductance, the property of a wire or circuit element to "induce" currents in isolated but adjacent conductive media. A detector consists of an insulated electrical wire, placed on or below the road surface, attached to a signal amplifier, a power source, and other electronics. Driving an alternating current (normal operating frequency between 10kHz and 200kHz) through the wire generates an electromagnetic field around the loop. Any conductor, such as the engine of a car, which passes through the field will absorb electromagnetic energy and simultaneously decrease the inductance and resonant frequency of the loop. However, it is the metal conductive surfaces which are closest to the loop, generally the body of the car, which have the greatest effect on inductance. This is why cars (which are closer to the ground) induce a larger change in inductance then trucks (which have more metal, but its higher up).

[11] Justin Black prepared the material in this appendix for the PATH LEAP website.

For most conventional installations, when the inductance or frequency exceeds a preset threshold in the detector electronics, this indicates that a vehicle has been detected. Many factors determine loop inductance, including wire size, wire length, the number of turns, lead length, and insulation.

Detector Configuration

The elements of a detector include:

- an inductive loop
- a pull box
- a lead-in cable
- a detector card, which normally consists of a tuning network, a signal amplifier,
- a data accumulator or controller, and other detector electronics

The inductive loop is an insulated electrical wire, usually several meters to a side, with several turns. Loops are installed in a variety of shapes such as square, rectangle, diamond, circular and octagonal, though each configuration produces a different electromagnetic field. For instance, diamond loops reduce the probability of detecting vehicles in adjacent lanes. The pull-box, usually located adjacent to the road, houses the splices between the lead-in cable from the controller and the lead-in wires from the loop. Lead-in wires are usually shielded and twisted to eliminate disturbances from external electromagnetic fields, such as adjacent loops. The controller electronics, usually housed in a rugged cabinet in a safer more accessible location, detect, amplify, and process loop signals. The controller orchestrates loop operation and provides power. A typical controller can handle up to forty loops, though in practice will probably oversee far fewer.

Inductive loops may be placed either on the road surface, or up to twenty inches or more into the pavement. While deep buried loops exhibit a longer life span, their electromagnetic field is weaker and detection becomes more difficult. Loop sensitivity, defined as the smallest change in inductance which will cause actuation, decreases around 5% for every inch into the pavement the loop is installed. Since the characteristic shape of a detection is not distorted, additional loop turns which emit a stronger loop field can compensate for pavement interference. Unless loops are installed during road construction, installation requires a saw cut, up to 10 mm wide, into the pavement. Unfortunately saw cuts have been found to undermine the structural stability of the pavement in some cases.

Wire type also influences loop operation. Most inductive loops are formed by wrapping a single wire strand around the loop shape a prescribed number of times. Because these turns are spaced randomly, the electromagnetic field, and consequently detection results may vary from detector to detector. It has been established that a multiconductor cable, which holds the loops in uniform proximity, will produce more accurate measurements. Installation of the multiconductor necessitates a wider cut into the pavement than the single wire configuration. Ruggedized, weather resistant, pre-formed loops that increase uniformity are available, but they can be more difficult to install.

Detectors operate in either the pulse or presence mode. Presence operation, often used with traffic signals, implies that detector output will remain "on" while a vehicle is over the loop. Pulsed detection requires the detector to generate a short pulse (e.g. 100 to 150 ms) every time a vehicle enters the loop, regardless of the actual departure of the detected vehicle.

Many detectors today employ digital technologies that sense a change in the resonant frequency of a loop due to a decrease in inductance. Digital techniques allow more reliable and precise measurements than their analog counterparts. Some digital units incorporate advanced electronics such as self-tuning amplifiers, open-loop test functions, and automatic or remote reset capabilities. These features can significantly reduce detector maintenance costs and calls. The newest detectors can actually output the digitally sampled inductance "signature" of each vehicle, allowing the development of flexible signal processing software to add considerable more robustness than the hardwired set "threshold" detectors.

Selection Criteria

Several parameters characterize the performance of loop detectors:

Response time - defined as the time between the when inductance crosses the preset "threshold" due to the arrival or departure of a vehicle, and when this is indicated on the digital output side of the detector. A consistent and fast response time is crucial for accurate speed measurements. Response time is affected by vehicle size, speed, detector type, sensitivity, and wire type. Response time decreases with smaller vehicles, which have lower ground clearance and a shorter distance to the engine block and axle. Faster speeds tend to reduce response time, as does increased sensitivity. Many loop detector cards can process multiple loops, although they usually do it sequentially (A-B-A-B) rather then in parallel. This can effect the response time.

Recovery time - the time required for a loop to return to normal operation after a period of sustained occupancy. Recovery time is particularly important for vehicle counting. Loop standards dictate that after a sustained occupancy of five minutes a detector return to at least 90% of the minimum sensitivity within one second after the zone of detection is vacated.

Installation and Maintenance Considerations

For comprehensive surveillance of mainline routes, detector stations (possibly pairs) should be installed every 600 to 1200 meters. Loops should also be stationed around access and egress points, and at any locations where operational problems occur. Provisions should also be made for maintenance access. Different degrees of processing of raw loop data may be performed at the TMC or remotely in the field. Remote analysis requires additional processing hardware for each detector station, but less sophisticated communication links, since only the data relevant to highway surveillance is transmitted back to the TMC. In contrast, centralized analysis requires the transmission of large quantities of raw data and greater hardware requirements (storage and processing) at the TMC. If transmission lines are leased, which is often the case, data transmission costs can become cost prohibitive.

For most California freeway installations, the raw loop detector data is processed locally in the controller cabinet and reported back to the TMC every 30 seconds. Hence, it is usually not possible to get loop data of higher resolutioin then 30 seconds.

Loop detectors are relatively inexpensive. Rough component costs are given below. The reader is referred to a list of commercial sites for more specific cost information.

- Loop with amplifier (purchase and installation) - $700 per loop
- Controllers - $2500 per unit
- Controller Cabinet - $5,000 per unit
- Fiber optic cable (purchase and installation) - $300,000 per mile

Annual maintenance costs average around 10% of the original installation and capital cost, adjusted for inflation.

Loop failure rates are strongly related to maintenance and installation procedures. Surveys of state DOTs indicate that failure rates vary significantly, ranging anywhere from three to fifteen percent per year. In practice, loop inspection procedures also vary substantially among DOTs. Loops may be inspected anywhere from one to twenty-five times per year. Highest inspection rates occur where loop operation is critical to the operation of other deployed traffic management systems. In many instances loop maintenance costs are sufficiently high that malfunctioning loops are replaced outright, without any diagnosis of the cause of failure. Because loops have been deployed extensively, consistent installation recommendations and primary causes of loop failure have been documented.

Installation

Installation recommendations for effective and long lasting loop systems include:

- Installation and calibration should be as uniform as possible.
- Saw cuts should be cleaned out and dried before loop installation. Saw cuts should also be of uniform depth.
- Loops should be properly sealed.
- Detectors should feed off the same power supply.

Mechanical Failure

Many factors contribute to physical loop failure. Pavement and sealant (of the saw cut) failure are commonly identified as the primary culprits. Pavement failure or deformation (cracking, rutting, potholes, or shoving) causes loop wires to be strained resulting in breakage, wire insulation wear, or the infiltration of foreign materials. Sharp bends in loop corners have also been found to cause problems, such that the insulation deteriorated or was broken.

Sealant failure poses additional problems. Once the sealant fails, the loop may become exposed or foreign materials may infiltrate the cut. In many cases the loop was found to have floated to the top of the cut, either before the sealant could cure or because it remained plastic. Other common sources of loop failure include poor installation and maintenance procedures, damage from utility repair or construction, lightning surges, detuned amplifiers, and corroded splices or wires.

Data Malfunction

Many sources of loop malfunction can produce erroneous detector data. These include stuck sensors, hanging (on or off), chattering, cross-talk, pulse breakup, and intermittent malfunction. Cross-talk involves the mutual coupling of magnetic fields that produces interaction between two or more detector units which are in the same cabinet or in close proximity to each other. Cross-talk results in erratic loop behavior and inaccurate detections. Pulse break-up involves gaps in detector actuation data, which may be incorrectly interpreted as different vehicles. As described below, many of these problems can be corrected with data filters.

Attainable Information

Loop detectors supply several pieces of information about prevailing traffic conditions, including vehicle presence, flow, occupancy, and velocity. A good loop detector system is cited as accurate to within 5%. The accuracy and consistency of detector output is a strong function of installation and calibration procedures. For example, it is possible that detectors with different sensitivities longitudinally separated by thirty feet give occupancy data that differs by 40%. Loop detectors are also limited by their inability to detect stationary vehicles.

Flow and occupancy may be extracted directly from loop data. Speed may be approximated from the data of a single detector using the fundamental theory of traffic flow:

flow = speed * density

where density is approximated from occupancy by:

density = occupancy * g

and

$$g = \frac{K}{\text{(vehicle length + detector length)}}$$

where K is a conversion factor. While it is possible to obtain reasonable speed estimates with this strategy, paired loops offer a more accurate approach. Velocity is calculated from the travel time between two loop detectors that are separated by a known distance. Accurately calibrated speed traps with loops of individual wire can expect to achieve measurement errors of 5-8 kph (3-5 mph) at low speeds and 16-19 kph (10-12 mph) at high speeds. Multi-conductor cable loops average errors about 0.3 kph (0.2 mph) at low speeds and 5-8 kmh (3-5 mph) at high speeds.

There are several considerations in speed trap design. For one, loop inductance is a strong function of vehicle speed. One study determined that a vehicle traveling at 20 mph produced a 3% inductance shift, while another at 80 mph yielded only a 1% inductance shift. Sensitivity settings may have to be adjusted when ILDs are used in high-speed freeway environments. The separation between loops is another relevant variable. In practice anywhere from 2 meters to more than 20 meters is feasible. However if detectors are too close cross talk may occur, while detectors spaced too far apart may be susceptible to vehicle lane changes. Suggested optimal spacing is around 9m.

Data Reliability

For traffic management strategies such as incident detection to be effective loop data must be reliable and accurate. Many TMCs use a combination of manual inspection and reliability tests to validate incoming data. Such tests serve a dual purpose, they flag erroneous data and identify malfunctioning loops. Various approaches are employed to identify inaccurate data.

Initial error detection often occurs in the field. The data may be filtered, where pulses or gaps in actuation less than some brief interval, say one-fifteenth of a second, are ignored. The data may be flagged as unreliable if a microprocessor sees more than two valid pulses (vehicle endings) in a second. These tests usually detect gross errors, but other malfunctions may go unnoticed.

More advanced filtering techniques are available to validate loop data. One approach is to compare a detector's "on" time to the average "on" time of all other detectors at that station. A second strategy compares detector data (volume, occupancy, and speed) against realistic thresholds at periodic time intervals. For example, detector data is flagged if occupancy exceeds a predefined maximum for a certain period of time (say more than 90% for five minutes). A more complex algorithm uses a multi-regime comparison of the flow - occupancy ratio to maximum and minimum expected speeds. Occupancy is converted to density using a variable *g* that varies as function of occupancy. Research has shown that a constant *g* can introduce significant error into speed estimates. These algorithms achieve good detection rates with low false alarms rates, and often identify malfunctioning detectors overlooked by manual inspection. In practice most TMCs operate in a hybrid fashion, using several elements from the tests described above.

Most existing loop detector cards have a hardware setting for the threshold sensitivity where they register a vehicle. Greater sensitivity requires longer data integration time, and this can effect performance at freeway speeds. An incorrect sensitivity setting may either double count truck tractor trailers, or not count them at all. The unoccupied loop inductance may change away from the detector card's hardware due to temperature, rain, corrosion, or mechanical ware. Vehicles switching lanes between loops are often missed entirely. For these reasons, the best volume accuracy that is normally expected from loops under ideal conditions is around 95%.

However, PATH researchers have noticed that as long as the loop is not entirely shorted out, there will always be a change in inductance with every passing vehicle. PATH is developing software that is examining the (digitally sampled) analog change in inductance, rather then relying on a hardware set threshold. By setting the threshold between what "is" and "is not" a vehicle dynamically in software, PATH has demonstrated that they can get accuracy much greater then 95%. By processing the loops as a lateral array, it is also possible to capture every vehicle - even those that are traversing lanes. The advantages to this mechanism are that it dynamically changes with any mechanical or weather induced change in the loop. If the loop is not performing adequately, it can report this, while simultaneously compensating for the malfunction.

An additional advantage of this system is that it can not only precisely determine the vehicle length, but it also generates a semi-unique inductive loop "signature" for every vehicle. These inductive loop signatures can be re-identified downstream to get the true point-to-point travel time. With PATH software and the new commercially available inductive loop signature detector cards, it appears possible to finally get consistently reliable data from loops.

Video Image Processing

Video image processing (VIP) systems employ machine vision technology to analyze traffic data collected with Closed Circuit Television (CCTV) systems. Although the technology to digitize CCTV images has been available for a number of years, downward trends in computer and image processing hardware costs have recently made VIP detection systems more attractive. The availability of relatively inexpensive high performance computers is particularly important because a full bandwidth, digitized television image provides a data stream of 6.25 Megabytes per second.

Though the technology is still developing and field deployment is somewhat limited, initial results seem promising. The advantage of VIP detection lies in its wide area detection capabilities, usually several lanes for one camera. This allows the efficient detection of shock waves and other spatial traffic parameters, such as density, queue lengths, and speed profiles, which can not be easily obtained by conventional devices. In addition, video detection provides ancillary information such as traffic on the shoulders, stopped vehicles, lane changing, speed differential, and traffic slow downs in the other direction. PATH is presently conducting research into machine vision based traffic surveillance. The remainder of the discussion is divided into the following sections:

Improved Detection Capabilities

Perhaps the major handicap of conventional automated incident detection (AID) systems is that they are designed to operate with data taken at a point rather than over space. This information alone, typically volume and occupancy, has not proven to be sufficient for effective and reliable incident detection. The data is deficient because volume is not a dynamic measurement, and because occupancy is an approximate rather than true measurement of a spatial traffic flow variable, namely density.

VIP detection systems offer numerous benefits. In addition to the detection of incidents and the collection and analysis of traditional traffic data, VIP detection systems can classify vehicles, monitor intersections, read license plates, and perform image compression. Because installation typically does not require lane closures, traffic personnel safety is enhanced and traffic disruptions are minimized. By repositioning cameras as road geometry varies, VIP detection systems can be used during realignment or resurfacing. If desired, video detection can be used to provide or supplement existing video surveillance, and is readily integrated with other management strategies such as variable message signs and adaptive ramp metering control.

Though initial capital outlays may exceed the costs of traditional detection methods such as loop detectors, video detection costs are predicted to decrease dramatically as the technology is refined and more systems are deployed.

VIP Incident Detection Systems

VIP detection systems generally fall into the three categories described below. A more precise classification is difficult, since in practice most systems operate in a hybrid fashion.

Tripwire Systems

In tripwire systems the camera is used to emulate conventional detectors by using small localized regions of the image as detector sites. Multiple detectors can be located within the image and detectors can be easily configured to suit the road geometry. The drawbacks to this approach are that the underlying incident detection algorithms remain the same as for conventional detectors, and that the accuracy of individual detectors is heavily dependent on the camera field of view. AUTOSCOPE and CCATS are examples of systems based on this approach.

Tracking Systems

These systems detect and track individual vehicles moving through the camera scene. This provides a microscopic description of vehicle movements which can reveal new data on events such as sudden lane changes, vehicles traveling in the wrong direction, and stationary vehicles. This increase in sophistication requires more computing power, requires individual vehicles to be discernible and can be even more restrictive in camera positioning. TRISTAR (INVAID),EVA, and IPVD are examples of such systems.

Spatial Analysis

The third approach, used in the Image Processing for Automatic Computer Traffic Surveillance (IMPACTS) system, concentrates on analyzing the two-dimensional information that video images provide. Instead of considering traffic on a vehicle by vehicle basis, the underlying strategy is to describe how the visible road space is being utilized at a particular instant in time. Disturbances in traffic flow can then be determined by analyzing how these descriptions vary over time. Use of road space is divided into three categories: no traffic present, moving traffic present, or stationary traffic. These are essentially qualitative decisions.

Measures of Performance

There are many metrics which characterize the performance of VIP detection systems. The detection rate, false alarm rate, and detection time of a VIP system are particularly important.

False Alarm Rate (FAR)-	Defined as the fraction of incorrect detections to the total number of detections. The FAR is typically expressed as a percentage, but may also be given as the number of false alarms per time period

Time to Detection (TTD)- Defined as the average time required to detect and incident. It applies at a given FAR and DR

Detection Rate (DR)- Defined as the number of detected incidents to the actual number of incidents in the data set, the DR is given as a percentage.

These parameters are not independent, rather there exists a trade off analogous to statistical hypothesis tests (i.e. alpha and beta statistics). Detection systems set to detect a large percentage of incidents must be highly sensitive, and consequently they also tend to produce a large number of false alarms. Similarly, while less sensitive systems produce fewer false alarms, they also detect fewer incidents.

The performance of VIP detection may vary with several environmental variables, including:

- variable lighting conditions, particularly during sunset and sunrise
- camera angle, height, and position
- adverse weather conditions (eg. rain, fog, and wind)
- direction of traffic flow
- the presence of camera vibration

Performance issues from the perspective of the transportation authority include:

- Measurement accuracy of traffic parameters
- Vehicle classification reliability
- Ease of setup and operation
- System failure rate and ease of recovery
- Compatibility with other TMC elements
- User interface, data storage, and data displays
- Cost

General Characteristics

Video detection is receiving attention at the international level, with evaluations underway in Europe, Japan, and the United States. Systems are most commonly deployed on major urban freeways that carry heavy traffic volumes or are incident prone. Though the detection strategy varies between systems, each shares a similar functional architecture.

Video is first collected with closed-circuit television cameras, usually mounted on roadside poles at heights from 5 to 15 meters. Each camera can monitor an area up to five lanes wide and several hundred meters in length. Image processing equipment then extracts relevant information from the digitized camera footage. A higher level processing unit orchestrates equipment operation, performs detection, and provides human interface functions. Processing equipment may be centralized at the traffic control center or located in the field adjacent to the cameras. Most systems rely on dedicated transmission lines connected to traffic management centers, which allow center personnel to multiplex between camera views.

Performance

Since the technology is still being developed, most evaluations focus on detection performance rather than cost-benefit analyses. Of the figures that are available, monetary benefits are calculated in terms of safer driving conditions (fewer secondary accidents) and reduced vehicle delay. For highly congested sites, video detection systems can offer benefits up to $98,000 per Km per year of monitored roadway. Results also suggest that system costs are amortized in four to ten years. System costs vary considerably depending on existing equipment along the roadway. With an existing transmission link, a camera, mast, and processing equipment can cost around $30,000. Video systems complement other management technologies and may be integrated into existing CCTV systems.

Because of site-specific variables and varying system configurations, detection performance is difficult to generalize. Under most conditions, VIP systems can achieve detection rates higher than 90% with a false alarm frequency as low as once every several days. VIP systems can also accurately measure traffic parameters such as density, speed and flow.

Though environmental variables can affect VIP detection performance, fairly consistent recommendations for robust operation have been developed. Cameras should be mounted as high as possible and should be fitted with a polarizing filter and auto-iris. Cameras should have a field of view of several lanes, and care should be taken to eliminate any objects that cast shadows or obstruct the camera field of view. It is not clear whether cameras should view approaching or departing traffic.

Assessment

Most results suggest that video detection systems are able to detect incidents at least as accurately as conventional systems. To develop a complete picture of their effectiveness, additional cost comparisons using side by side tests along the same corridor are required. Several evaluations into these areas are presently underway. Since highly technical systems can also be fairly unstable, additional research into system reliability, maintenance requirements, and ease of recovery seems warranted. It seems that VIP detection might be useful as an element in the surveillance systems of roadway.

Deployed VIP Detection Systems

The table on the following page describes VIP systems presently deployed or under evaluation.

Table B-1 Reported Performance, Benefits and Costs

	System Location	System Performance, Benefits, and Costs
AID System	Osaka and Kobe, Japan	Under normal operating conditions, the system measured speed and volume with an accuracy exceeding 90%. The cameras installed at 6.5 meters proved more accurate, as the lower cameras had problems differentiating obscured vehicles in the lane furthest away. For optimal performance, cameras should cover two to three lanes of traffic and view traffic from the rear. During the evaluation the system detected 69 out of 79incidents, and because incident footage was automatically recorded (80frames of video kept in memory), accidents could subsequently be analyzed. The system significantly reduced detection time, allowing authorities to more quickly activate VMS signs on the approaches (such as "ACCIDENT AHEADON LEFT"). Such notification reduced the distribution of average approach speed to the curve by at least 5 km/hr.
AUTOSCOPE	Minneapolis, Minnesota	Installation of the video system was completed in 1994, though no quantitative results were available at the reference date of publication. Several evaluations were underway, including a life-cycle comparison of the video system with loop detectors, and the potential for integration with adaptable ramp metering and VMS systems.
IPVD	Tokyo, Japan	The IPVD measured speed within +/-2% and volume with at least 90% accuracy, even during congested periods. Field engineers with no knowledge of image processing were able to initialize the system in 30 minutes. The performance evaluation will be performed once the IPVD units are connected with the Tokyo Metropolitan Police Control Center.
IMPACTS	London, England	The system appears robust and can be effective even with difficult views from relatively low camera positions. IMPACTS successfully detected queues of stationary traffic, stopped isolated vehicles, and empty road. Detection better than 90% can be expected, and the false alarm frequency is low, around 1 every 24 hours. Comparison with the HIOCC loop detector system was not completely clear, although there was some correlation between the results. Although it was not possible to say which was better with the data collected, IMPACTS more accurately detected stopped traffic, isolated stationary vehicles, when traffic had started to move, and when congestion had ceased. Researchers encountered several problems during the evaluation. Adjacent artifacts which cast strong shadows led to the false detection of stationary traffic, and rapid variations in contrast due to the functioning of the auto-iris caused image content to change faster than the system cycle rate, distorting results.
	Glagsgow, Scotland	At the date of publication no quantitative results were available. During a video data collection exercise in January of 1994 the system correctly identified the build up of queues. Problems were reported with shadows during periods of low angle sunshine.
	Kent, England	Initial results show that some detections have been made, but that the number of false alarms can be highly variable. For one particular week two cameras showed only three false alarms in a seven day period, while an adjacent camera had more than 200 false alarms during the same period.

System Title	System Location	System Performance, Benefits, and Costs
INVAID-TRISTAR	Antwerp, Belgium	Researchers estimated benefits from accident detection, in terms of vehicle hours saved, at 78.6 KECU (approximately $98,000 US as of October 1996) per Km per year. This figure is about ten times the cost of the system per Km in a depreciation period of 10 years. Approximately 5-10% of incidents along the roadway are secondary ones, and of these, traffic authorities estimated that 30 to 80% could be avoided with an adequate detection procedure including the INVAID system. Under this scenario, 6to 32 major incidents per year could be avoided on the 40 Km roadway. The system was integrated fairly easily into existing CCTV installations of varying age and quality. Image quality factors such as definition, contrast, and white level seemed to have little impact on performance. In fact, by acting as a low-band pass filter, poor image definition had a beneficial impact on the FAR. The system operated at a 77% DR and with a 4% FAR, or0.02 false alarms per camera per day.
	Lyon, France	Over a three-month period INVAID detected 86% of 2400recorded incidents, with a FAR of 13%, or 0.26 false alarms per camera per day. Considering reduced vehicle delay and improved safety, the system investment was amortized in 4 years.
	Madrid, Spain	No quantitative results available at the date of publication. An analysis of improvements in the overall system performance and of ease of integration with other traffic management strategies is being conducted.
	Nice, France	INVAID operated with a 91% detection rate, a mean time to detection of 22 seconds, and a 6% FAR, or 0.17 false alarms per camera per day. Based on this performance, ESCOTA subsequently expanded the number of cameras along the route to 40. For optimal performance, it was necessary to: remove nearby tree branches which, with wind, caused false alarms; eliminate reflections by adding a polarizing filter; protect transmission cables from electromagnetic disturbances generated by a nearby railway; and eliminate the situation where the camera shutter frequency and public lighting current appear to be in sync.
		In tunnel conditions INVAID achieved an 87-90% detection rate, with one false alarm per camera every five days. Cameras should view traffic from the rear, have polarizing filters, and should be cleaned at regular intervals. If possible, minimum lighting should exist in the tunnel.
	Tours, France	The system achieved a detection rate greater than 90%,with a FAR of 25%, or 0.8 false alarms per camera per day. False alarms were generated by alternate dark-sunny areas within the detection zone, and at night, by parasitic reflection or raindrops on the protective glass cover of the camera case. The cost of the system varies considerably depending on existing equipment along the motorway. On a section already equipped with a means of transmission, the cost of a mast, camera, and picture analyzer is somewhere around 200,000 Francs (approximately $38,000 US as of October1996). This represents the pilot costs, which is expected to decrease considerable once marketing is launched.
EVA	N/A	EVA provides a complete set of traffic parameters including counts, volume, speed, density, occupancy, and spatial headway. EVA can also relay static or moving images. The system is reported to be accurate to within 2% under most conditions, monitors several lanes of roadway, and is able to track vehicle lane changes.

Author: Justin Black

Other Vehicle Detection Technologies

A wide variety of devices have emerged on the market which maybe used to replace or complement inductive loop systems. Although the technologies described below are primarily point detectors, in many cases they address deficiencies of loop detectors. Many of these new technologies are roadside (or vertical sensors) which do not require pavement cuts or the disruption of traffic for installation. Many of the sensors are cost competitive with ILD systems. Future freeway management systems will be able to integrate a wide variety of detector technologies to produce more robust surveillance and AID systems
In some areas transportation management authorities have also established cellular phone numbers for incident reporting. Such programs have been successful, complementing TMCs automated incident detection systems. Detection technologies discussed below include:

- Magnetometers
- Microwave Radar
- Ultrasonic Detectors
- Acoustic Detectors
- Infrared Detectors
- Cellular Phones

Magnetometer

The magnetometer uses magnetic anomaly detection, the principle that ferrous vehicles collect and distort the fairly uniform magnetic field lines which should the earth. It is a in-road sensor which picks up the magnetic disturbance sin the earth's field as a vehicle (ferrous metal) passes over. It is a passive devices, and does not radiate a field. The magnetometer is normally used where the only information required is vehicle presence. It exhibits excellent detection rates, and is effective for counting vehicles The magnetometer is often used in place of loops on bridge decks, where ILDs can not be installed, and in heavily reinforced pavement, where steel adversely affects loop performance. ILDs and magnetometers each have their respective applications and tend to complement each other.

The magnetometer consists of a small, in-road probe, a lead-in cable, and an amplifier. The typical magnetometer probe is cylindrical, encased in polyurethane, and about 1 inch by 4 inches. Installation is achieved by drilling a hole into the pavement, though fewer linear feet of pavement saw cut is required than for ILDs. Smaller, solid-state magnetometers have been developed, which may see more use in the future.

Microwave Radar

Microwave radar vehicle detectors transmit electromagnetic energy at the speed of light in frequency bands between 2.5 to 24.0 GHz. They are able to count vehicles, measure speed, and, in some configurations detect vehicle presence. Experimental models have been used for vehicle classification by measuring the vertical profile of a vehicle. They are generally insensitive to weather, provide day and night operation, and perform best on fairly open road where long-range capabilities can be taken advantage of. Advanced units can measure target distance, thus one unit can have multiple detection zones.

The waveform used to transmit the microwave energy determines the type of traffic data which may be obtained. Continuous-wave energy, transmitted at one frequency allows vehicle speed to be measured from the Doppler shift in the received signal. The frequency shift is proportional to vehicle velocity. Stationary vehicles can not be detected with this wave form. The frequency modulated, continuous wave signal allows the measurement of both speed and presence, as well as the detection stationary vehicles. Pulse waveforms are used with ultrasonic and laser radar sensor technologies. These technologies measure distances to the road and vehicles, providing vehicle count, presence, and occupancy information. By transmitting pulse energy at two calibrated, closely-spaced incident angles, vehicular speed may be measured by recording the time at which the vehicle crosses each beam.

Ultrasonic Detectors

Pulsed-Doppler Ultrasound

These sensors operate on the same principles as Doppler radar, except that they emit sound waves with frequencies between 20 and 200 KHz, which are above the human audible range. They are pressure waves that travel through the air at about 740 mph at sea level. They can measure speed (so-so), occupancy, presence, and, in some configurations, queue length. Because of the complexity of the returns from moving vehicles, the effective range of the sensor (around 40ft) is smaller than the microwave detectors.

Because ultrasonic waves propagate through the air, they are subject to attenuation and distortion from a number of environmental factors including changes in ambient temperature, air turbulence, and humidity. Nearly all targets reflect ultrasonic sound waves, but textured or porous surfaces produce a weaker echo.

Pulsed Ultrasound

Vehicle profiling can be achieved by installing a pulsed ultrasonic detector above the roadway. The vertically aligned (downward looking) transducer measures wave travel time to the pavement or to the top of passing vehicle. A fast pulse repetition rate (~13 Hz) allows a minimum of 2 to 3 measurements with up to 1" resolution to be made. Excellent classification performance can be achieved for most vehicle types, though the sensor can have difficulty differentiating between cars and vans. Air turbulence and temperature adversely effect performance.

Passive Acoustic Detectors

This detector is usually configured as vertical dipole array of microphones that listen to noise produced by approaching vehicles. The time delay between the arrival of sound at the upper and lower microphones changes with time as the vehicle emitting the sound passes under it. When the vehicle is distant from microphones the sound arrives almost instantaneously at both phones. When the vehicle is under the microphones, sound reception at the upper microphone is delayed by the inter-sensor distance. Vehicles are tracked using cross-correlation between phones, and best results are achieved when the data is filtered to a bandwidth of 50-2000Hz. In this band frequency content includes both engine and tire noise (though most of the acoustic noise is produced by the tires). Interference between the noise of multiple vehicles is a key limitation to acoustic technology. Detection and speed performance are poor and fair, respectively.

Infrared (IR) Detectors

Laser IR Radar

Laser detectors operate on the same principles as microwave radar, but transmit energy at higher frequencies (shorter wavelengths). The detector senses a portion of the reflected energy in its field of view. The distance of an object from the detector is found by measuring the two-way travel time of the infrared pulse, from the detector to the target and back. Lasers provide presence, speed, volume, occupancy, and classification information in day and night conditions. IR detectors are vulnerable to weather conditions such fog, mist, rain, and snow which scatter and attenuate wave energy. Problems also may arise when radar locks onto and measures the speed of the strongest back-scattered signal, excluding smaller vehicles in the same area.

Passive Infrared Detectors

Passive IR detectors do not transmit energy themselves, but rather measure energy emitted by objects in their field view. By measuring the difference in emitted energy (temperature) from the road and from vehicles, vehicle presence is detected. Passive IRs provide volume, occupancy, and presence information. These sensors have difficulty measuring speed because the extended nature of the vehicle distorts the IR signature, making velocity less clear. Weather such as fog, snow, and precipitation which scatter energy can have adverse effects on operation.

Cellular Phone Technology

Several DOTs and local transportation authorities have implemented programs which allow drivers to call and report incidents from their cellular phones. Evaluations have found cellular detection to be effective, and that a the large proportion of major incidents are first identified from cellular reports. In the Chicago program more than 100,000 calls are received annually. About two-thirds of the calls are originating (provide new information),while about a quarter are duplicate calls, providing previously received information. These programs have received favorable media reaction and public acceptance. Other unexpected benefits also occurred, such as the reporting of malfunctioning traffic signals, debris in the roadway, and other incident causing conditions. Cellular detection is most effective during peak periods when coverage is highest. As cellular phone ownership rates continue to increase, cellular detection promises to become a useful component in freeway surveillance systems.

Reported Detection Performance

Detector Type	**Vehicle Presence**	**Detection Performance-Vehicle Counts**	**Speed Estimation**	**Vehicle Classification**
Magnetometer	Yes	Excellent	N/A	N/A
Microwave (Doppler) Radar	No	Fair	Excellent	Poor
Doppler Ultrasound	No	Good	Fair	N/A
Pulsed Ultrasound	Yes	Very Good	N/A	Good
Passive Acoustic	No	Poor	Fair	Good
Active Infrared	No	Very Good	Excellent	N/A
Passive Infrared	No	Very Good	Poor	N/A

Author: Justin Black