The final report of ITS Center project: Cost-benefit analysis for ITS: CORSIM.

UVA Center for Transportation Studies

A Research Project Report

For the Center for ITS Implementation Research

A U.S. DOT University Transportation Center

Performance-Cost Evaluation Methodology for ITS Equipment Deployment

Authors

David C. Clark
Dr. William T. Scherer
Dr. Brian L. Smith

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein.   This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange.   The U.S. Government assumes no liability for the contents or use thereof.

Abstract

Although extensive Intelligent Transportation Systems (ITS) technology is being deployed in the field, little analysis is being performed to evaluate the benefits of implementation schemes. Benefit analysis is particularly in need for one popular ITS technique: ramp metering systems. The reported benefits of ramp metering systems range widely and little analysis has been performed to determine the magnitude and cause of this discrepancy. There is such a wide range of opinions that some transportation professionals are currently questioning whether ramp meters generate any benefits at all.

The variation in opinions is largely due to the fact that the relationship between the performance and cost of ramp metering systems is not well understood. Without understanding what factors cause ramp meters to perform well, it is very difficult to efficiently design a new ramp metering system. Additionally, insufficient analysis of prior benefits causes difficulty for transportation engineers attempting to make rational design decisions based on past results.

This study investigates the benefit-cost relationship of ramp metering systems by constructing a performance-cost tradeoff curve that displays the maximum achievable performance for any budget. A simulation-optimization methodology was developed to generate the tradeoff curve. This methodology employs heuristic search techniques to locate the "optimal" ramp metering deployment scheme (quantity and location of equipment) for each budget. All deployment schemes are evaluated using CORSIM, a traffic simulation tool.

The performance-cost curve constructed by the simulation-optimization methodology has the shape of a step function. Significant benefits are possible with only

a small initial investment, however, beyond this initial step, diminishing returns are experienced as benefits slowly increase with additional spending.

Analysis of the tradeoff curve provides valuable information about the performance-cost relationship and ramp metering systems in general. Recommendations for improving ramp metering implementation are developed using this information. Also offered are additional recommendations for improvement of the simulation-optimization methodology and suggested areas of future research.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Purpose

The relationship between the costs and benefits of deploying Intelligent Transportation System (ITS) equipment has not been sufficiently explored. The purpose of this research is to infer the cost-benefit relationship of ITS deployment via analysis of empirical simulation data. Once this relationship is understood, transportation professionals will be able to objectively evaluate ITS designs based on their expected costs and potential benefits.

1.2 Intelligent Transportation Systems

The United States surface transportation system accommodates four trillion passenger-miles and three trillion ton-miles of freight per year. With travel demand expected to increase thirty percent between 1999 and 2009, more than 4,400 lanes-miles of roadway need to be added each year in order to maintain current levels of congestion. However, high construction costs and diminishing available land make building new roads unattractive, and roads are currently only being added at a rate of 3,000 lane-miles per year [26].

One solution is to build fewer new roads and invest money in Intelligent Transportation Systems. Intelligent Transportation Systems integrate technologies in fields such as information processing, communications, control, optics, and other electronics to increase the effective capacity of existing transportation infrastructures. ITS includes a diverse range of program areas such as Advanced Traffic Management Systems (ATMS), Advanced Traveler Information Systems, Commercial Vehicle

Operations, Electronic Toll Collection, and Intelligent Vehicle Initiatives [30]. These systems increase the effective roadway capacity and improve safety on the roadway while costing far less than building additional roads. A 1997 study conducted for the United States Department of Transportation and the Intelligent Transportation Society of America found that ITS systems produce a benefit-cost ratio of more that 8:1 in the nation's 75 largest metropolitan areas [4].

1.2.1 Ramp Metering

Ramp metering is a ATMS technique that was first implemented in the early 1960s in Chicago and is now in place in 23 metropolitan areas in the United States [27]. This control strategy places traffic signals on freeway entrance ramps to regulate the number of vehicles entering a freeway. The traffic signal reduces the flow of cars onto the freeway. Ramp metering benefits arise from the smoothed flow of traffic on the freeway and the storage of excess vehicles on the entrance ramp instead of on the freeway mainline. Reduced congestion on the mainline leads to increased speeds and improves overall corridor flow despite the extra delay vehicles experience on the ramp.

Ramp meters can be quite simple or extremely complex. Pre-timed ramp meters have a constant metering rate and simply allow one vehicle to pass every set number of seconds. These meters are programmed to turn on and off at predetermined times and are unaware of the current traffic conditions. Conversely, traffic responsive meters communicate with vehicle detectors in the roadway and calculate an appropriate metering rate based on the current traffic conditions. Figure 1.1 shows a schematic of a traffic responsive ramp meter.

Direction of
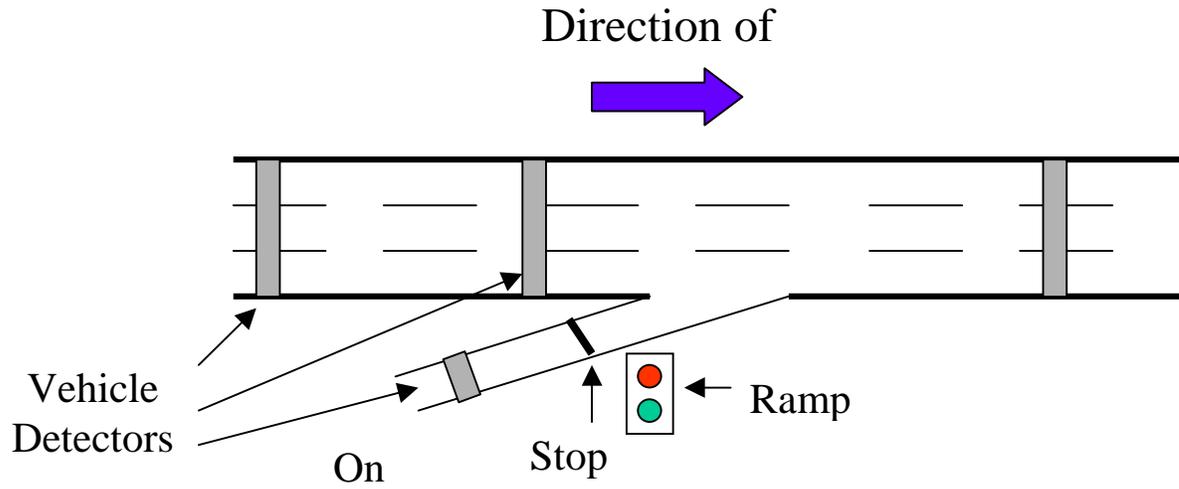
Figure 1.1: Schematic of a Traffic Responsive Ramp Meter.

Traffic sensitive ramp meters can be classified as local or system wide. Local ramp meters act independently of all other ramp meters when determining their metering rate. They only consider the traffic conditions within a mile or two of the on-ramp and select a metering rate with the objective of optimizing traffic flow in the ramp's local area. Conversely, system wide ramp meters communicate with vehicle detectors and other ramp meters outside of the local area. Each ramp's metering rate is based on traffic conditions throughout the system and the metering rate at other ramps. Table 1.1 contains a two-by–two matrix which displays available ramp metering algorithms for each class of ramp meter.

	Local	System Wide
Pre-timed	• Clock-time	• System wide Clock-time metering
Traffic Responsive	• Demand-Capacity • Speed • Occupancy • ALINEA	• Bottleneck • Predictive • Neural Networks • Fuzzy Logic

Table 1.1: Available Ramp Metering Algorithms

1.3 Rationale

Ramp metering systems across the country are providing a wide range of benefits. Ramp metering systems in Detroit, for example, have reduced accidents by 50% while ramp meters on Long Island have reduced the accident rate by only 15%. Additionally, ramp meters in Portland have increased mainline speeds by over 60% on some stretches of road, but Detroit only reports an 8% increase in speed on its metered sections of freeway [26]. Unfortunately, like ramp metering, most other forms of Intelligent Transportation Systems are also experiencing a large range of benefits.

Compounding this situation is the fact that some parties are questioning the accuracy of past evaluation studies. The local press claims that the Minnesota Department of Transportation (MnDOT) has been overstating the benefits of ramp metering. According to a 1997 MnDOT report, ramp meters in the Minneapolis/St. Paul metro area typically increase speeds and flow rates by thirty percent while decreasing the accident rate approximately forty percent [33]. However, in March 2000, the Transportation Committee in Minnesota's House of Representatives endorsed a bill to

turn off the ramp metering system for one month to perform an independent benefits study [1].

Thus far, researchers have had very little success determining why some ramp metering systems have provided drastic improvements, while other systems have done little to reduce congestion and increase safety. Some of this difficulty can be attributed to the large number of variables that differ from system to system. Site specific factors such as flow rates, level of congestion, driver aggressiveness, and road geometry likely all affect the observed benefits. Additionally, design decisions such as the number of meters installed, the spatial location of the meters, and the algorithms used to determine the metering rates affect the level of return.

Making the situation even more difficult is the fact that ramp meters are often installed at the same time as other freeway improvements such as variable message signs (VMS), additional travel lanes, and HOV lanes [25]. This makes it quite difficult to evaluate the benefits of the individual components.

Unfortunately, individual municipalities generally do not perform thorough evaluations of ITS deployment activities. Any evaluation is almost always performed to justify expenditures instead of attempting to understand why and under what conditions benefits occur [34]. ITS benefit studies rarely provide information concerning the pre-existing conditions or the specific implementation details. Therefore, it is extremely difficult for an outside researcher to combine data from several different systems and determine what factors lead to greater improvements.

These factors cause ITS professionals to have a very difficult time determining which parameters determine the level of benefit generated by ramp metering. Without

understanding what drives benefits it is extremely difficult to predict the benefits that will be generated by a ramp metering system design.

The failure to perform and share ITS benefit analyses has led to two significant problems. First, appropriation committees are unable to predict accurately what budget is needed to achieve their ITS goals. Additionally, once a budget is selected, engineers are unable to recognize an inefficient ITS design. Earlier mistakes are repeated and inefficient designs are copied. If this situation is not improved, public money will continue to be wasted.

1.4 Goals and Objectives

The goal of this research is to investigate the relationship between the costs and benefits of ITS projects. However, due to time and feasibility constraints this study will only analyze ramp metering systems. Choosing ramp metering systems allows this exploration to focus on one area of ITS technology and reduces the complexity and state space of possible deployment schemes. For the purpose of fulfilling this goal, two specific objectives for this study have been established:

1 To generate a performance-cost tradeoff curve (i.e. optimal frontier) which clearly illustrates the relationship between the cost and achievable performance of a ramp metering system.

2 To develop useful ramp metering guidelines from the information gained while building and analyzing the tradeoff curve.

The relationship between the benefits and costs of ramp metering has not been sufficiently explored and the shape of the optimal frontier between performance and cost is unknown. Once generated, the optimal frontier curve will be an extremely helpful tool that will allow financial funding committees of ramp metering systems to objectively choose an appropriate budget for a project given their desired benefits and monetary constraints. Figure 1.2 illustrates possible shapes of the performance-cost tradeoff curve.
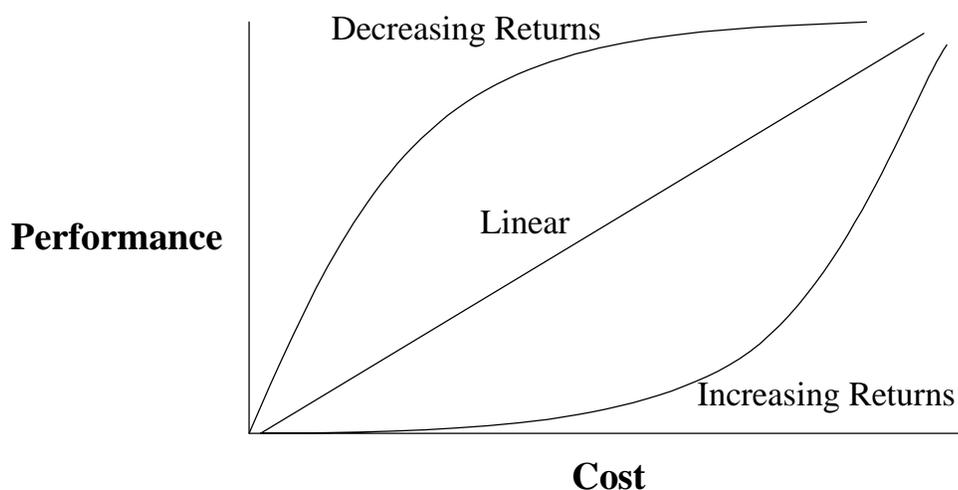
Figure 1.2: Possible Shapes of performance/cost tradeoff curve

The ramp metering guidelines will contain information that the transportation professional can reference when designing or evaluating a proposed ramp metering system. It will inform the reader which parameters drive ramp metering benefits and how other parameters affect the system. By following these guidelines, ramp metering designers will be able to choose an efficient design which generates the maximum

possible benefits for their budget and lies on the optimal portion of the performance-cost tradeoff curve.

1.5 Scope

This study investigates the relationship between the costs and benefits of ITS projects. However, in order to maintain a manageable level of complexity and increase the likelihood of generating useful recommendations, the scope of this research was limited to one type of ITS project: ramp metering.

Additionally, due to time and data constraints, all ramp metering deployment schemes were tested on models of roadways in Southeastern Virginia. Precautions were taken to ensure that the results from this study are transferable to other freeways, but it is beyond the range of this study to test multiple freeway locations in different geographic areas.

Multiple ramp metering algorithms need to be implemented in order to determine the maximum possible benefits of a ramp metering system. However, some metering algorithms are extremely complicated and are still in the developmental stage. Advanced metering algorithms that incorporate neural networks, fuzzy logic, and traffic prediction were not used in this study since their implementation and testing are significant research endeavors in themselves.

1.6 Overview of Technical Report

The remainder of this report details the study performed to generate the performance-cost tradeoff curve and ramp metering guidelines. Chapter Two provides a Literature Review focusing on prior research in the measurement and evaluation of ITS benefits. The simulation-optimization methodology employed to generate the tradeoff

curve and the measures of performance used are described in Chapter Three. Chapter Four provides details of the implementation of the simulation-optimization methodology. This chapter focuses of the development of the traffic simulation model and the pseudo-random search methodology applied to locate the "optimal" deployment scheme. The results of the study are presented in Chapter Five, where the shape of the tradeoff curve is discussed. Chapter Six is dedicated to testing assumptions and performing a sensitivity analysis. The affects of modeling assumptions and site-specific variables on the results are discussed in this chapter. Chapter Seven contains conclusions about the performance-cost relationship and lessons learned from the curve can lead to more efficient ramp metering designs. This section also contains conclusions about the simulation-optimization methodology and recommends areas of future research.

Chapter 2: Literature Review

2.1 Benefit-Cost Ratio Method

Many government agencies and private businesses use the benefit-cost ratio method to evaluate and compare capital investment projects. The ratio is calculated by dividing the present value of the expected benefits of an investment by the present value of its expected costs. The value of the ratio is directly related to the investment's rate of return, where higher ratio values correspond to larger rates of return. Investments with a benefit-cost ratio value less than one should not be considered since they produce a rate of return lower than the risk-free interest rate.

Analysts employing this method to choose between mutually exclusive projects must exercise caution since the alternative with the largest benefit-cost ratio will not necessarily produce the largest profit. It is possible that an alternative requiring a larger investment will produce a greater profit than an alternative with a larger benefit-cost ratio but requiring a smaller investment. Despite these shortcomings, the benefit-cost ratio is an excellent method for comparing the rate of return between many possible investments [6].

2.2 ITS Benefits Measurement

Benefits from ITS projects typically include reductions in delay, decreased accident rates, and lower levels of pollution. Since most benefits are not monetary it is often difficult to measure and directly compare the benefits of ITS projects. The ITS Joint Programs Office of the United States Department of Transportation (USDOT) has identified six metrics which it uses to measure the benefits of ITS projects. Termed "a

few good measures," the metrics are "robust enough to represent the goals and objectives of the entire ITS program, yet are few enough to be affordable in tracking the ITS program on a yearly basis" [4]. These six metrics track the magnitude of safety improvements, delay reduction, cost savings, effective capacity improvements, customer satisfaction, and energy and other environmental impacts.

Each metric is tied to one of the ITS program goals. The safety measure focuses on reducing the number of accidents and lessening the probability of a fatality if a crash does occur. The delay reduction metric measures the total travel time as well the variability of travel times. The cost metric generally measures the difference in cost between an ITS solution and the preexisting condition. It is more meaningful when applied to ITS programs that lower operating costs such as electronic toll collection. The effective capacity metric tracks the maximum potential rate at which vehicles can traverse a network. Since the true potential is difficult to measure, the throughput of a roadway is often used as a surrogate measure. The customer satisfaction metric attempts to quantify the public's happiness with the ITS implementation. It is often measured using surveys, questionnaires, or focus groups. Environmental impacts cannot be measured directly and must be estimated through simulation or other analysis. It is believed that, at least in the short term, smoother more efficient traffic flows will create a positive environmental impact. Long-term impacts, however, have not been measured and are not well understood [26].

2.2.1 ITS Benefits Evaluation

The non-monetary metrics used to measure the benefits of ITS projects make benefit evaluation difficult. The vast majority of ITS projects have shown positive

improvements indicating that some positive level of return is being generated by investments in ITS. However, without formal benefits analysis it is impossible to determine at what cost these improvements come. Additionally, it is impossible to compare the benefit per dollar of different ITS strategies or different implementations of the same technique.

The need for more rigorous ITS benefits analysis has been recognized for several years. In the paper "Models for Assessing the Impacts and Potential Benefits of Intelligent Transportation Systems," Ajay K. Ruthi of the Oak Ridge National Library states that the absence of a systematic analysis of the benefits and costs of ITS is well recognized in the ITS community [29]. This paper, written in 1995 for the Joint Programs Office (JPO) of the United States Department of Transportation, is the first call to action for the evaluation of ITS benefits. Although the paper lists several research organizations that are formulating benefits models, nothing yet has been produced that can accurately point Transportation Engineers towards efficient ITS designs.

This lack of benefit analysis, and consequential lack of design tools for ramp metering systems, was confirmed in December of 1998 when the Virginia Transportation Research Council (VTRC) performed its own thorough literature review of ramp metering systems. In the resulting publication, *Ramp Metering: A Review of the Literature,* VTRC senior research scientist E.D. Arnold Jr. states "although some attempts have been made to develop warrants, criteria, or guidelines for the implementation of ramp metering, few have been successful due to the many factors involved with ramp metering" [5].

Fortunately, a few researchers have recognized the need for thorough ITS benefits analysis and the lack of ramp metering system design tools. In their paper "Benefit/Cost Analysis on Ramp Metering: Optimization of Travel Time and Cost Factors," Haikun Dong and Bin Ran evaluate the benefit-cost tradeoffs that ramp metering system designers should be making. They developed a mathematical programming model with an objective function that considers throughput and cost while adhering to constraints that ensure the physical laws of traffic are upheld [13]. Although this model addresses a large need, it usefulness is diminished by the complex mathematics needed to evaluate the model, the deterministic nature of the model, and the lack of a high level description of the general shape of the tradeoff curve.

Vassilios Alexiadis and James Schmidt, supported by JHK & Associates, provided a look at the general shape of the benefit-cost tradeoff curve in their paper "Ramp Metering: A System Concept Design Methodology." During their research, Alexiadis and Schmidt performed traffic simulations on a 500-mile freeway network in California using four different ramp metering deployment schemes. Their empirical results showed that as the amount of equipment deployed increased, the benefit-cost ratio decreased [3]. This suggests that the tradeoff curve is concave down and that benefits rise quickly with lower budgets but do not increase at the same rate as cost at higher budgets. Although this is a significant finding, we must keep in mind that only four budgets were considered and no time was spent attempting to locate the optimal equipment locations or metering algorithm for each budget.

2.3 Microscopic Traffic Simulation

Traffic simulation is a powerful tool for modeling traffic flows along a roadway. It is useful for estimating the effects on traffic flow caused by events such lane closures due to construction and vehicle accidents or unusually high traffic due to an event such as a professional football game. Additionally, simulation can be used to predict the effects of proposed improvements such as lane additions, a new traffic light, and ramp metering.

Traffic simulations can be classified into two types: microscopic and macroscopic. In a microscopic simulation each vehicle is modeled as an individual entity. In a macroscopic simulation the high-level traffic flow is modeled. During each step of a microscopic simulation every vehicle will travel (accelerate, decelerate, maintain speed, change lanes, etc) based on its desired freeflow speed, desired following distance, and any upcoming exits. Sophisticated car-following logic controls each vehicle and directs its actions. In macroscopic simulations flow is modeled according to macroscopic traffic laws and parameters. Microscopic simulation is preferable for smaller scale simulations and simulations in which vehicle interactions, such as those found in merging and weaving sections, are significant. Macroscopic simulation is preferable when vehicle interactions are not significant or the roadway model is too large to make microscopic simulation practical [18, 21]

Developing a traffic simulation model is a very time-consuming task. The user must construct an accurate model of the roadway using blueprints, ariel photography, maps, or other data sources. Additionally, a large amount of traffic information such as the flow rate distributions of vehicles entering the simulation, vehicle destinations, and desired speed must be collected. Furthermore, the user must tune simulation parameters such as the roadway capacity and driver characteristics. Lastly, the user must compare

simulation output to actual traffic data in order to validate the simulation is an accurate representation of the real world.

Despite its high level of labor intensity, the benefits of traffic simulation are numerous. Results are often obtained more quickly and less costly with traffic simulation than by performing field tests. Also, traffic simulation does not cause the disruptions that are often associated with field experiments. Simulation is occasionally the only option when field tests are impossible or their disruptions are unacceptable. Furthermore, traffic simulation is often more useful than field experiments. Simulation allows the user to hold variables constant and isolate only the parameters of interest. This can yield insight into which variables are truly important and how these variables interrelate. Lastly, the flexibility of simulation allows a user to quickly evaluate many proposed solutions and test "what if" scenarios [18, 21].

2.4 Accident Rates

One negative aspect of traffic simulation is that accidents do not randomly occur during the simulation as they do on actual freeways. The user can simulate the effects of an accident by closing travel lanes, but the simulation will never spontaneously create an accident. One goal of ITS improvements is to lessen the likelihood and severity of a crash. Since accidents do not occur, traffic simulation alone cannot be used to evaluate the safety improvements of a potential ITS improvement.

When using traffic simulation to evaluate the benefits of a proposed improvement, a surrogate measure for accident rate must be used. The surrogate measures attempt to estimate the actual accident rate by using other traffic statistics. The estimators can be as

simple as the historical average or can be complex equations taking into account factors such as flow per lane, mean speed, variance of the speed, lane width, etc. [14].

The Oak Ridge National Laboratory uses a relatively simple accident rate model in its ITS Deployment Analysis System (IDAS). The model considers only the percentage of the freeway capacity being utilized (volume / capacity) and the total number of vehicle miles traveled. Despite using only two predictor variables, the model is more sophisticated than most since it separately predicts accidents involving fatalities, accidents involving injuries, and property damage only accidents. The complete model is listed below.

Fatalities: 0.004 Fatalities per Million Vehicle Miles (MVM)

Injuries: 0.5156 Injuries per MVM (if volume / capacity is less than 0.78)

0.5757 Injuries per MVM (if volume / capacity is between 0.79 and 0.88)

0.7329 Injuries per MVM (if volume / capacity is between 0.89 and 0.98)

0.7642 Injuries per MVM (if volume / capacity is greater than 0.98)

Property Damage Only: 0.8551 Property Damage Only per MVM

[2]

Prior research has shown that the variance of the speed is one of the most effective predictors of accident rate [15]. Using hourly speed data collected from data stations throughout the Virginia highway system, Nicholas Garber and Ravi Gadiraju employed linear regression analysis to create a model for the accident rate. The model obtained describes about 60% of the variance observed and has the form:

$$\text{ACCRT} = 43.2 + 0.00347(\text{SPVA})^2 \qquad \textbf{Equation 2.1}$$

Where

ACCRT = Number of accidents per 100 million vehicle miles
SPVA = Speed Variance

[15]

In 1999 Angela Ehrhart extended Garber's work by creating a more complex model to estimate the accident rate. Like Garber's study, Ehrhart also used hourly traffic data collected from data stations throughout the Virginia highway system. Using a heuristic process called the multivariate ratio of polynomial procedure, Ehrhart generated the following model (Ehrhart, 52):

CRASHRATE = ((-0.4468269) – (3.13093E-03)*(SD*SD) – (1.469674E-06)*(SD*SD)^2 + (2.797139E-07)*(FPL*FPL) – (6.315968E-10)*(SD*SD)*(FPL*FPL) + (2.384377E-14)*(FPL*FPL)^2 + (3024.788)*(1/(MEAN*MEAN)) + (15.15044)*(SD*SD)*(1/(MEAN*MEAN)) – (5329379*(1/(MEAN*MEAN))^2)

$$\textbf{Equation 2.2}$$

Where

SD   = Speed Variance
FPL   = Flow Per Lane
MEAN= Mean Speed

[14]

This model has a R^2 value of 0.5816 and an Akiake information criterion (AIC) of –475.68. However, given the complexity of the equations and number of predictor variables, it seems likely that this model may be overfit to the training data used to generate the model.

2.4.1 Model Selection

All three models presented above were implemented in the simulation model. The estimates of the accident rate generated by each of the three models were compared to each other and to the historical accident rate in the test area.

Ehrhart's model was eliminated due to the poor quality of its results. When traffic data generated by the simulation run were used as the input, the model predicted a negative accident rate. Since negative accident rates are impossible, the model was no longer considered.

The results returned from the ORNL model and Garber's model were similar, although the ORNL prediction tended to be higher. However, when the standard deviation of the speed was around twelve miles per hour, the two models predicted the same accident rate.

The ORNL model was discarded since it is contradictory to historical accident data . The ORNL predicts that the accident rate will increase as use of the road reaches capacity. However, field results show that ramp metering not only brings traffic flows closer to their theoretical capacity, it also decreases the accident rate. Therefore, the ORNL model predicts that implementing a ramp metering system will increase the accident rate although historical data clearly shows that this is not the case.

Before implementing Garber's model as the predictor for accident rate, predictions generated by the model were compared to historical accident data. Simulations using actual historical traffic data were performed and Garber's model was used to predict the number of accidents that would occur between 6:30AM and 8:30AM on weekday mornings on the test section of road. Garber's model predicted 16.5

accidents during the course of one year, an average of 1.375 per month. The Smart Travel Lab's incident database was queried to see the actual number of accidents that occur on this stretch of road between July 1998 and June 1999. Except during May of 1999, whose six accidents were at least twice as high as any other month, the average number of monthly accidents was 1.55. The May data appear to be an anomaly and may be due to construction, severe weather, or other factors. However, even if the May data are included, it is obvious that the prediction generated by Garber's model is satisfactorily close to actual accident levels. Therefore, Garber's model will be used to predict the accident rate in all future simulation studies.

2.5 Ramp Metering Algorithms

In addition to the simple clock-time metering algorithm, five traffic responsive metering algorithms were implemented as control policies for the ramp meters. The five algorithms are speed control, occupancy control, demand-capacity control, ALINEA, and a bottleneck algorithm. The following sections detail the logic embedded in each of these traffic responsive algorithms.

2.5.1 Speed Control

The speed control algorithm utilizes real-time speed data collected from the vehicle detectors. Measurements from the immediate upstream detector are often used, but data from downstream detectors can also be used.

The speed measurements are compared to a preset lookup table. Table 2.1 provides an example of a speed lookup table. The measurements are compared to the values in the table and the appropriate metering rate is applied. If the designer desires a

more precise return from the lookup table, the metering rate can be interpolated if the speed value falls between two thresholds.

Speed (mph)	Metering Rate (Seconds per vehicle)
≥ 60	4
50 - 59	6
40 - 49	8
35 - 39	10
≤ 35	12

Table 2.1: Speed Control Lookup Table

2.5.2 Occupancy Control

The occupancy control algorithm is extremely similar to the speed control algorithm. Real-time traffic measurements are again taken from the immediate upstream sensor and compared to a lookup table. However, in this case the metering rate decreases as the occupancy of the mainline increases. Table 2.2 provides an example of a Occupancy control lookup table.

Occupancy (%)	Metering Rate (Seconds per vehicle)
≤ 10	4
11 – 16	6
17 – 22	8
23 – 29	10
≥ 29	12

Table 2.2: Occupancy Control Lookup Table

2.5.3 Demand-Capacity Control

The demand-capacity algorithm attempts to prevent the demand downstream of the entrance ramp from exceeding capacity. The current demand is measured using vehicle detectors placed upstream of the entrance ramp. The metering rate is determined

by subtracting the current upstream demand from the maximum downstream capacity. For example, assume downstream capacity is 7200 vehicles per hour (vph). If the upstream vehicle detectors record a demand of 6500 vehicles per hour, then the meter will release vehicles at the rate of 700 per hour.

In times of unusually high or low demand, the minimum metering rate is often set to 240 vehicles per hour and the maximum metering rate is often set at 900 vph. Releasing fewer than 240 vehicles per hour usually leads to excessive delays and the vehicles will often become impatient and proceed through a red light. On the other hand, releasing over 900 vehicles an hour leads to a cycle time of less than four seconds, and it is difficult for vehicles to keep up with the signal.

2.5.4 ALINEA Algorithm

Asservissement Linéaire d'entrée Autoroutière (ALINEA) is a local feedback ramp metering strategy developed by M. Papageorgiou, H. Hadj-Salem, and F. Middelham. The algorithm attempts to keep the mainline occupancy near at an optimal rate known as the "critical occupancy." However, unlike most traditional metering algorithms, ALINEA uses the occupancy measurement as well as the metering rate during the previous period when calculating the metering rate for the current period. This allows the ALINEA algorithm to react smoothly to large differences and respond to small fluctuations in the occupancy measurement. The metering rate is for any given period is calculated as follows:

$$r(k) = r(k-1) + K_R[\hat{o} - o_{out}(k)]$$ **Equation 2.3**

Where

$r(k)$ = number of vehicles to be released this period

r(k-1) = actual ramp volume last period
K_R = regulatory parameter (units: vehicles per hour)
ô = critical occupancy
$o_{out}(k)$ = most recent occupancy measurement

[23]

2.5.5 Bottleneck Algorithm

Unlike the four local ramp metering algorithms described above, bottleneck algorithms are used in system-wide ramp metering. Instead of attempting to optimize traffic flow in the meter's local area, multiple meters are controlled with the objective of optimizing the flow through a downstream bottleneck. Although many different control equations can be used, almost all attempt to ensure that the demand entering the bottleneck area does not exceed the capacity of the bottleneck. The number of vehicles released from entrance ramps added to the upstream mainline demand must be less than the capacity of the bottleneck. This algorithm often causes upstream meters to have more restrictive metering rates than they would if a local metering algorithm were used [33, 35].

2.6 Pseudo-Random Search

Many problems exist in which a researcher is attempting to maximize or minimize an objective function. In general, there exists an objective function $\mathbf{f(x)}$ that translates decision variables into an objective function value. If $\mathbf{f(x)}$ is known, a maximum or minimum value can be located through the use of calculus or more sophisticated programming methods. However, cases exist where the objective function is in the form of a "black box" and the analytical description of the function is unknown. An example of this situation is a simulation model.

Several different methods are available that attempt to find the set of decision variables which optimizes the value of the objective function. If time were not a constraint, all possible solutions would be evaluated. The solution with the highest (or lowest) value of the objective function would easily be identified as the optimal solution. Unfortunately, time constraints often cause a complete search of the state space to be infeasible. Another method is to test a randomly selected subset of all possible solutions. Although it cannot be stated with complete confidence, the solution in the subset with the highest (or lowest) value of the objective function could be considered the optimal solution. This method requires less time than a complete search, but drastically reduces the probability of finding the optimal solution. A third option is to use a pseudo-random search method. Pseudo-random search methods attempt to locate the optimal solution without searching the entire search space. These methods use the objective function value of previously evaluated solutions to determine where to search next. The idea is to evaluate solutions that are similar to, or in the same neighborhood as, previously tested solutions that performed well.

Simulated annealing is one proven pseudo-random search technique. In order to use simulated annealing it must be possible to represent the decision variables as a string of binary digits. The first step is to generate an initial solution. The initial solution, named S1, can be randomly generated or based on a good guess. The initial solution is evaluated and the objective function value, or measure of merit, for this solution is termed MOM(S1). This solution is then randomly mutated by probabilistically flipping the value of each digit in the binary string. The probability of a flip should be low such that only relatively small changes are made to this solution. The new solution, termed

S2, is evaluated. If S2 is an improvement over S1, it is always accepted and S2 becomes S1. If S2 is not an improvement over S1, then it is still accepted with a probability based on the difference between MOM(S1) and MOM(S2) and the value of a variable called the temperature.

This process is then repeated. S1 is mutated in order to form a new string called S2. The new string is evaluated and accepted if MOM(S2) is greater than MOM(S1). If the measure of merit is not greater, then S2 may still be accepted with some probability. This cycle continues until a stopping criterion is reached.

The purpose of probabilistically accepting an inferior solution is to avoid getting stuck in a local maximum. Often, steps in the wrong direction are required to locate the global maximum. If S2 is inferior to S1, the probability of still accepting S2 is:

$$p = \exp(delta/TEMP_k),$$ **Equation 2.4**

Where
delta	=	MOM(S1) – MOM(S2)
$TEMP_k$	=	value of the temperature at the current step (step k)

[22]

The temperature is a parameter initially set by the user and reduced each iteration. The higher the temperature, the greater the probability of accepting an inferior solution. Thus, early in the search the probability of accepting an inferior solution is relatively high, but late in the search this probability approaches zero.

The search ends when a stopping criterion is met. The stopping criterion may be a maximum number or iterations or may be a number of iterations in which no improvement is found.

Simulated annealing is more efficient than random searches since the search focuses around solutions that have performed well. This focus increases the probability of finding the "optimal" solution without testing the entire state space. Additionally, simulated annealing is more robust than a strict uphill search since it probabilistically accepts inferior solutions. This keeps the search from getting stuck in a local optimum and increases the probability of finding the global optimal solution.

Chapter 3: Methodology

3.1 Overview

Figure 3.1 provides an illustration of the circular simulation-optimization methodology used to create the optimal tradeoff curve. The curve is generated by determining the maximum benefits possible from a variety of budgets. At each budget level the "optimal" deployment scheme is located using a pseudo-random search technique. All deployment schemes tested by the search technique are evaluated via traffic simulation.
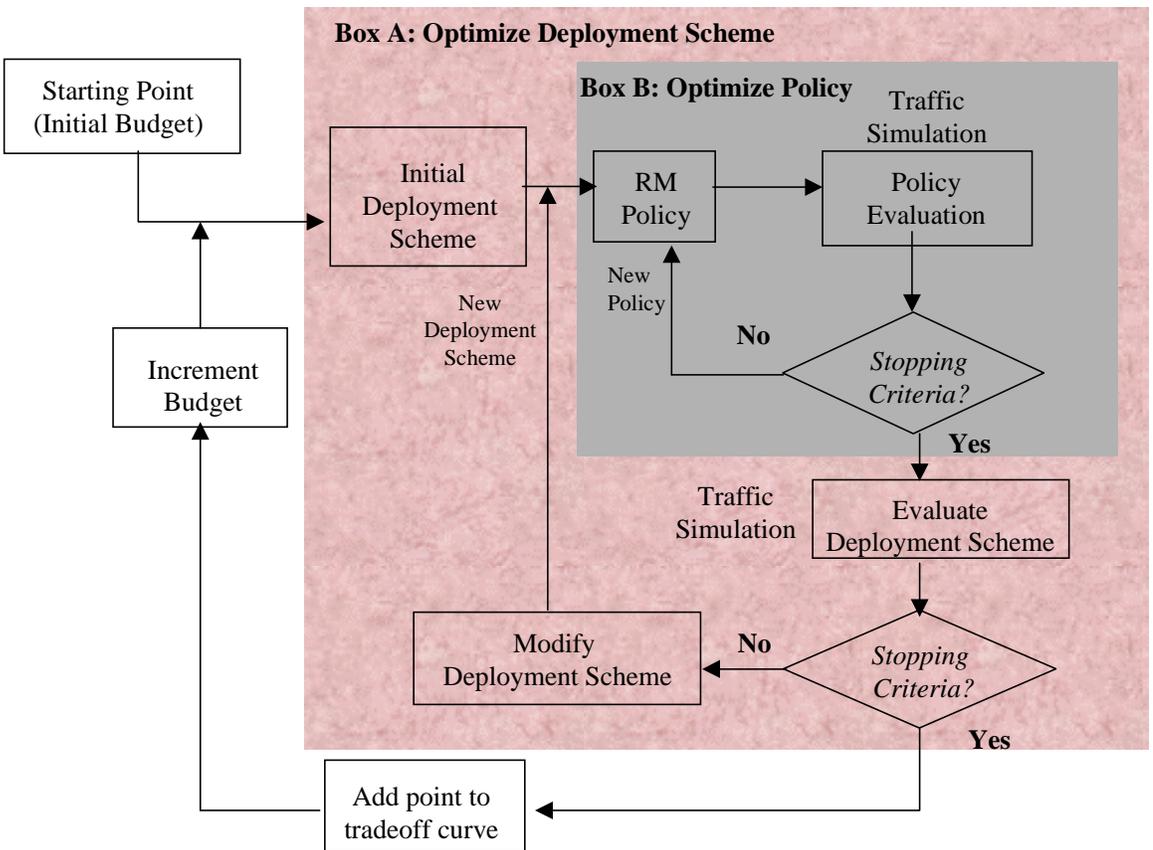
Figure 3.1: Simulation-Optimization Methodology

Within each budget level, the search technique attempts to locate the deployment scheme that produces the largest benefits by exploring many different possible schemes (this occurs in box A). The type of metering algorithm and algorithm parameters used to control ramp metering equipment has a large effect on the level of benefits returned by the deployment scheme. Therefore, for each deployment scheme it is necessary to determine which ramp metering algorithm and parameter values generate the maximum benefit from the scheme (this occurs in box B). Once the best metering algorithm and parameter values are located, traffic simulation is used to evaluate the performance of the deployment scheme. Several different traffic scenarios are used to ensure the robustness of the scheme.

After the scheme is evaluated, it is mutated to form a new deployment scheme that meets the budgetary constraints. This new scheme is optimized, evaluated, and modified in the same manner as the previous scheme. This cyclical process continues until the pseudo-random search technique has located an optimal equipment deployment scheme or a stopping criterion is reached. If the stopping criterion is reached, the algorithm will consider the deployment scheme that generated the largest benefits to be the optimal scheme.

Once the optimal deployment scheme for a given budget is located, a point is added to the tradeoff curve. This point indicates the maximum benefits that can be achieved for that budget level. At this point the budget is incremented by a fixed amount and the above activities are repeated until an optimal scheme is located for this new budget. The maximum benefit level for this incremented budget is also marked on the

tradeoff curve.   This process continues until there are enough points that can be connected with a smooth curve to display the optimal tradeoff frontier.

3.2 Decision Variables

The two technologies that can be deployed are ramp meters and vehicle detectors. A ramp meter may be located at each entrance ramp to the freeway. If no meter is present, vehicles continue onto the freeway without delay. If a meter is present, vehicles are released onto the highway one at a time. The rate vehicles are released is determined by the metering algorithm used.

In order to bring the number of possible deployment schemes to a manageable number, the mainline has been broken into discrete zones one-half mile long.   Each mainline zone (one-half of a mile in length) may contain one vehicle sensor. Traffic data gathered by the vehicle detectors are communicated to ramp meters allowing the use of traffic responsive metering algorithms.

The decision variables are binary digits that indicate whether a technology is deployed at each potential location. Therefore, a decision variable is needed for each ramp. If the decision variable has a value of one, a ramp meter is deployed on this entrance ramp. If no meter exists, the decision variable is given a value of zero. The same representation is used for vehicle detectors. If a mainline zone contains a vehicle detector, the corresponding decision variable has a value of one. If no detector exists, the decision variable has a value of zero.

When these binary decision variables are placed in order from the furthest upstream point of the freeway to the furthest downstream, a binary string is created that completely defines the deployment scheme.   Figure 3.2 provides an illustration of

representing a section of roadway as a binary string. The scheme in this figure can be represented by the string "1011."
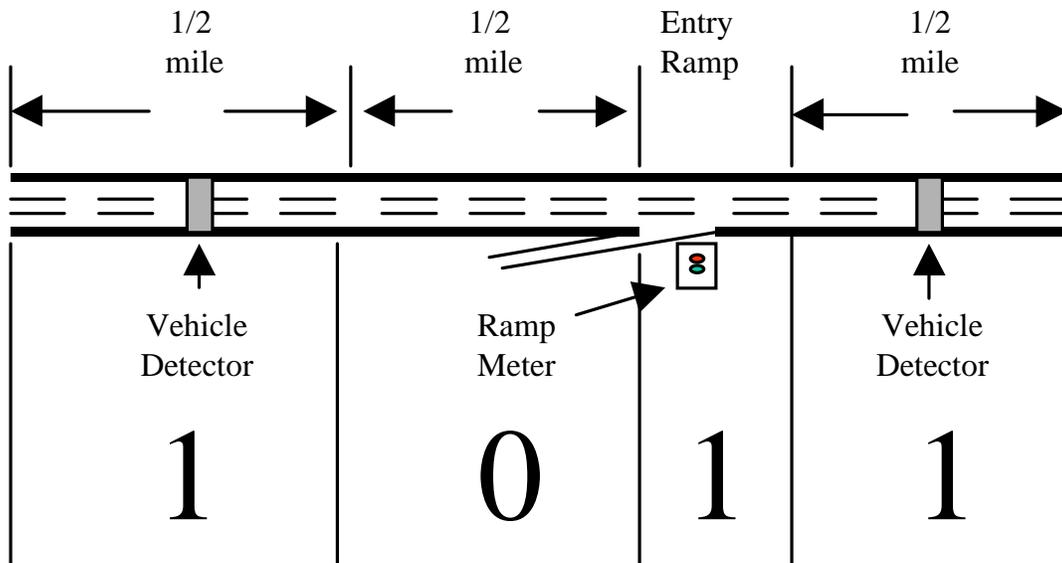
Figure 3.2: Binary Representation

3.3 Traffic Scenarios

The performance of all deployment schemes is evaluated in a variety of different traffic scenarios. This is done to ensure the robustness of each scheme and to more accurately model the demand fluctuations experienced by freeways.

Three distinct scenarios are simulated: normal demand, heavy demand, and accident occurrence. The traffic demand in the normal scenario is the actual traffic demand experienced on the test days. The traffic demand experienced in the heavy scenario are higher than the demand used in the normal scenario. The arrival rate at each entrance ramp is increased by twenty percent, and the arrival rate at the upstream mainline entrance of the freeway is increased by ten percent.

The accident scenario is modeled with two different simulation runs. Each run has the same traffic demand as the normal scenario, but contains a fifteen minute incident on a three-lane section of road. During the first ten minutes, the left-most lane is completely closed and the middle lane has a reduced capacity due to rubbernecking. During the final five minutes the middle lane is back to full capacity, but the left-most lane is partially blocked due to rubbernecking and/or debris on the road. In the first accident simulation run the incident is located in the first one-third of the modeled freeway. In the second accident run the incident is located in the final one-third of the modeled freeway. The different locations are needed to ensure that certain deployment schemes are not favored by the location of the accident.

Each deployment scheme will by evaluated via simulations of all four traffic conditions (one normal, one heavy, and two accident). The performance during each scenario will be used to calculate a total performance score for the scheme.

3.4 Measures of Performance

The USDOT uses six measures of performance (MOP's) to compare the benefits of ITS projects. These six metrics track the magnitude of safety improvements, delay reduction, cost savings, effective capacity improvements, customer satisfaction, and energy and other environmental impacts. The delay reduction, effective capacity improvement, customer satisfaction, and energy and environmental impact metrics are all highly correlated. An increase in effective capacity reduces delay which leads to more satisfied travelers, less wasted gasoline, and less pollutants outputted by the vehicles.

To avoid redundancy, this study will use three independent measures of performance instead of the six correlated metrics used by the USDOT. The three MOP's

are average system speed, accident rate, and cost. The average system speed is the average speed of all vehicles, in all parts of the model (including entrance and exit ramps), throughout the duration of the simulation. This metric is easy to calculate and accurately measures congestion on the freeway and the delays being experienced by travelers. The accident rate is a good measure of the safety improvements produced by the deployed equipment. The accident rate is estimated using Garber's model which was discussed in the literature review. The cost of each deployment scheme includes the funds needed to install and maintain the ramp meters, vehicle detectors, and communications hardware. Maintenance costs are calculated assuming a five-year lifecycle. A complete description of the rules used to calculate costs is provided in Appendix A.

Many ramp metering systems are part of a larger Advanced Traffic Management System (ATMS). Traffic information is valuable to an ATMS because it may be used for incident detection, traffic predictions, or simply to alert commuters of the present traffic conditions. Therefore, the quantity of traffic information collected by the deployment scheme is the fourth, and final, MOP. The amount of information gathered is based on the number of vehicle detectors in the deployment scheme.

Chapter 4: Implementation

4.1 Simulation Model

CORSIM, part of ITT Industries' Traffic Software Integrated System (TSIS) suite, was chosen as the traffic simulation package used to evaluate all equipment deployment schemes. A combination of ITT's old FRESIM and NETSIM models, CORSIM is a well documented and widely used microscopic simulation package. A significant portion of the benefits generated by ramp meters comes from smoothing out the randomness of traffic flows on the mainline and vehicles entering the freeway. Therefore, the car-level detail provided by a microscopic simulation package such as CORSIM is needed to accurately assess the benefits of ramp metering.

Other reasons for choosing CORSIM include its built-in ramp metering functionality and Run-Time Extension capability. The built-in ramp metering allows the user to easily implement basic ramp metering algorithms. The Run-Time Extension (RTE) is a dynamically linked library (DLL) written by the user which interfaces with COSRIM every second. The RTE allows the user to implement more sophisticated ramp metering algorithms and is also helpful for extracting performance data from the simulation.

4.1.1 Test Area

The test area used for this research is a six and one-half mile stretch of Interstate 64 in the Hampton Roads area of southeastern Virginia. The westbound travel lanes from 3,300 feet south of the Military Highway Interchange to the Bay View Boulevard crossing are modeled. The roadway contains seven entrance ramps and eight exit ramps.

The six and one-half miles of freeway are partitioned into fourteen mainline zones, each approximately one-half mile long. Describing the state of the seven entrance ramps and fourteen mainline zones requires twenty-one binary decision variables. Therefore, there are 2^{21}, or 2,097,152, possible deployment schemes. Figure 4.1 provides a diagram of the test area.

End

Ramp 7

Ramp 6

N

W   E

S

Ramp 5

Ramp 3

0   0.5   1.0 miles

Ramp 4

Westbound Lanes of I-64
Hampton Roads, Virginia

Ramp 2

Ramp 1

Start
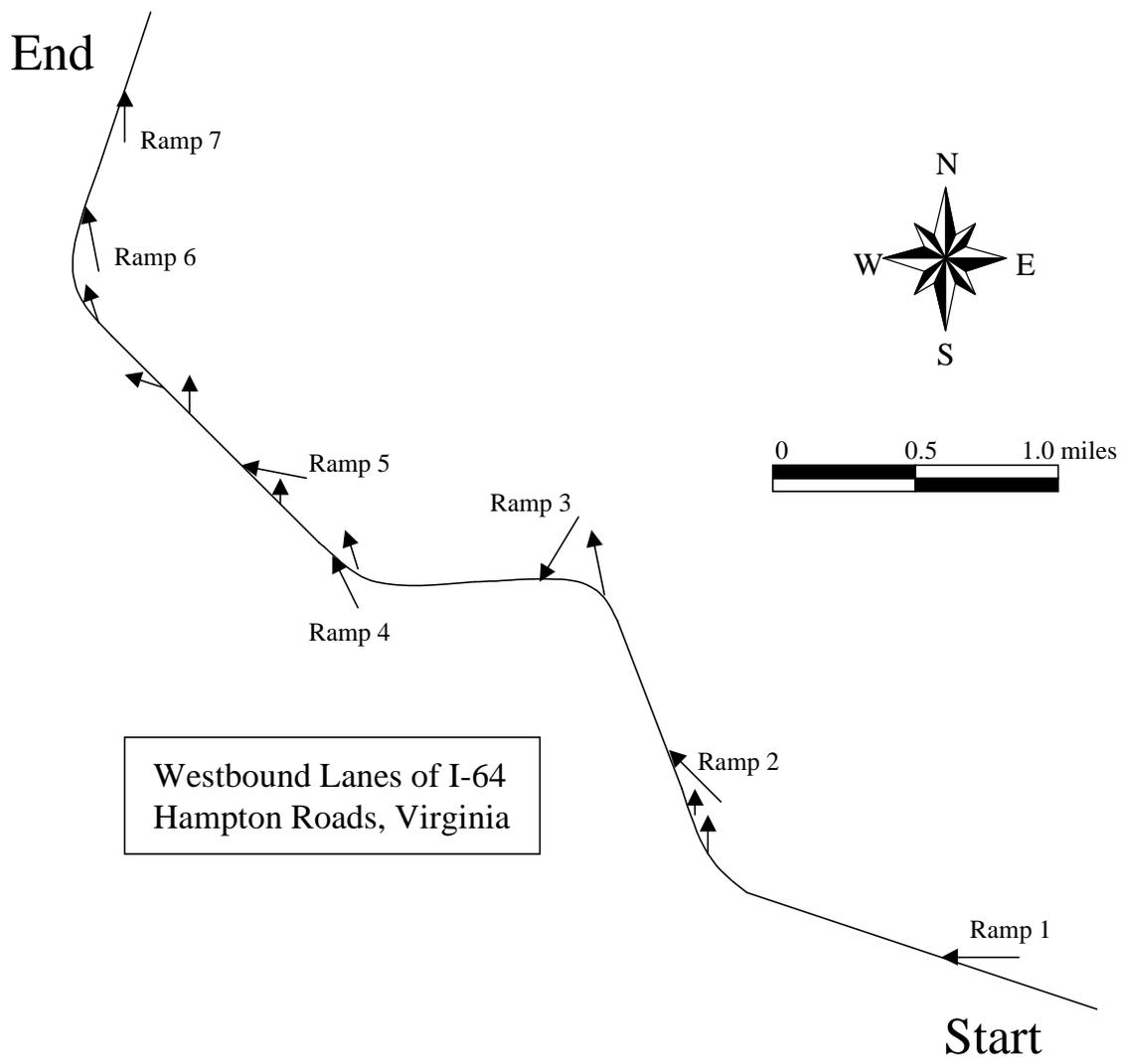
Figure 4.1: Diagram of the Test Area

Figure 4.2 provides an example of representing a deployment scheme as a binary string. Each digit of the binary string represents an entrance ramp or a section of road that may contain a vehicle sensor. If the area contains a piece of equipment the corresponding digit receives a value of one. If no equipment is located in the area, the digit is set to zero.
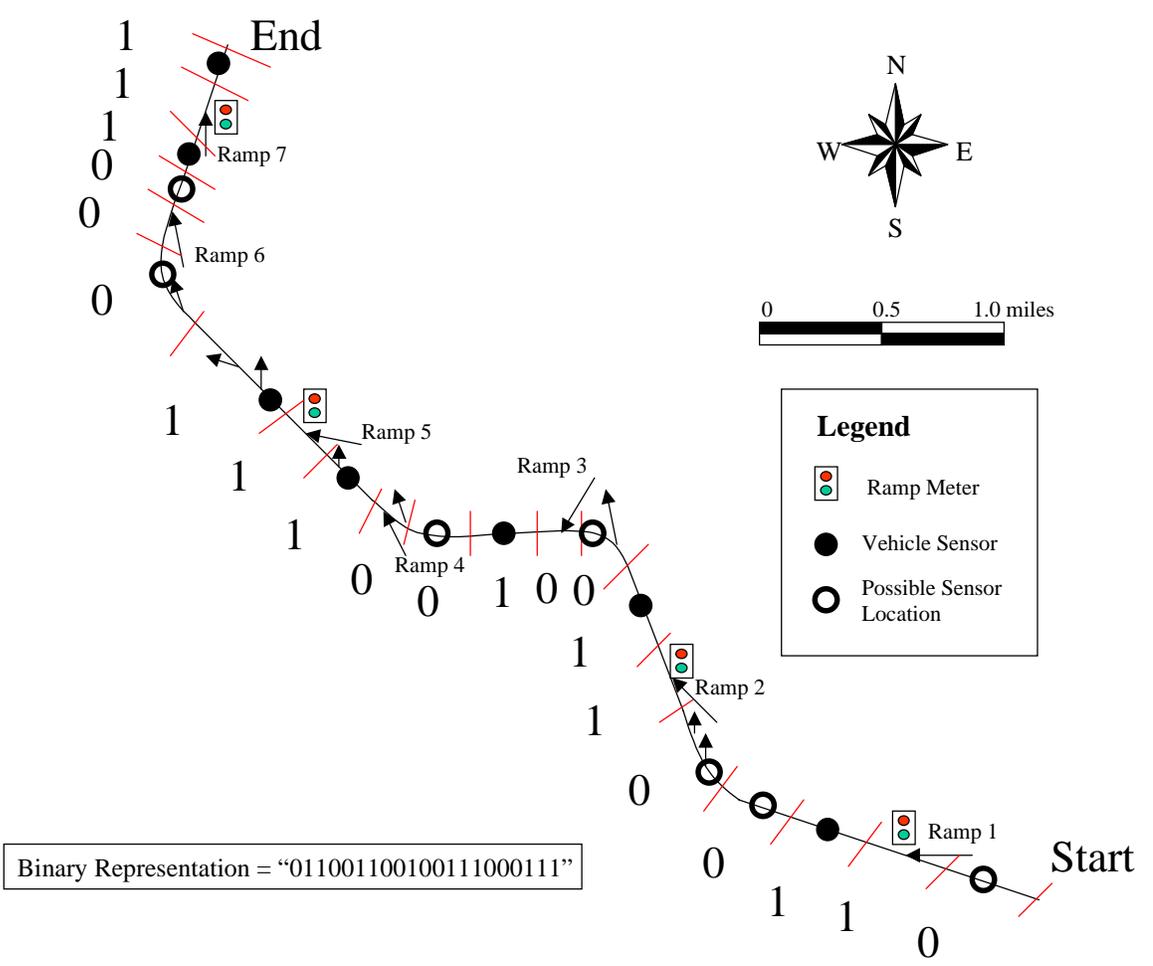
Figure 4.2: Binary Representation of a Deployment Scheme

4.1.2 Test Time and Duration

Traffic patterns on this section of I-64 are largely influenced by the United States Naval base located on the Elizabeth River. The westbound travel lanes lead to the Naval Base and the freeway exit for the base is located near the western-most end of the model. The flow of commuters traveling to the base causes the morning traffic levels to be much higher than the afternoon traffic levels.

Ramp metering is most useful when implemented on roadways with high levels of congestion. In practice, ramp meters are only turned on during the morning and/or evening rush hours when traffic levels are at there highest. Since the modeled section of roadway has it highest levels of congestion in the morning, the simulation runs use morning traffic data.

If ramp meters were implemented on this stretch of freeway, they would likely be operating every weekday during the morning rush hour. Rush hour, which actually tends to last more than one hour, is the time of day when increased traffic demand leads to reduced speeds and delays for travelers. If simulation time were not a constraint, it would be logical for the simulation runs that evaluate each deployment scheme to last the entire duration of the morning rush hour. However, simulating this entire rush hour period is quite time consuming and requires nearly ten minutes per simulation run.

One timesaving solution is to assume that the performance of the deployment scheme can be accurately evaluated without simulating the entire morning rush hour period. It is reasonable to assume that the performance of the deployment scheme during a subset of the rush hour should be highly correlated to the performance of the deployment scheme during the entire rush hour period. Therefore, instead of modeling

the entire rush hour, each run simulates the portion of the morning rush hour from 6:30AM to 6:50AM. This twenty-minute span was chosen because it has a higher demand than any other period during the day. The effects of modeling a twenty-minute time period instead of the entire morning rush hour are analyzed and discussed in the "testing assumptions" section of Chapter 6.

4.1.3 Input Data

The traffic data used to populate the simulation runs were obtained from the Smart Travel Laboratory at the University of Virginia. The Smart Travel Lab possesses a data warehouse that contains daily traffic data dating back to June of 1998. Originally, historical averages of daily weekday volumes from 6:30AM to 6:50AM were going to be used. However, once implemented, the ramp metering produced no benefits when the simulation used these historical averages of traffic data. The averaging process smoothed most of the randomness that is found in a single day's data. As mentioned before, smoothing this randomness is one of the features that cause ramp meters to produce benefits. Additionally, the historical averages tended to be lower than the actual traffic demand found on single days. The lower average traffic flow is likely caused by days of reduced travel due to holidays and severe weather events.

Since historical averages do not replicate actual daily traffic conditions, the simulation runs had to be populated with traffic data from individual days. Wednesday, August 11, 1999 and Monday, Aug 23, 1999 were chosen as the two test days. All traffic simulations use actual traffic data from one of these two days. The days were chosen because the traffic on these days followed the "normal" traffic pattern on the freeways of the Hampton Roads region of Southeastern Virginia. The Smart Travel Laboratory has

scrutinized weekday traffic flows in the area and traffic demand on these two days follows the normal pattern. The plot of each day's traffic data appeared as expected and neither day had reduced traffic flows due to weather, construction, accidents, or holidays. Additionally, most of the vehicle detectors were operational and each day had extremely few data points removed by the Smart Travel Lab's data filters.

All simulation runs used to evaluate a performance scheme use data from the same test day. Before the first run the control program randomly decides which day's data (August 11 or August 23) will be used in the simulation runs. It was assumed that two different days were enough to ensure the shape of the performance-cost curve would not be skewed by the peculiarities of a single day's data. Chapter 6 tests this assumption and examines in closer detail what effect the input data has on the performance of the deployment schemes and the shape of the tradeoff curve.

4.1.3.1 Volumes and Turn Percentages

CORSIM requires the user to specify the volume at all entrance points to the model and the turn percentage at all points in the model where vehicles have the opportunity to exit. The volumes for each entrance ramp and at the mainline entrance to the model are the actual volumes measured during the two test days and stored in the Smart Travel Laboratory's database.

The turn percentage data is calculated directly from the volumes in the test data. Each exit ramp in the test area has a sensor (which contains a vehicle detector in each lane) on the exit ramp, immediately upstream of the exit ramp, and immediately downstream of the exit ramp. As long as two of the three sensors are functioning correctly, the percentage of vehicles that exit can be easily calculated. Historical

averages are used for turn percentages in the two locations where failed sensors make the calculation of turn percentages impossible.

The data stored by the Smart Travel Lab are aggregated into two-minute intervals. Thus, every two minutes during the simulation runs both the input volumes and turn percentages are updated to reflect the changing traffic demand.

4.1.3.2 Arrival Distribution

CORSIM requires the user to enter a statistical distribution to model the vehicle entry headways. The user can choose between the uniform distribution, normal distribution, or Erlang distribution. The traffic data in the Smart Travel Laboratory contains volume counts, but does not include any information regarding the distribution of the vehicle arrivals. Additional research and data collection needed to be performed in order to determine which distribution most accurately modeled the actual arrival distribution.

Inter-arrival times were collected and analyzed during two morning rush hours in February of 2000. Based on this analysis, the Erlang distribution with a shape parameter (K) value of two was selected to model the inter-arrival times. Of the three distributions that CORSIM makes available, the Erlang distribution with K equal to two most accurately models the actual inter-arrival times. Appendix B provides more detail on the data collection and analysis procedures. Additionally, Chapter 6 explores how the arrival distribution affects the performance of deployment schemes.

4.1.4 Model Parameters

CORSIM's internal logic contains many parameters that influence the behavior of individual drivers and affect the flow of traffic as a whole. Prior research has shown that

using the default values for these parameters often leads to simulations that do not closely mimic the actual traffic pattern of the roadway [11, 36]. Therefore, a calibration process was performed to determine what parameter values would lead to simulation runs that most accurately modeled actual traffic flows.

Model parameters were iteratively varied and evaluated with a training set of traffic data. Each parameter set was evaluated by calculating the percentage deviation between the volume data and speed data from the simulation and actual measurement found in the training data. The process continued until the parameter set which minimized the deviation between the simulation traffic measurements and actual traffic measurements was found. This "best" parameter set was then evaluated against a test set to generate an unbiased estimate of how well the actual traffic flows were modeled. Appendix C contains a complete description of the calibration procedure.

The calibration process resulted in the following six actions:

1. Freeflow speed set at 70 mph on all mainstream links (default is 55 mph).

2. Car following aggressiveness set to 1.0 – 0.1 with a Pitt Car following constant of 5 feet (defaults are 1.5 – 0.6 and 10 feet).

3. Time required to make a lane change set at one second (default is 3 seconds)

4. Courtesy factor (percentage of drivers willing to yield the right of way to vehicles attempting to merge in front of them) set at 40% (default is 20%)

5. Location of one exit warning sign moved to two thousand feet before the exit (default location is twenty-five hundred feet before the exit).

When the final parameter set was evaluated against the test data set, the average speed error was 2.5% and the average absolute speed error 24.9%. The average volume error was 3.1% and the average absolute volume error was 11.6%. Although the speed error is higher than desired, the error levels are comparable to error levels experienced by past researchers. In his 1998 paper, Cheu describes using a genetic algorithm to test over four hundred different parameter sets for a FRESIM model of a freeway in Singapore. Even with all these tests he was never able to reduce the average speed error below 11% or the average volume error below 30%.

4.1.5 Ramp Metering Implementation

The ramp metering algorithm used to determine the metering rate can have a large effect on the performance of a ramp metering system. Fortunately, the Run-Time Extension provided with CORSIM allows the user to implement any ramp metering algorithm, or combination of algorithms he chooses. The following algorithms were implemented and available to ramp meters during the generation of results: clock-time metering, speed control, occupancy control, demand-capacity control, and ALINEA.

Although each algorithm uses different logic to calculate the metering rate, all the rates have some common attributes. In all algorithms traffic data is aggregated into thirty-second periods, and the metering rates are updated every thirty seconds using the previous period's data. Each algorithm contains parameters that must be tuned in order to maximize the performance of the algorithm. Additionally, each algorithm contains a queue override feature to ensure the queue of vehicles does not spill back onto arterial streets. An advanced queue detector measures the length of the queue, and all algorithms will increase the metering rate if the queue is in jeopardy of backing up onto city streets.

The extremely large number of possible deployment schemes and the sizeable number of algorithms available to control each scheme makes searching the state space a daunting task. Two assumptions were made that reduces the time required to search the state space. The first assumption allows similar deployment schemes to be placed in equivalence classes which receive a single performance score. The second assumption greatly reduces the amount of time required to locate the optimal control policy for each deployment scheme. The following sections discuss each assumption in more detail.

4.1.5.1 Assumption: Equivalence Classes

The methodology described in Chapter Three states that time will be spent locating the best policy to govern each deployment scheme. If you refer back to Figure 3.1, this procedure takes place in box "B." With as many as fourteen different data sensors to choose from it would, however, be extremely time consuming to determine which sensor, or combination of sensors, should provide the data needed by a single ramp meter for a traffic responsive metering algorithm. Given that a single deployment scheme can include up to seven ramp meters, it could take hundreds of test runs to determine the optimal policy.

In practice, a sensor located immediately upstream or downstream of the ramp meter usually provides the data required by the traffic responsive metering algorithm. It is wasteful to spend a large amount of time considering all available sensors, when the best performing policy most often uses a sensor in close proximity to the ramp meter. The time spent testing all available sensors simply is not justified by the small increase in expected performance. Therefore, when being controlled by a traffic responsive algorithm, ramp meters only use data from a sensor located immediately upstream or

downstream from the meter. This assumes that the increased performance from considering all sensors is insignificant and is not worth pursuing.

This assumption greatly limits the number of control policies that need to be evaluated for each deployment scheme. The majority of sensors have one upstream sensor and one downstream sensor. Therefore, there are only four possible states for each entrance ramp:

0. No meter available.

1. Meter controlled by a clock-time algorithm.

2. Meter controlled by a traffic responsive algorithm using data from the immediate upstream sensor.

3. Meter controlled by a traffic responsive algorithm using data from the immediate downstream sensor.

The benefits derived from this assumption are twofold. First, considering fewer sensors greatly decreases the time required to determining the best control policy. Second, having a discrete number of states for each ramp allows deployment schemes to be placed into an equivalence class. A single equivalence class contains many different schemes that are controlled in the same manner and should generate the same level of benefits.

For example, consider a deployment scheme in which only the first ramp contains a meter. Additionally, assume that the only sensor in this scheme is immediately upstream of this ramp. Obviously, the optimal policy for this scheme would be to control the first ramp with a traffic responsive algorithm that receives data from the sensor

immediately upstream. Now consider a second deployment scheme, which is identical to the first, except that it has a second sensor located five miles downstream of the ramp meter. Most likely, the optimal policy for this scheme would also be to control the first ramp with a traffic responsive algorithm that receives its data from the sensor immediately upstream. These two deployment schemes are in the same equivalence class, since they have identical locations of ramp meters and are controlled by the same policy.

Each equivalence class contains dozens of different deployment schemes. In fact, while there are 2.1 million different deployment schemes, there are only 15,360 different equivalence classes. And since every scheme in an equivalence class should generate the same level of benefits, there is no need to test more than one scheme from each equivalence class.

4.1.5.2 Assumption: Algorithm Choice

While the equivalence class assumption reduces the amount of time needed to determine the optimal policy for a deployment scheme, the task is still far from trivial. After a sensor is selected to provide data for a traffic responsive algorithm, time must be spent determining which algorithm generates the best results. This process is made more complex by the fact that each algorithm contains parameters that can greatly influence the performance of the algorithm. An exhaustive search would first determine the optimal parameter values of each algorithm and then compare the performance of the algorithms. With four traffic responsive algorithms under consideration and no less than five appropriate parameter values for each algorithm, it is obvious that an exhaustive search would require a tremendous amount of time.

Since the traffic responsive metering algorithms receive data from sensors located near the meter, they are effectively attempting to optimize traffic flows in the meter's local area. Thus, it is reasonable to assume the same metering algorithm and parameter values will perform best regardless of whether other entrance ramps have meters on them. If the same algorithm and parameter values are always optimal for a meter, then it is unnecessary to spend time determining the best policy for each equivalent class. The best performing algorithm for each state of each meter can be determined ahead of time and implemented whenever the state occurs.

Appendix D contains the results of the traffic responsive algorithm evaluation tests. The top performing traffic responsive algorithms (with appropriate parameter values) for each state is provided for each of the seven ramp metering locations. States zero and one are not included in Appendix D, since state zero indicates no meter is present and state one indicates clock-time metering.

It is also assumed that the same clock-time metering parameter value will be optimal regardless of the presence and control policy of other meters. Appendix E contains the results of the clock-time algorithm evaluation results.

With these assumptions in place, there is no longer a need to determine the optimal policy for each deployment scheme. The optimal policy can be determined from the algorithm evaluation test results located in Appendices D and E. Each meter will use a traffic responsive metering algorithm as long as one of the local sensors is available. If multiple local sensors are available, the meter will receive data from the sensor that generated the best test results. The clock-time algorithm parameter values will also be selected based on the results of the algorithm evaluation tests.

For example, consider a deployment scheme in which a meter exists on the second entrance ramp. As Appendix D indicates, if the immediate upstream sensor is available, it will be used. In this case an occupancy control algorithm with lookup-table 1 will control the meter. If this sensor is not available, the immediate downstream sensor will provide data for the traffic responsive algorithm. However, if the downstream sensor is used, the meter will be controlled by the speed control algorithm utilizing lookup-table 6. If neither sensor is available, a clock-time algorithm will control the meter. Results located in Appendix E indicate that a metering rate of one car every six seconds will be used.

The results provided in Chapter Five were generated using both the equivalence class and algorithm choice assumptions. These assumptions will be further analyzed in Chapter 6 and their impacts on the results will be discussed.

4.2 Pseudo-Random Search

The costs of deployment schemes range from a minimum of $281,560 (if no sensors or ramp meters are installed) to a maximum of $922,210 (if a sensor and ramp meter is installed at every possible location). This total cost range was split into six budgets of equal width.

Simulated annealing was used to search for the optimal deployment scheme in each of the six budgets. Each search was seeded with a deployment scheme that was expected to perform well.

4.2.1 Performance Evaluation

The process used to evaluate each deployment scheme is non-trivial since it considers three MOP's that are measured during four different traffic scenarios (normal congestion, heavy congestion, and two accident occurrence scenarios). During each simulation run, the average system speed, accident rate, and amount of information collected are measured. These three MOP's are then normalized on a scale from zero to one hundred. Each normalized score is then multiplied by the MOP's weight and the products are summed to generate a single score for that simulation run. The weights assigned to the three MOP's are provided in Table 4.1.

Measure of Performance	Weight
Average System Speed	0.475
Accident Rate	0.475
Amount of Information Collected	0.05

Table 4.1: Weight Assigned to each MOP

This process is repeated for each of the other three scenarios until four scores are generated, one score for each scenario. These four scores are then multiplied by the scenario weights and the products are summed to generate a single score for this four scenario iteration. Table 4.2 contains the weights assigned to each traffic scenario.

Traffic Scenario	Weight
Normal Congestion	0.7
Heavy Congestion	0.2
Accident Occurrence – Location 1	0.05
Accident Occurrence – Location 2	0.05

Table 4.2: Weight Assigned to each Traffic Scenario

The scenario weights were chosen to reflect the relative frequency of each scenario occurring. Thus, it is implied that normal flows will occur about three and one-half times as often as heavy flows are experienced. Likewise, it is implied that an accident will occur roughly one out of each ten morning rush hours. Accidents occurred during twenty-three morning rush hours between July 1998 and June 1999. A full year has approximately two hundred and fifty weekday morning rush hours. Therefore, accidents occurred on roughly 9.2% of all rush hour mornings.

This iteration score captures how well the deployment scheme fulfills each MOP during all four different traffic scenarios. If traffic simulation was a deterministic evaluation method, the iteration score could be used as the final score for the deployment scheme. However, due to the stochastic nature of traffic simulation, deployment schemes will receive different iteration scores due to the randomness in the simulation. Multiple replications must be performed in order to accurately estimate the actual performance score of a deployment scheme.

The number of iterations performed to evaluate a deployment scheme is dependent on the score of the current deployment scheme. If the current deployment scheme (S1) is scoring worse than the current max (S2), the current scheme is only subjected to three iterations. If the current deployment scheme is scoring higher than (S2), five iterations are performed. This rule ensures that high performing deployment schemes are sufficiently tested, but valuable simulation time is not wasted on poor performing deployment schemes. An analysis was performed to determine the sensitivity of the results to the number of iterations performed. The results of this sensitivity analysis are located in Chapter 6.

4.2.2 Accepting/Rejecting Candidates

After five full iterations are complete, if the average iteration score of the candidate scheme is higher than the current maximum, the candidate scheme becomes the current maximum. This new maximum is then be mutated to form a new candidate deployment scheme.

If three or more iterations are completed and the candidate deployment scheme has a lower score than the current maximum, no more iterations are be performed. However, this inferior candidate may still be accepted as the new maximum. Equation 2.4 illustrated how the probability of acceptance can be calculated based on the current temperature and the difference between the scores of the current maximum and the candidate solution. Equation 2.4 assumes a deterministic evaluation process. However, traffic simulation is a stochastic process and the scores assigned to the candidate and maximum are only estimates of their true values. In fact, a candidate that appears inferior after a few iterations may actually have a higher performance score than the current maximum.

Equation 2.4 must be modified to account for the fact that we are not certain of the actual difference between the two performance scores. Although other methods could be used, the law of total probability was chosen to incorporate the uncertainty of the true difference in performance scores.

Based on the mean score and the variance of the scores of each deployment scheme, we can state with a certain degree of confidence that the current maximum score is actually higher than the candidate's score. The confidence level can be determined by testing the hypothesis that the actual value of the two scores (u1 and u2) are equal. Since

the number of replications performed is less than thirty, this is done by calculating the appropriate t score and comparing it to the t-value table. Equations 4.1 and 4.2 show how to calculate the t-score.

Equation 4.1

$$t = \frac{(x_1 - x_2) - (u_1 - u_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where s_p^2 is calculated as:

Equation 4.2

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Recall that a superior candidate will always be accepted and that an inferior candidate will be accepted with a probability equal to $\exp(delta/TEMP_k)$. Equation 4.3 presents the probability of accepting an apparently inferior candidate when a stochastic evaluation process is used.

Equation 4.3

Prob(accept candidate) = Prob(candidate is an improvement)*1 +
            Prob(candidate is not an improvement)*$\exp(delta/Temp_k)$

4.2.2.1 Temperature

The value of the temperature drops throughout the simulation. This allows many backward steps to occur early in the search, but very few backward steps late in the search. The temperature was set so that early in the search an inferior, yet reasonable, candidate will have a forty percent chance of being accepted, and late in the search (around the 100th iteration) a reasonable candidate will have a one percent chance of being accepted. A reasonable candidate is defined as having a score no more than ten points below the current maximum. This correlates to an average system speed approximately 0.75 miles per hour slower and approximately one more accident per year than the best solution.

These specifications require a temperature of 10.9 on the first iteration of the search and a temperature of 2.17 at the one-hundredth iteration. Each iteration, the temperature is decreased by multiplying it by .9838. Figure 4.3 provides the value of the temperature during each iteration.
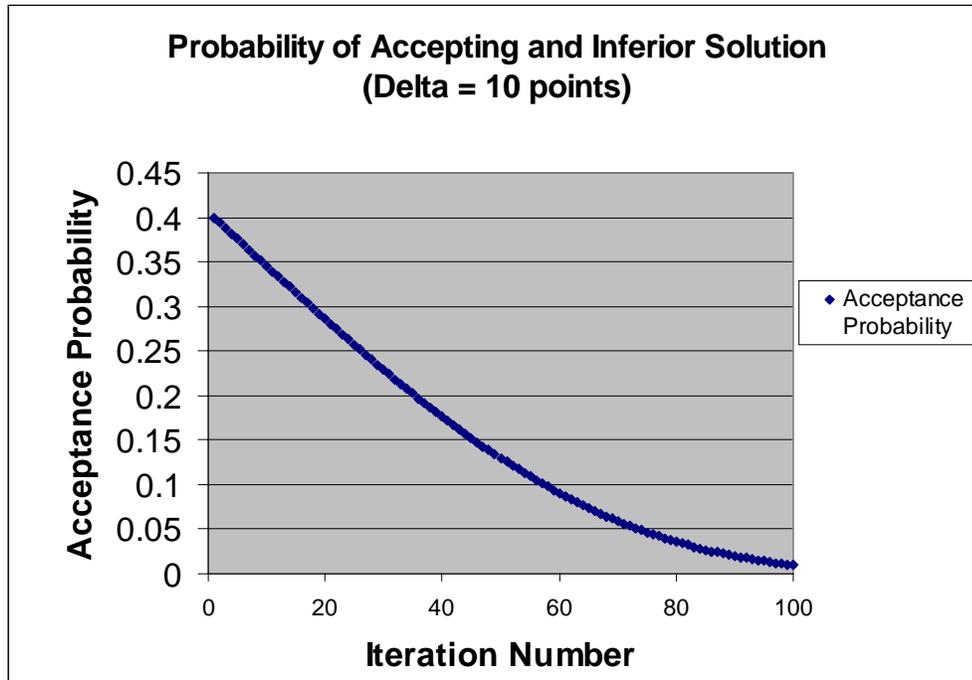
Figure 4.3: Temperature at each Iteration of the Search

4.2.3 Mutation

If a candidate solution is accepted as the new maximum, it will be mutated to form the next candidate. If it is not accepted, the current maximum will again be mutated to form the new candidate. In both cases, the probability of mutating each binary digit in the string is 0.1. Since there are 21 bits in the string, the expected number of mutations is 2.1 bits. It is felt that an expected value of two bits is large enough to generate a variety of candidates, but small enough to ensure that the candidates do not mutate too rapidly.

Chapter 5: Preliminary Results

Eight hundred and twenty-eight deployment schemes were search and evaluated using the methodology described in Chapters 3 and 4. The performance of each deployment scheme and its cost is displayed in Figure 5.1. This figure provides our first look at the performance-cost tradeoff curve for ramp metering systems.
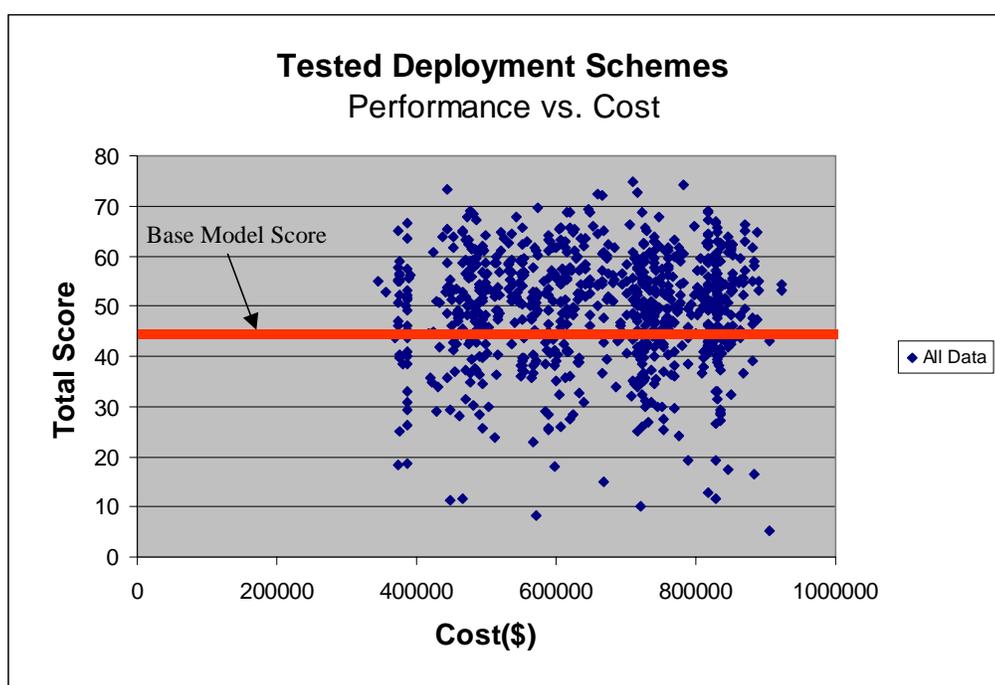
Figure 5.1: Performance and Cost of all Tested Deployment Schemes

Although the data appear quite noisy and the curve is surprisingly flat, several important observations can be drawn from these preliminary results. First, notice that in many cases the implementation of ramp metering produced positive benefits. The base model (no metering) received a performance score of forty-five. This level, marked with a thick red line, represents the traffic flow as it currently exists without any ramp metering equipment. It is obvious that well over half of the deployment schemes tested

earned a performance score higher than forty-five. Surprisingly, if ramp metering is implemented, the amount of spending does not seem to strongly influence the performance score. The level of benefits returned from a $400,000 investment is only slightly lower than the benefits returned from an investment of $800,000. Even more surprisingly, for spending levels higher than $800,000, the performance of the ramp metering system appears to degrade as spending increases.

The relationship between performance and cost can be better understood if the performance score is broken down into its three components: speed, accident rate, and information collected. Figure 5.2, Figure 5.3, and Figure 5.4 display the individual scores of the three components.
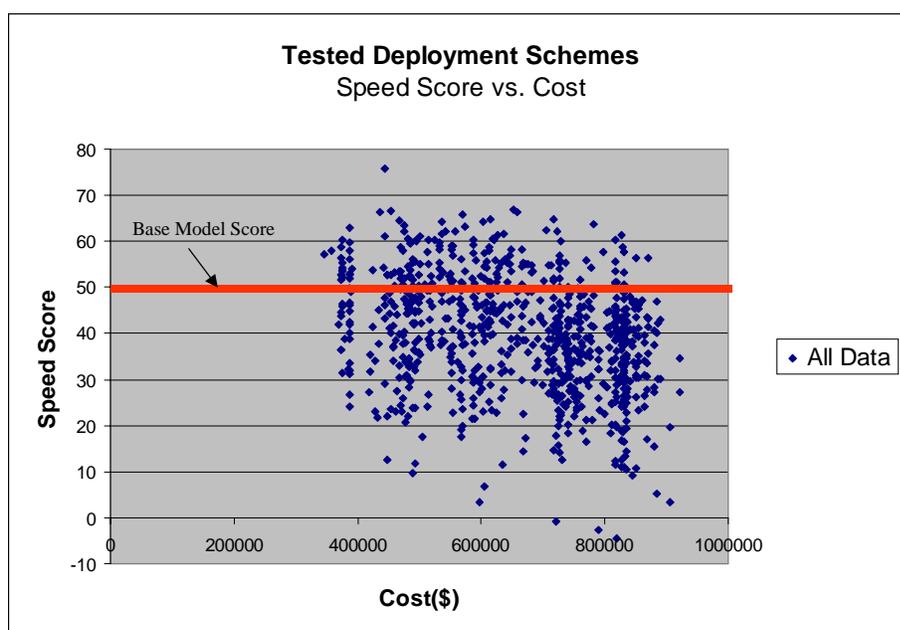
Figure 5.2: Speed Score and Cost of all Tested Deployment Schemes
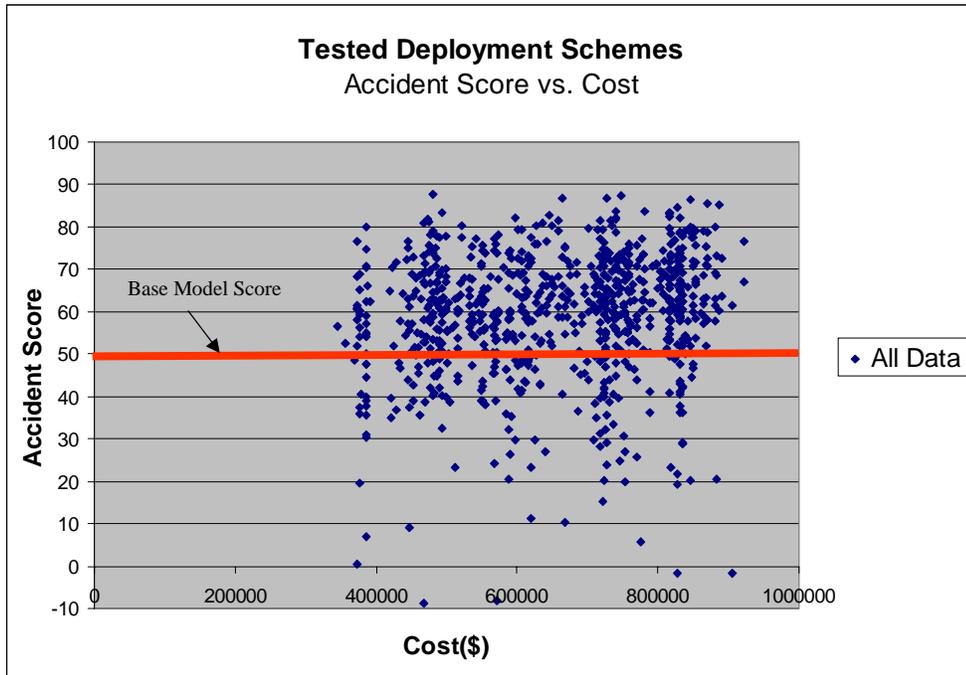
Figure 5.3: Accident Score and Cost of all Tested Deployments
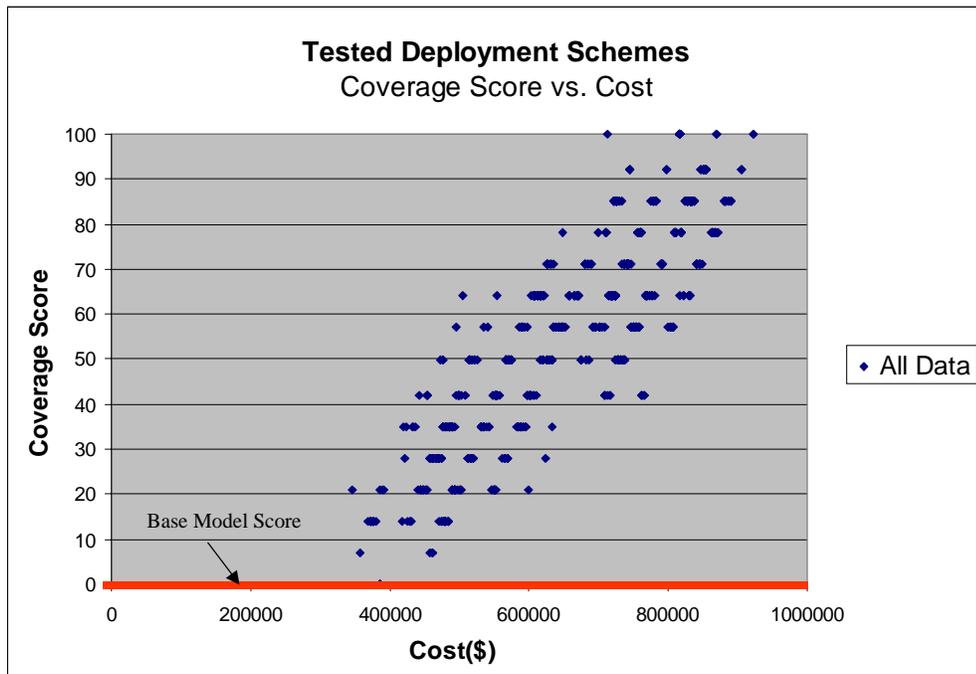
Figure 5.4: Amount of Data Collected and Cost of all Tested Deployment Schemes

From these figures it is apparent that the level of spending affects the individual speed and accident scores even less than the performance score. Although each has a slight upward trend, the increased performance is negligible compared to the additional spending. Without the data collection measure of performance the shape of the performance-cost tradeoff curve in Figure 5.1 would be almost entirely flat.

The results also indicate that the vast majority of deployment schemes are inferior to another deployment scheme. This means that a dominant deployment scheme exists which can provide better performance for a lower cost. In fact, some schemes are so poor that negative benefits are generated. These schemes are not only a poor value given their cost, they actually decrease the performance of the roadway.

Table 5.1 lists ten deployment schemes that cost between $495,000 and $505,000. All these schemes require approximately the same budget, but they produce drastically different results.

ID	Deployment Scheme	Cost	Score
1	000001001000000111000	$477,610	34.86
2	00000100001100010100	$474,310	42.30
3	00000011010101101001	$472,285	50.06
4	00001000100100110000	$474,310	50.17
5	00000101000001010110000	$474,310	55.42
6	00001000001100110100	$476,335	57.47
7	01010101000000110000	$479,635	61.35
8	00001011110010010101000	$475,585	62.35
9	00000101100000110000	$474,310	63.25
10	00000011010001001000	$473,560	67.78

Table 5.1: Deployment Schemes Costing Approximately $475,000

Some of the deployment schemes are nonsensical and would never be selected as a ramp metering design. For instance, scheme three consists of seven sensors and only

one ramp meter. Illogically, none of the sensors is located directly upstream or downstream of the ramp meter. On the other hand, deployment scheme nine appears to be a logical design. Figure 5.5 illustrates the location of the three ramp meters and two sensors deployed in scheme nine. On paper, this deployment scheme appears to be a reasonable design for this budget and could possibly be selected.
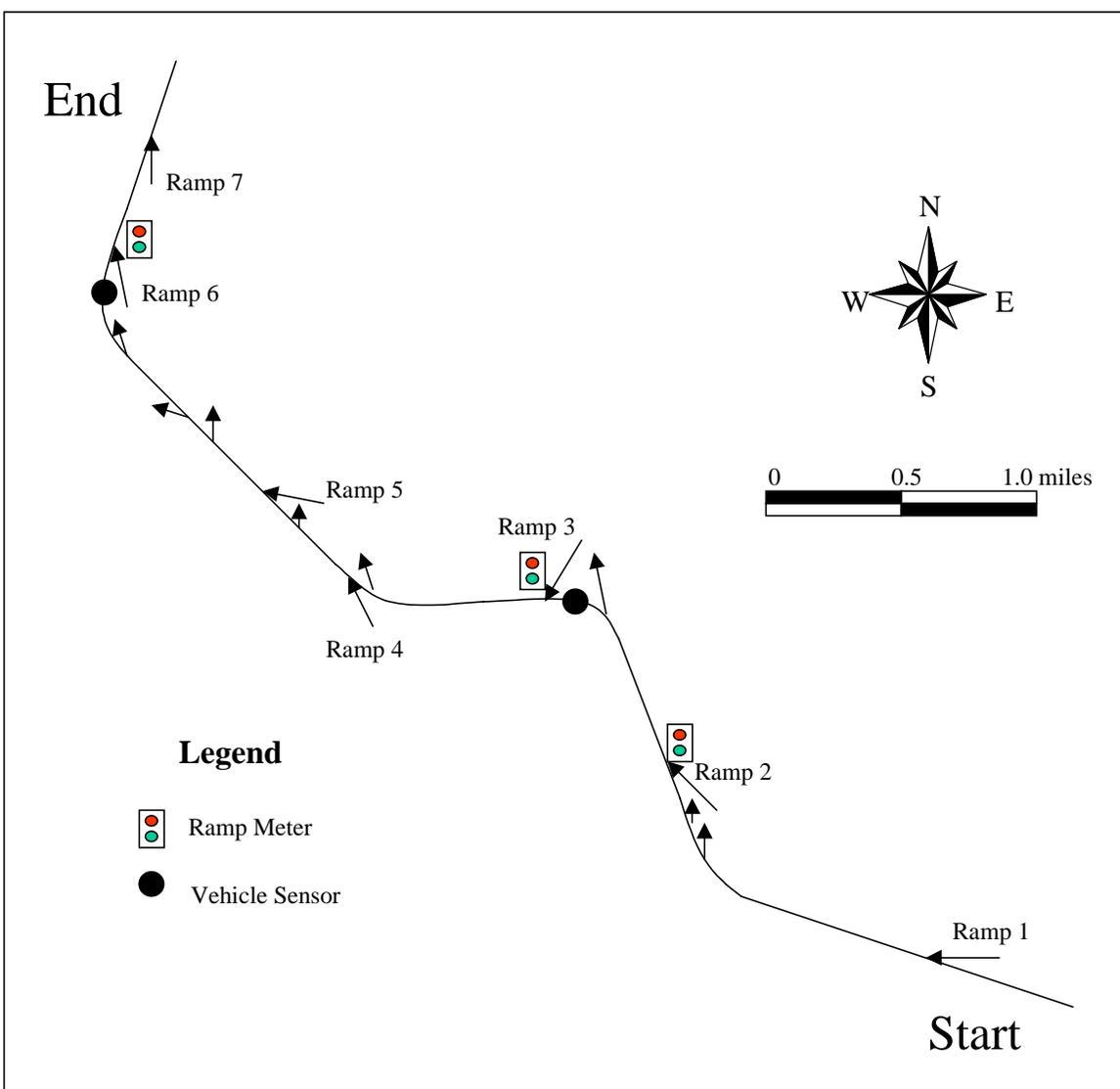
Figure 5.5: Diagram of Deployment Scheme Number Nine

Deployment scheme ten, diagramed in Figure 5.6, outperforms scheme nine. Even though its equipment placement is not as intuitive as locations in scheme nine, the results of five iterations (twenty total simulation runs) indicate that scheme ten dominates scheme nine. Without proper guideline or design tools, planners would likely follow their intuition and unknowingly implement an inferior design.
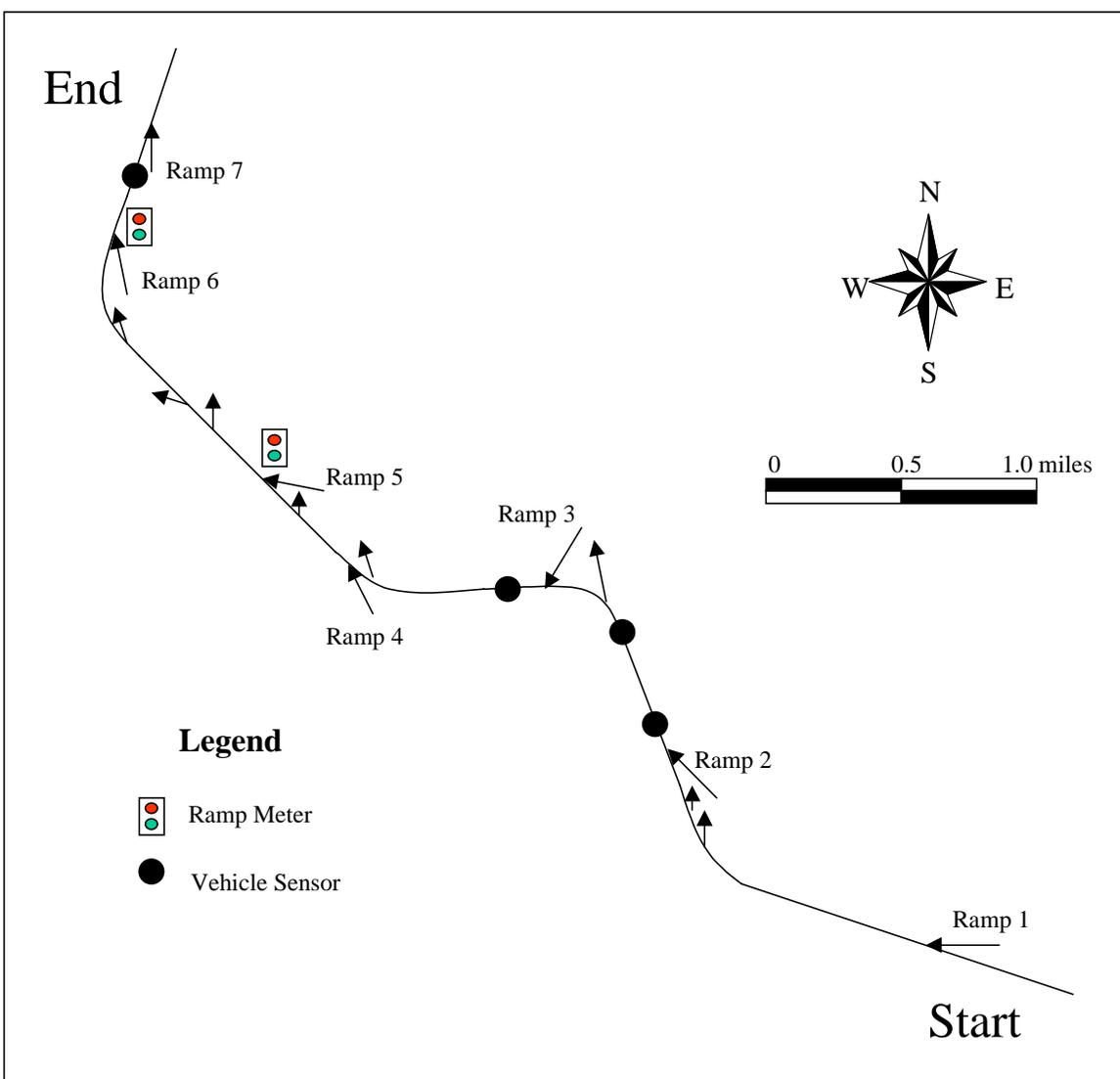
Figure 5.6: Diagram of Deployment Scheme Number Ten

This chapter has presented the preliminary results from the simulation-optimization methodology. Although the results appear to contain a significant amount of noise, some important observations were made. The next chapter improve provides additional insight by further examining both the results and the methodology used to generate them.

Chapter 6: Analysis of Results

This chapter provides a critical analysis of the results presented in Chapter 5. The first part of this chapter will focus on the assumptions made during the implementation of the simulation-optimization methodology. The second section of the chapter performs five sensitivity analyses that investigate the effects of important design decisions and simulation model parameters.

6.1 Testing Assumptions

This section revisits the assumptions made during the implementation process and investigates what effect each assumption has on the results. The five following assumptions will be tested:

1. Five replications are enough to accurately evaluate the performance of each deployment scheme.

2. The entire rush hour period does not need to be simulated to accurately evaluate the performance of each deployment scheme.

3. Traffic data from two separate days is a large enough test set to accurately evaluate the performance of each deployment scheme.

4. Ramp meters will not perform significantly better is they have access to traffic data from sensors outside of a meter's local area.

5. Ramp meters will not perform significantly better if a new optimal control policy is determined for each new deployment scheme.

6.1.1 Number of Simulation Replications

All deployment schemes were evaluated during four different traffic scenarios (normal congestion, heavy congestion, and two accident occurrence scenarios). A single performance score for all four scenarios was calculated using a weighted average of the four performance scores of the individual scenarios. This one score, which encompasses the deployment scheme's performance throughout the entire four-scenario suite, can be thought of as the result of a single replication.

Between three and five replications were performed for each scenario. The final performance score assigned to each deployment scheme is simply the average of the three to five replication scores. This investigation attempts to determine whether three to five replications was sufficiently large enough to accurately evaluate the performance of the deployment schemes.

Figure 6.1 displays the performance scores of the fifty highest scoring deployment schemes. Along with their total score, the figure shows the upper and lower bounds of the ninety-percent confidence interval of the scores.

As can be seen from this figure, the average half-width of each confidence interval is 15.5. This very large half-width is due to both the large standard deviation (average of 12.95) and the small number of replications.

In order to determine the effects of the large variance, the seventeen highest scoring deployment schemes and all deployment schemes in the smallest budget were reevaluated using eight to ten replications.
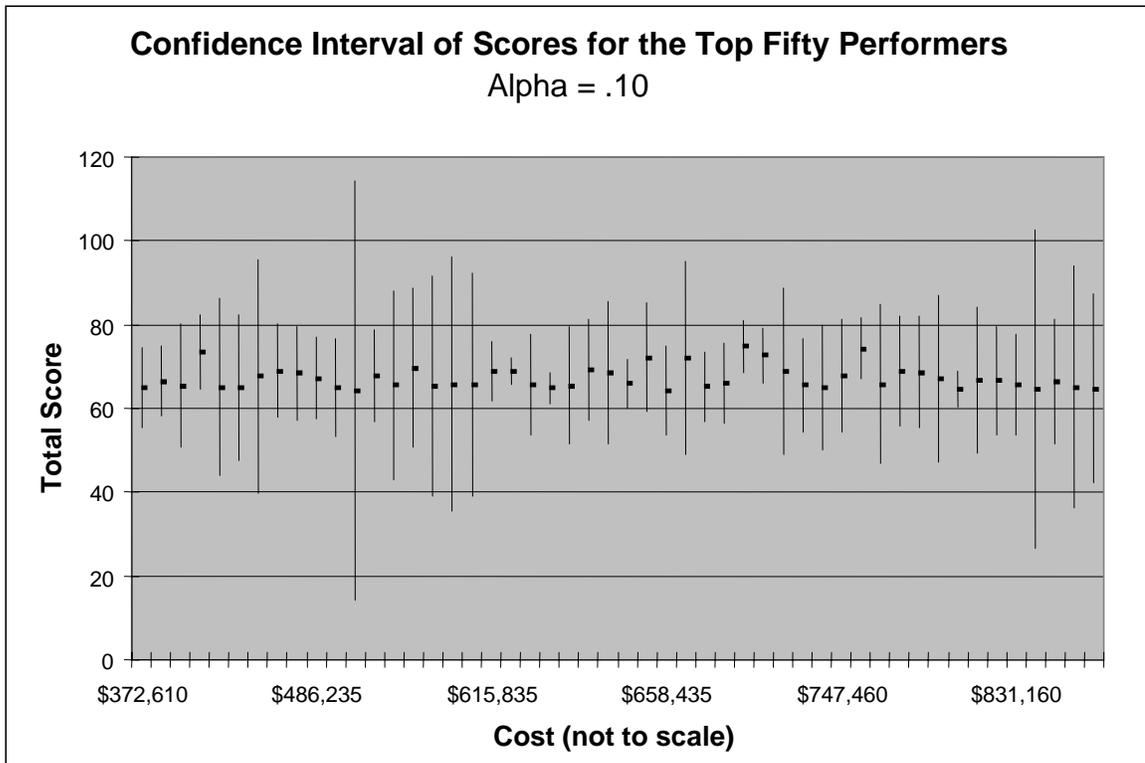
Figure 6.1: Ninety-percent Confidence Interval Around the Total Score of the
Top Fifty Performers

average absolute difference between the preliminary results (with three to five replications) and the additional testing (eight to ten replications) was 10.35 points. Additionally, when ranked by score against the other deployment schemes, the difference between the two rankings had an average of 6.4. It is obvious that the additional testing not only had a large effect on the total score of the deployment schemes, but it also affected how well the schemes performed relative to each other. Even more dramatic results were found when the seventeen top performers were tested further. Every deployment scheme scored worse during the second testing and the average total score decreased nearly fifteen points. The complete results of these tests are located in Appendix F.

The additional testing has shown that performing three to five replications is often not sufficient to accurately estimate the total score of a performance scheme or to definitively conclude that one performance scheme outperforms another. In order to correct for this, the top performing deployment schemes were reevaluated using twenty replications each. As shown in Figure 6.2, the confidence intervals are now drastically narrower and have an average half-width of 6.2.
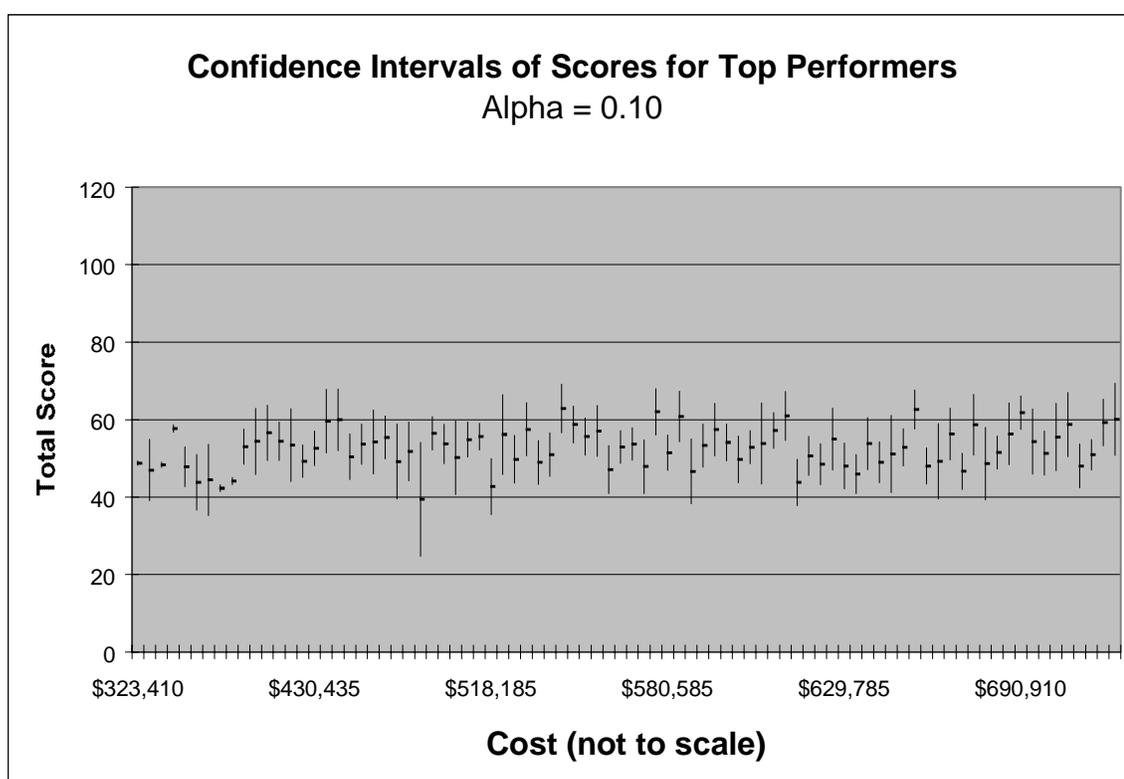
Figure 6.2: Ninety-percent Confidence Intervals of Total Score for Top Performers After Twenty Replications

The shape of the performance-cost tradeoff curve constructed from the new, high replication scores is very similar to the shape of the original curve. The largest difference is that none of the deployment schemes received a total score greater than sixty-three when evaluated using twenty replications. In the preliminary results, however, six

different deployment schemes had a score higher than seventy. It is apparent that the very high scores seen in the preliminary results were elevated due to the randomness in the simulations.

Figure 6.3 displays the total performance scores of the deployment schemes that were reevaluated using twenty replications. As before, the performance-cost tradeoff curve is similar to a step function that increases steeply once funds are invested. However, once the spending level reaches $400,000, the rate of improvement drastically decreases and very little benefits are returned from the additional investment. Figures 6.4 and 6.5 show the annual benefits of each ramp metering deployment schemes in terms of annual hours saved and accidents avoided. As before, the speed benefits are practically flat, but the accident benefits have a slight upward slope. The coverage score, shown in Figure 6.6, again appears to be the major cause of the small positive slope in the performance-cost curve.
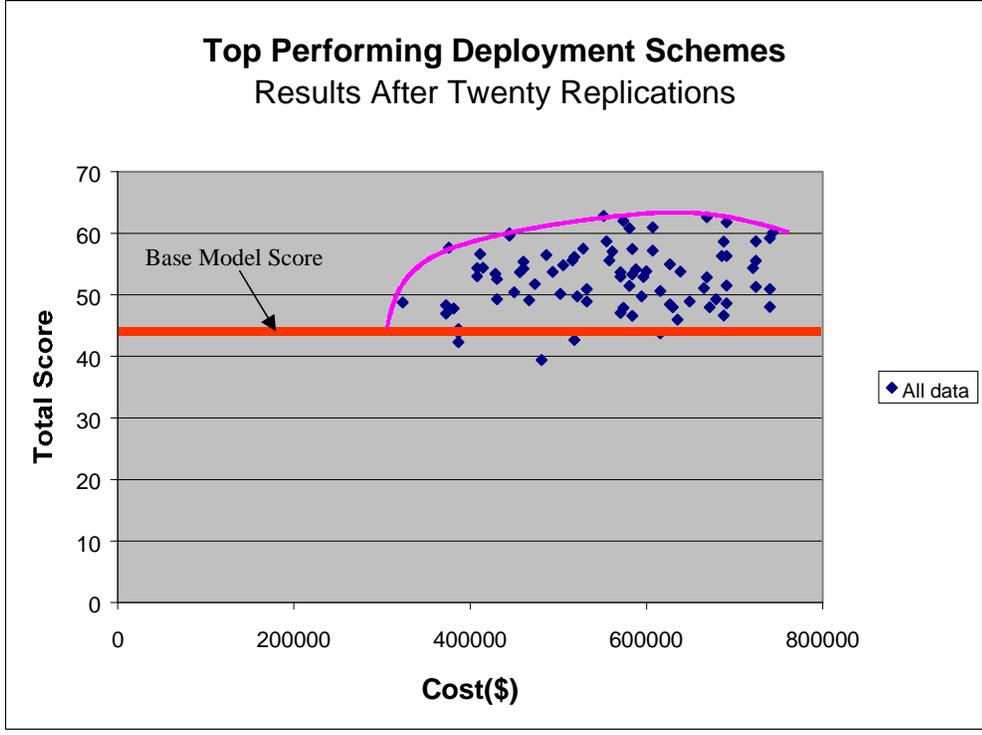
Figure 6.3: Total Score of Top Performing Deployment Schemes
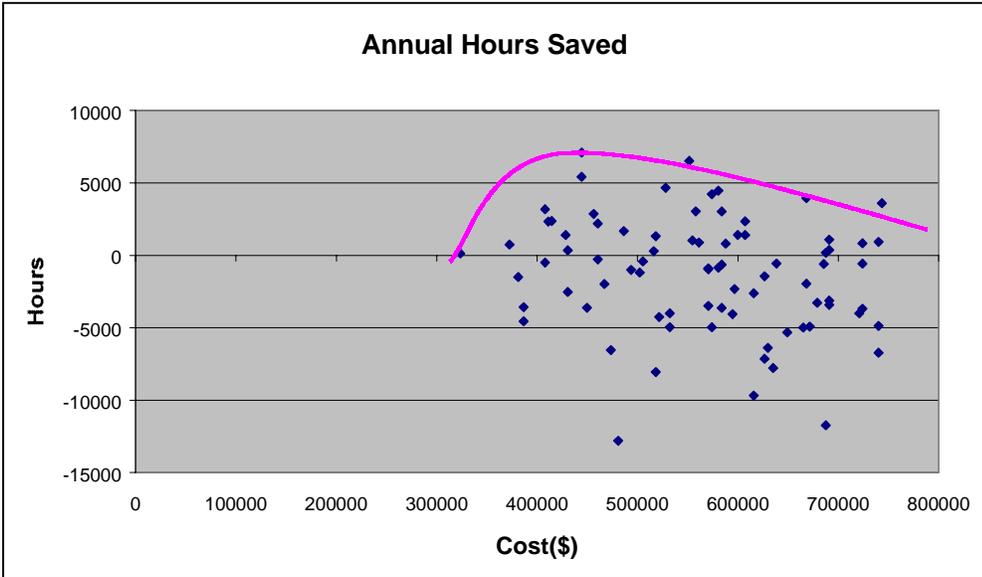
Figure 6.4: Annual Hours of Traveled Saved due to Ramp Metering
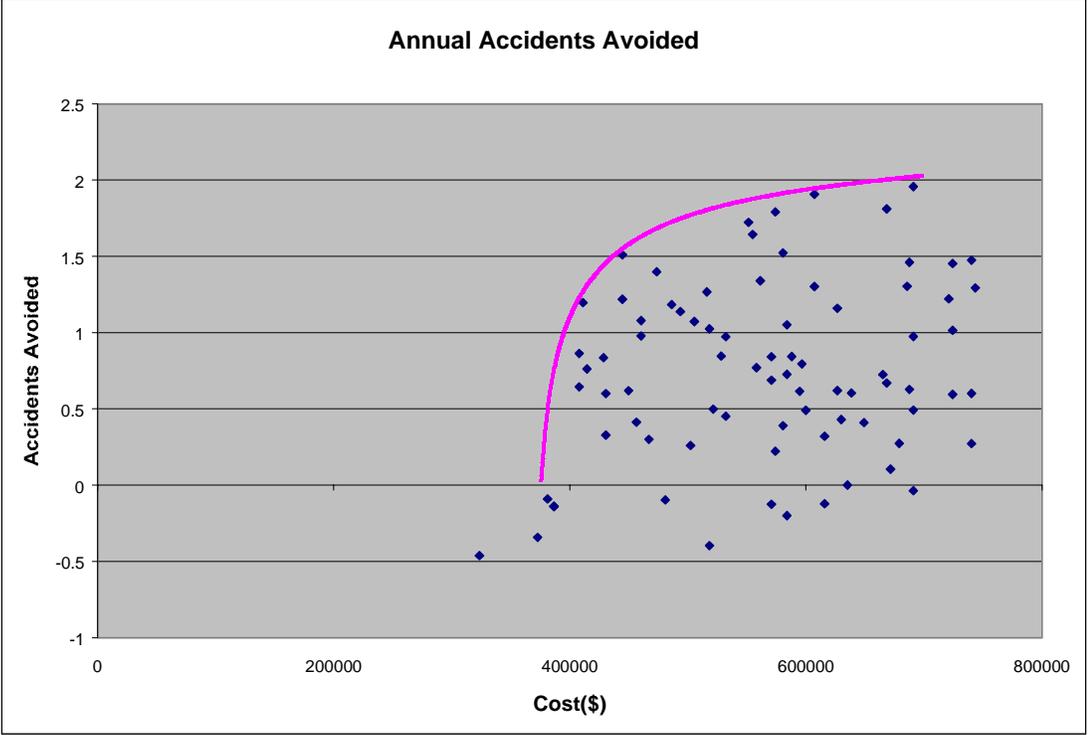
Figure 6.5: Number of Accidents Avoided Annually due to Ramp Metering
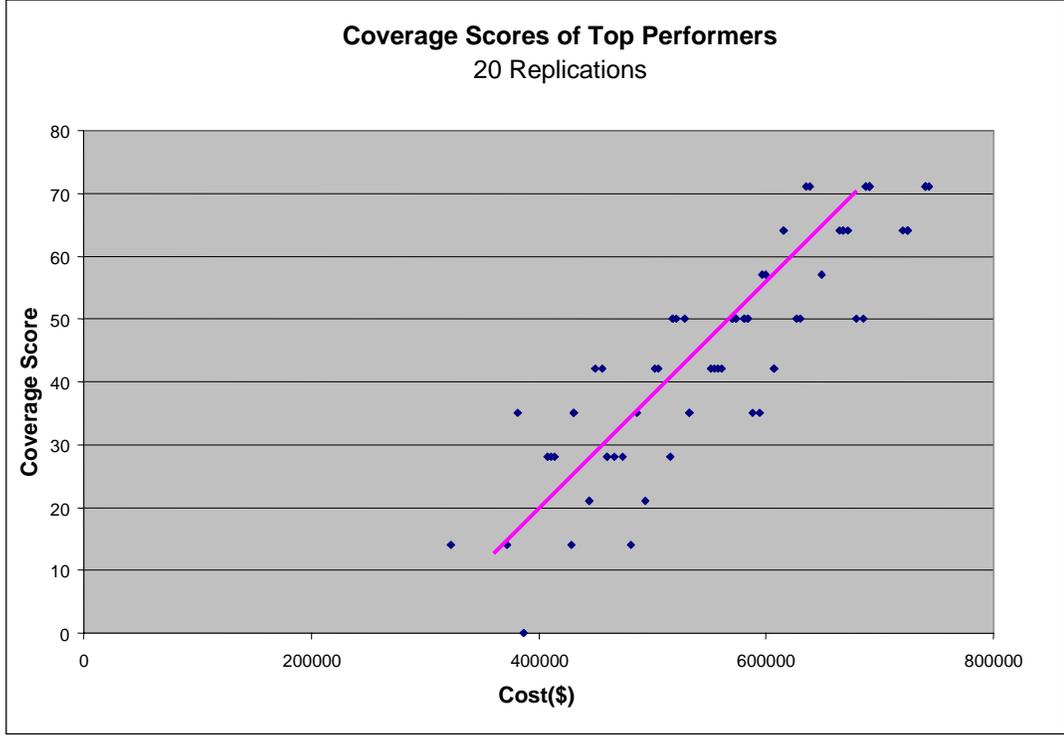
Figure 6.6: Coverage Score of Top Performing Deployment Schemes

6.1.2 Duration of Simulation Runs

All simulation runs performed to evaluate the deployment schemes were twenty minutes in length. It was assumed that twenty minutes was long enough to accurately judge the performance of the deployment scheme during an entire morning rush hour.

This assumption was tested by reevaluating five different deployment schemes with simulation runs that spanned the entire length of the morning rush hour. The five deployment schemes reevaluated, along with their results from the twenty-minute runs, are shown in Table 6.1.

Deployment Scheme	Cost	Ramp States	Total Score	Variance	Rank
111011111101011000000	$685,585	2231200	61.64	226.60	1
011110010011010100000	$554,710	3002100	58.73	405.22	2
101111111011111000000	$668,335	0223200	54.23	406.51	3
011010011010111100000	$570,685	3020200	53.10	392.62	4
000000110100010011000	$473,560	0000130	51.91	305.90	5

Table 6.1: Results of Deployment Schemes from Twenty-Minute Evaluations

The metering algorithms and parameter values controlling these five schemes were not adjusted. They were simply reevaluated using data from the entire morning rush hour of August 11, 1999 instead of only a twenty-minute period. The results of the entire rush hour evaluations are shown in Table 6.2. The scores had to be renormalized due to the longer evaluation period. Although the scores still give an excellent indication of the relative performance of each deployment scheme, the scores themselves cannot be compared to the scores from the twenty-minute evaluations.

Deployment Scheme	Score	Rank	Change in Rank
111011111101011000000	63.33	1	0
011110010011010100000	50.57	2	0
101111111011111000000	32.65	5	-2
011010011010111100000	32.83	4	0
000000110100010011000	47.28	3	2

Table 6.2: Results from Evaluations over Entire Rush Hour

Results from the entire rush hour evaluations are not greatly different than the results from the twenty-minute evaluations. All of the deployment schemes except two are ranked in the same position as they were by the twenty-minute evaluations. Most importantly, the two schemes that performed best during the twenty-minute simulations also performed best during the entire rush hour period. This test provides strong evidence that you can accurately predict the performance of a ramp metering scheme without evaluating the scheme over its entire operational period. This test also implies that using an evaluation period only twenty minutes in length did not significantly alter the performance scores of the deployment schemes.

6.1.3 Traffic Input Data

Each deployment scheme was evaluated by traffic simulations populated with data from either August 11, 1999 or August 23, 1999. Data from actual days were used instead of historical averages since the averaging process smoothes out the fluctuations in traffic demand which partly cause the need for ramp metering. It was assumed that evaluating a deployment using traffic data from one of the two test days was sufficient to accurately assess the performance of the deployment scheme.

This assumption was tested by evaluating ten different deployment schemes using traffic data from both test days. Each scheme was subjected to twenty replications with each set of test data. The test results are presented in Table 6.3.

Configuration	Data Set	Score	Rank	Data Set	Score	Rank	Change in Rank
111011111101011000000	11-Aug	61.64	1	23-Aug	52.25	3	2
011110010011010100000	11-Aug	58.73	2	23-Aug	52.82	2	0
010101110011011000000	11-Aug	57.97	3	23-Aug	54.33	1	-2
011101010110111011100000	11-Aug	55.66	4	23-Aug	48.06	7	3
101111111011111000000	11-Aug	54.23	5	23-Aug	46.57	8	3
111110111111111000000	11-Aug	53.87	6	23-Aug	46.36	9	3
011010011010111100000	11-Aug	53.10	7	23-Aug	51.82	4	-3
000000110100010011000	11-Aug	51.91	8	23-Aug	45.53	10	2
011010011011110100000	11-Aug	50.10	9	23-Aug	50.77	5	-4
000000000001000000010	11-Aug	48.14	10	23-Aug	49.43	6	-4
	average =	54.53		average =	49.79	average =	2.6

Table 6.3: Results Using Both Traffic Data Sets

The average change in ranking indicates that the relative performances of the deployment schemes are dependent on the traffic data set used. However, the same three deployment schemes finished in the top three for both traffic data sets. It appears that robust deployment schemes will perform well for different traffic data sets.

While the change in rank of each deployment scheme is an important observation, the difference between the average scores resulting from each data set is very significant. The average score when the evaluation used traffic data from August 11 is nearly five points higher than the average score calculated using data from August 23. It appears that the traffic demand on August 11 was more conducive to ramp metering. This makes it difficult to accurately compare the performance of deployment schemes that were tested using different data. It is possible that an inferior deployment scheme evaluated with data from August 11 will score slightly higher that a superior deployment scheme evaluated using August 23 data.

Performance evaluations using data from the two different days cannot be compared until they are normalized. One simple normalization technique is to add 4.74 to each score generated using traffic data from August 23. This will make the average scores generated by each data set equal. Table 6.4 displays the results of this test if the scores are normalized in this manner. Notice that when the scores are normalized, the difference between the two performance scores for the same deployment scheme is nearly six percent.

This test provides strong evidence that multiple data sets should be used during the evaluation phase. Even though four different traffic scenarios were used (normal congestion, heavy congestion, and two accident occurrence scenarios), they must be constructed using data from multiple sets. Without multiple data sets it is impossible to

Configuration	Data Set	Score	Data Set	Score	Change in Score
111011111101011000000	11-Aug	61.64	23-Aug	56.99	-4.65
011110010011010100000	11-Aug	58.73	23-Aug	57.56	-1.16
010101110011011000000	11-Aug	57.97	23-Aug	59.07	1.10
011101010101110111100000	11-Aug	55.66	23-Aug	52.80	-2.85
101111111011111000000	11-Aug	54.23	23-Aug	51.31	-2.92
111110111111111000000	11-Aug	53.87	23-Aug	51.10	-2.78
011010011010111100000	11-Aug	53.10	23-Aug	56.56	3.46
000000110100010011000	11-Aug	51.91	23-Aug	50.27	-1.64
011010011011110100000	11-Aug	50.10	23-Aug	55.51	5.41
000000000001000000010	11-Aug	48.14	23-Aug	54.17	6.03
	average =	54.53	average =	54.53	avg = 3.20

Table 6.4: Normalized Results from Both Data Sets

evaluate how a deployment scheme will perform in a variety of different demand levels. Additionally, direct comparison of scores is only possible if each deployment scheme is evaluated using the same data set or groups of data sets. In the future, all deployment schemes should be evaluated via simulations containing traffic data from two or more different test sets.

6.1.4 Equivalence Classes

During the execution of the simulation-optimization methodology, ramp meters only used vehicle detectors immediately upstream or downstream of the meter as sources of data for the traffic responsive metering algorithms. It was assumed that collecting data from sensors outside the meter's local area would not significantly improve the performance of the meter. This assumption decreased the time needed to locate the optimal control policy for each ramp meter and greatly reduced the state space of possible deployment schemes.

This assumption was tested by selecting one previously evaluated deployment scheme and attempting to improve its performance by using data collected from sensors outside the local area of each ramp meter. The deployment scheme selected to test this assumption contains a ramp meter on the second, third, and fourth entrance ramps. Each meter has vehicle detector located nearby and is controlled by a traffic responsive metering algorithm that uses data from the local detector. When evaluated using this local control policy, this deployment scheme received a total score of 57.4.

Minnesota's sophisticated bottleneck algorithm was also implemented to see if the performance of this deployment scheme could be improved by using data from sensors outside of each meter's local area [33]. However, the bottleneck algorithm was never able to generate larger benefits than the local algorithms. A statistical test of the means indicates that there is a less than forty percent probability that the highest scoring bottleneck algorithm outperforms the local algorithm. Table 6.5 contains the results from local algorithm evaluation as well as the bottleneck algorithm evaluations.

Deployment Scheme	Score	Variance	N	Description
00000000110111000000	57.42	300.59	20	Local Algorithm
00000000110111000000	55.91	356.20	20	Bottleneck Algorithm 1
00000000110111000000	54.18	435.90	20	Bottleneck Algorithm 2
00000000110111000000	48.18	360.52	20	Bottleneck Algorithm 3

Table 6.5: Results of Equivalence Class Assumption Tests

For this deployment scheme, the optimal control policy required each ramp meter to use a local metering algorithm. No additional benefits were generated when the meters received data from additional vehicle detectors further upstream or downstream. This provides solid evidence that the equivalence class assumption is valid.

6.1.5 Metering Algorithm Choice

When generating results, the optimal control policy for each deployment scheme was not determined during the search. Instead, the optimal control policy for each individual meter was determined ahead of time. It was then assumed that this policy would remain optimal regardless of whether other meters were installed nearby.

This assumption was tested by selecting two previously evaluated deployment schemes and attempting to locate a new control policy that outperforms the control policy used during the previous evaluation. This assumption was made for both clock-time and traffic responsive meters. Therefore, two deployment schemes were tested, one using clock-time algorithms and one using traffic responsive algorithms.

The first deployment scheme tested contains meters at the second, third, and fifth entrance ramps. However, there are no sensors installed so each meter must use a clock-time metering rate. When the metering rates previously determined to be "optimal" (located in Appendix E) were used, the deployment scheme received a total score of 48.1. Then, several different metering rates that attempted to account for the interactions

between the ramp meters and produce larger benefits were implemented. Two of these modified metering rates scored higher than the metering rate previously determined to be "optimal". Table 6.6 contains the results from each of these clock-time algorithms. Statistical inference indicates that there is a 0.72 probability that the third deployment scheme outperforms the first deployment scheme.

Deployment Scheme	ID	Score	Variance	N	Description
000001001000010000000	1	48.14	398.13	20	Rates Previously Determined
000001001000010000000	2	45.17	798.65	20	New Rates
000001001000010000000	3	51.98	461.61	20	New Rates
000001001000010000000	4	42.78	835.26	20	New Rates
000001001000010000000	5	49.75	528.17	20	New Rates

Table 6.6: Results for Clock-time Algorithm Choice Assumption Test

A similar experiment was also conducted for the traffic responsive metering algorithms. A deployment scheme was evaluated using the previously determined "optimal" algorithms (located in Appendix D). New metering rates were then implemented to see if the performance of the deployment scheme could be increased. Table 6.7 contains the results from the traffic responsive tests. Statistical inference indicates there is a 0.81 probability that the second deployment scheme has a higher actual score than the first deployment scheme.

Deployment Scheme	ID	Score	Variance	N	Description
000001101100110000000	1	50.79	379.27	20	Algorithm Previously Determined
000001101100110000000	2	55.92	296.49	20	New Algorithm
000001101100110000000	3	52.32	247.04	20	New Algorithm
000001101100110000000	4	49.10	317.82	20	New Algorithm
000001101100110000000	5	36.97	632.84	20	New Algorithm

Table 6.7: Results for Traffic Responsive Algorithm Choice Assumption Test

This test was able to locate new local clock-time and traffic responsive control policies that outperformed the policies that were previously determined to be optimal. Although, the increases in performance were not statistically significant at a 0.1 level, there is still strong evidence that the algorithms which best control an isolated ramp meter are not optimal when the ramp meter is part of a larger system. It appears that performance can be improved if time is spent determining the optimal control policy for each deployment scheme. Future evaluations should not assume that algorithms, which optimize the performance of an isolated meter, will be optimal when other ramp meters are installed.

6.2 Sensitivity Analysis

During this project several decisions were made that may have affected the results. Two important examples are the weights selected to calculate the total score of each scenario and the seed used to start the simulated annealing search. Additionally, simulation parameters were set to values observed in the Hampton Roads test area. The results cannot be considered transferable to other test areas unless the affects of using these traffic parameters are known.

Sensitivity analyses have been performed on five critical parameters in order to assess their impact on the final results and recommendations. The objective of these analyses is to probe each parameter and determine how sensitive the results are to the value of the parameter. These analyses are not intended to be complete experiments which identify the relationship between the results and each parameter.

6.2.1 Weighting Scheme

The three measures of performance used to calculate the performance score of each deployment scheme were the average system speed, the accident rate, and the amount of information collected. The total score assigned to each scheme was simply a weighted average of these three metrics. The weights used are shown in Table 6.8.

Measure of Performance	Weight
Average Speed	0.475
Accident Rate	0.475
Amount of Data Collected	0.05

Table 6.8: Weight of each MOP

The weights are correlated to importance the researcher assigns to each MOP. Although this weighting scheme was used, there are certainly other reasonable weighting schemes that could be employed. In an area plagued by frequent accidents, the accident rate may be the most important MOP. Conversely, the average speed would be most important in an area with terrible delay but few accidents. Additionally, locations that are planning to provide extensive forecasting or traffic information services may feel that the information collected should be weighted much higher than 0.05.

This analysis will investigate how varying the weighting scheme affects the results. A robust solution will perform well regardless of the weighting scheme used. However, the performance score of a deployment scheme that only fulfills one or two of the MOP's well will be highly dependent on the weighting scheme used.

The performance scores of the top eighty performing deployment schemes were computed using three new deployment schemes. Table 6.9 lists the individual weights comprising each of the three new weighting schemes.

Weighting Scheme	Average Speed	Accident Rate	Data Collected
1	0.60	0.35	0.05
2	0.35	0.60	0.05
3	0.40	0.40	0.20

Table 6.9: Three Alternative Weighting Schemes

For each new weighting scheme the average change in rank was calculated as well as the number of deployment schemes whose rank was identical to its ranking in the original data set. Additionally, the number of top ten deployment schemes remaining in the top ten was counted. Table 6.10 displays the results of these analyses.

Weighting Scheme	Avg. Change in Rank	No. With Unchanged Ranking	No. Remaining in the Top Ten
1	2.0	26 out of 80	10 out of 10
2	2.175	14 out of 80	9 out of 10
3	10.35	4 out of 80	7 out of 10

Table 6.10: Results with Alternative Weighting Schemes

Weighting schemes one and two, which varied the weights assigned to the average speed and accident rank, had little effect on the results. In most cases, the deployment schemes scored either well or poorly on both MOP's. This indicates that deployment schemes which increase the average speed also tend to reduce the accident

rate. This is logical since the accident rate increases when vehicles tend to slow well below the speed limit.

Varying the weight assigned to the amount of data collection greatly changed the results. Apparently, deployment schemes which collect a large amount of data often do not increase the average speed or decrease the accident rate. This is in agreement with the decrease in performance at very high budgets that can be seen on the right side of the performance-cost tradeoff curve.

This sensitivity analysis indicates that unless the amount of data collected MOP is heavily weighted, the results are not greatly dependent on the weighting scheme used to calculate the final result. Therefore, if another reasonable weighting scheme had been chosen the relative performances of the deployment schemes would be very similar.

6.2.2 Car-Following Aggressiveness Parameter

The car-following aggressiveness is one of the parameters used in CORSIM's car-following logic. The value of the parameter influences how closely drivers will follow each other and affects the effective capacity of the roadway. The car-following aggressiveness that minimized simulation error during the model calibration was used during all simulation runs.

The purpose of this analysis is to determine if and how the value of the car-following aggressiveness parameter affects the performance of deployment schemes. The conclusions drawn from this research will be much more transferable to other test areas if the parameter has very little affect on the performance of the schemes. Conversely, if the aggressiveness parameter seems to largely affect the performances, the results of this research may only apply to other roadways with similar driver behavior.

The effect of the car-following aggressiveness parameter was probed by reevaluating five deployment schemes using a different, less aggressive parameter. The costs and performances of the five deployment schemes chosen are listed in Table 6.11.

Deployment Scheme	Cost	Ramp States	Total Score	Variance	Rank
111011111101011000000	$685,585	2231200	61.64	226.60	1
011110010011010100000	$554,710	3002100	58.73	405.22	2
101111111011111000000	$668,335	0223200	54.23	406.51	3
011010011010111100000	$570,685	3020200	53.10	392.62	4
000000110100010011000	$473,560	0000130	51.91	305.90	5

Table 6.11: Car-Following Aggressiveness Set at Original Values (1.0 – 0.1, 5)

The results of the evaluations using a different driver aggressiveness parameter are shown in Table 6.12. The scores had to be renormalized due to the capacity decrease caused by less aggressive drivers. Although the scores still give an excellent indication of the relative performance of each deployment scheme, the scores themselves cannot be compared to the scores obtained with the original car-following aggressiveness parameter.

Deployment Scheme	Total Score	Rank	Change in Rank
111011111101011000000	60.87	2	1
011110010011010100000	55.73	3	1
101111111011111000000	44.71	5	2
011010011010111100000	54.09	4	0
000000110100010011000	65.20	1	-4

Table 6.12: Car-Following Aggressiveness Set at New Values (1.2 – 0.3, 10)

Surprisingly, with a different car-following parameter value, the deployment scheme that previously scored worst has now scored the highest. Additionally, the scheme that was ranked third best performed much worse than all the other schemes.

The results of this analysis indicate that the car-following aggressiveness parameter can have a significant result on the performance of deployment schemes. Therefore, a researcher should always carefully calibrate the simulation model before evaluating any potential improvements. Additionally, this test indicates that the performance of deployment schemes may be dependent on the aggressiveness of the drivers in the area. Thus, a deployment scheme that has been effective in one area may not perform as well if implemented elsewhere.

6.2.3 Amount of Traffic Congestion

The level of traffic congestion present in the simulations was based on historical traffic data gathered from the test area. Like the car-following aggressiveness parameter, the ability to transfer the results to other test areas greatly depends on how the level of congestion affects the results. The objective of this test is to determine if the relative performance of different deployment schemes is heavily dependent on the amount of congestion present.

The relationship between deployment scheme performance and traffic congestion was investigated by reevaluating five deployment schemes. The traffic demand in the simulations used for this test were ten percent higher than the demand in the simulations used to generate the results. The five deployment schemes shown previously in Table 6.11 were used as the test group.

Table 6.13 contains the results of the tests with a higher congestion level. Interestingly, the deployment scheme that had performed best in the group at the original demand level now performed worst at the increased levels. This is quite surprising since the other four deployment schemes all scored the same relative to each other.

Configuration	Total Score	Rank	Change in Rank
11101111101011000000	47.98	5	4
01111001001101010000	65.35	1	-1
10111111011111000000	61.88	2	-1
01101001101011110000	57.53	3	-1
00000011010001001100	53.13	4	-1

Table 6.13: Results with Traffic Demand Increased Ten Percent

The mixed results generated by this test make it difficult to draw firm conclusions. It appears that generally a deployment scheme's performance is not greatly affected by minor differences in the level of congestion. However, it that seems varying the level of congestion can test the robustness of a deployment scheme. A ten percent increase in traffic level greatly lowering the performance of a deployment scheme is an indication that they scheme is not a robust solution and should not be implemented. Although the level of congestion does not have a large affect on the performance of a solution, potential improvements should be tested at both the current traffic demand and predicted levels of future demand. This will force designers to look past short-term results and ensure that a robust solution is being implemented.

6.2.4 Variance due to Search

Every simulated annealing search will generate different results even if the same seed and parameter values are used. The randomness found in the results of the search is directly caused by the randomness within the search technique itself.

The objective of this analysis is to investigate the magnitude of the variance from search to search. A large variance indicates that multiple searches should be performed

in order to generate thorough results. A small variance will provide evidence that one search is enough to generate accurate results.

The variance due to the search was explored by performing another search of the third budget. This search had identical parameters, but used a different seed. Figure 6.7 contains the results from the original search of the third budget. Figure 6.8 contains the results of the second search of the third budget.

The results from the two different searches are quite similar. In both searches the majority of the deployment schemes scored between forty and seventy. Additionally, during each search the top performing schemes scored in the low seventies. If viewed in the context of a larger performance-cost tradeoff curve, the results of the two searches would appear almost identical.
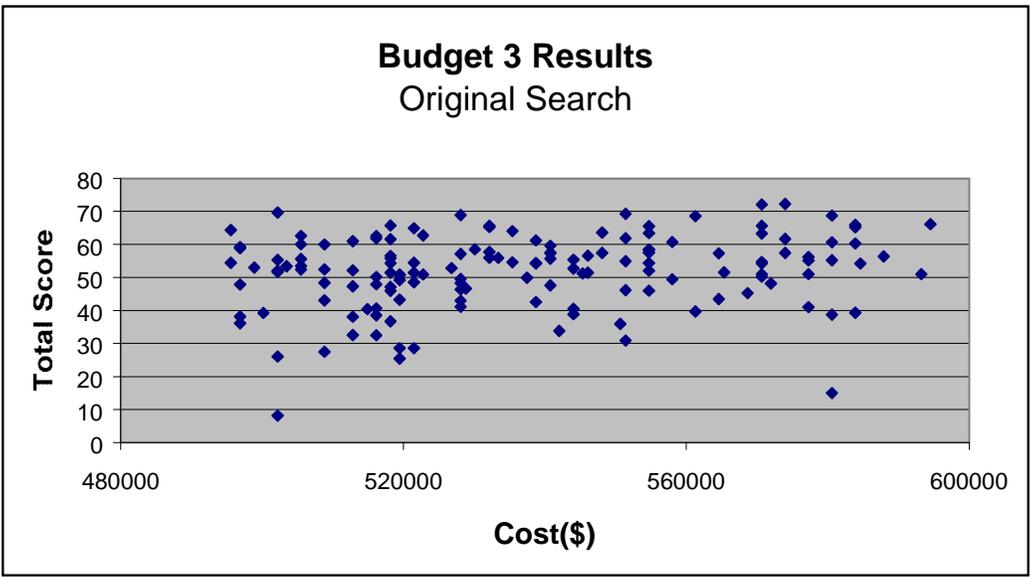
Figure 6.7: Results From Original Search of Budget 3
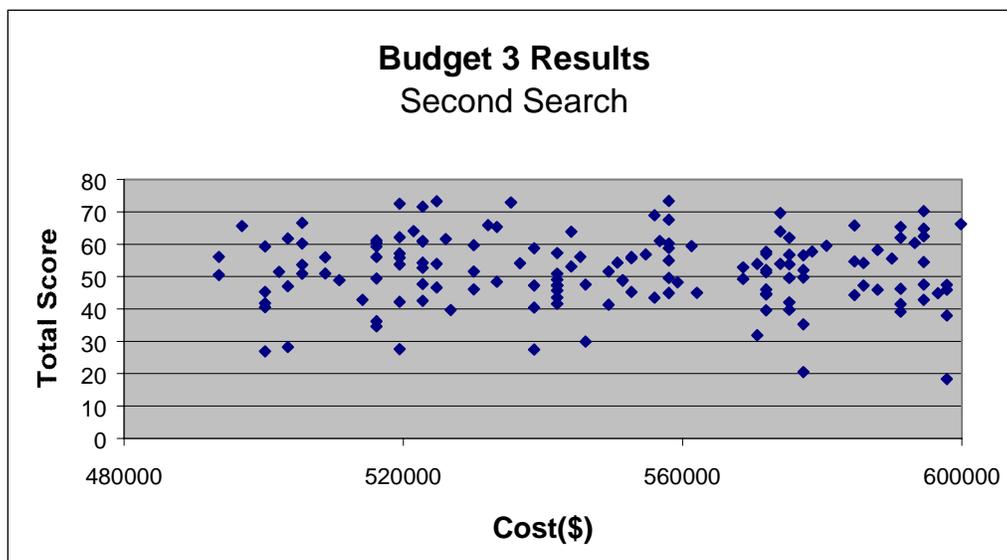
Figure 6.8: Results From Second Search of Budget 3

Not only is the shape of the curve similar, but the top performing schemes in each search are quite alike. Almost all of the top performers have meters on three ramps: ramp two, ramp five, and one additional ramp. Given the similarity of the resulting curves and the top performing deployment schemes, it appears that the variance between different simulated annealing searches does not have a large effect on the results.

6.2.5 Traffic Arrival Distributions

Throughout this research, the Erlang distribution with shape parameter equal to two was used to model vehicle inter-arrival times. The Erlang distribution was chosen because it provides the best fit of the three arrival distributions allowed by CORSIM.

Arena's Input Analyzer, supplied with the Arena Simulation Package, was used to analyze the inter-arrival data. The Input Analyzer indicated that both the Lognormal distribution and the Gamma distribution fit the collected data better than the Erlang distribution. Table 6.14 displays mean square error of the distributions the Input

Analyzer attempted to fit to the arrival data. Mean square error is one metric used to analyze goodness-of-fit.

Theoretical Distribution	Mean Square Error
Lognormal	0.00661
Gamma	0.0175
Erlang	0.0182
Beta	0.0379
Normal	0.0508
Exponential	0.0829
Triangular	0.101
Uniform	0.1333
Weibull	0.357

Table 6.14: Square Error of Distributions Fit to Arrival Data

Although the mean square error appears low, all of these distributions scored horribly when tested by the Chi-squared and Kolmogorov-Smirnov goodness-of-fit tests. In fact, none of these distributions had a p-value larger than 0.01.

Since none of the distributions appear to be a good fit, it is important to see how different arrival distributions affects the results. If using different arrival distributions does not significantly affect the results, then it is largely inconsequential that a loose fitting distribution was used. However, if the arrival distribution significantly affects the results then the accuracy of the deployment scheme evaluations must be questioned.

Both the base case (no meters) and one deployment scheme were evaluated using two different arrival distributions: the Erlang with shape parameter 4 and the Normal distribution. The results were then compared to the original evaluation using a two-tailed t-test to see if there was a significant difference. Tables 6.15 and 6.16 contain the results of these tests.

Deployment Scheme	Arrival Distribution	Score	Variance	N	t	p-value
00000000000000000000	Erlang(2)	50.00	184.16	20		
00000000000000000000	Erlang(4)	53.45	460.29	20	0.608	0.547
00000000000000000000	Normal	54.87	662.46	20	0.748	0.459

Table 6.15: Base Model Results with Different Arrival Distributions

Deployment Scheme	Arrival Distribution	Score	Variance	N	t	p-value
01000001110000100000	Erlang(2)	49.13	473.72	20		
01000001110000100000	Erlang(4)	50.41	702.39	20	0.167479	0.867837
01000001110000100000	Normal	48.74	410.83	20	0.05777	0.954219

Table 6.16: Deployment Scheme Results with Different Arrival Distributions

Statistical inference tests indicate that no difference exists between the results within any reasonable significance level. It appears that the arrival distribution has almost no impact on the performance on either the base case or the tested deployment scheme. This could be due to a general insensitivity to the arrival distribution or caused by the long straight away at the very beginning of the model. The first half-mile of the simulation model is a straightaway that contains no entrances or exits. In this straightaway, vehicles will shift around relative to each other based on the behavior characteristics of the driver. This movement increases the randomness in the traffic flow and likely greatly reduces the importance of the arrival distribution being used.

6.3 Summary

Table 6.17 summarizes all the tests and sensitivity analyses discussed in this chapter and the results of each test.

Test Name	Section	Purpose	Result
Number of Simulation Replications	6.1.1	Determine if five replications are enough	Not enough. At least 20 replications should be performed
Duration of Simulation Runs	6.1.2	Determine if a twenty-minute simulation run is long enough	Yes, it is long enough
Traffic Input Data	6.1.3	Determine if two data sets are enough	At least two should be used. Every scheme should use every data set
Equivalence Classes	6.1.4	Determine if the Equivalence Classes Assumption is valid	Yes, it appears to be valid.
Metering Algorithm Choice	6.1.5	Determine if the Metering Algorithm Choice Assumption is valid	No, it appears that performance is increased if the control policy is continually optimized
Weighting Scheme	6.2.1	Determine sensitivity of the results to the weighting scheme	Weighting scheme does not greatly affect results
Car-Following Aggressiveness Parameter	6.2.2	Determine the sensitivity of the results to the car-following aggressiveness parameter	The parameter appears to affect the results
Amount of Traffic Congestion	6.2.3	Determine the sensitivity of the results to the level of traffic congestion	The level of traffic appears to affect the results
Variance due to Search	6.2.4	Determine the sensitivity of the results to the pseudo-random search process	The results do not appear to be sensitive to the search process
Traffic Arrival Distributions	6.2.5	Determine the sensitivity of the results to the traffic arrival distribution	The results do not appear to be sensitive to the arrival distribution

Table 6.17: Summary of Tests Performed

6.4 Favorable Deployment Schemes

Several mathematical models were built using linear regression in an attempt to better understand what factors are driving the ramp metering benefits. The models used

data from previous simulation runs and attempted to predict the performance of a deployment scheme based on the location of equipment in each ramp metering system.

It proved impossible to build an accurate model using the preliminary results which were evaluated using five replications or less. The best model constructed with the preliminary data only had an adjusted R-square value of 0.117. These data are obviously too noisy to be mathematically model at an acceptable level of accuracy. This reemphasizes the fact that five replications are simply too few to accurately estimate the true value of the performance score.

Linear regression models built using the results with twenty replications were able to achieve a moderate level of accuracy. The best model constructed using this data had an adjusted R-square value of 0.503. Although it is not completely accurate, this model is able to explain more than half of the deviation seen in the performance scores. Appendix G contains a complete description of the linear regression process and all models constructed.

The much larger adjusted R-square value indicates that the second data set (twenty replications) contains much less noise than the preliminary results. It seems reasonable that the adjusted R-square value will continue to increase as the number of replications performed increases. However, at some point the adjusted R-square will reach an asymptote and no longer increase as the number of replications increased. If this were not the case, an extremely accurate mathematical model could be created and there would no longer be a need for the traffic simulations. However, traffic flow is a random process and it is unlikely that a mathematical model could outperform a simulation package.

The linear regression models provide insight concerning why some configurations perform better than others do. These models indicate that the greatest positive impact is generated from ramp meters placed on the first, third, and fifth entrance ramps. Conversely, placing meters on the second and sixth entrance ramps causes a decrease in performance. Placing meters on the fourth and seventh entrance ramps does not cause a significant change in performance.

There are also significant interactions between the ramp meters. The benefits generated from placing meters on ramps one and three is much smaller than the sum of the benefits generated by each meter individually. These two meters interact in a way that degrades performance when they are both in operation. There is also a negative interaction between the third and fifth meters. These interactions validate the earlier conclusion that performance is not significantly increased as more ramp meters are installed. In both these cases the negative interaction negates any additional benefits that would be derived from installing the second meter.

The models also indicate that the ramp metering algorithm used by the meters can influence the benefits generated. The ramp meter on the second entrance ramp performs much better when control by a traffic responsive metering algorithm. The fifth ramp meter, however, has a superior performance when it is controlled by a constant clock-time metering algorithm.

As mentioned earlier, the greatest benefits are generated when the first, third, and fifth entrance ramps are metered. These ramps were analyzed to determine what common characteristics are shared by the three ramps. The common characteristics of these ramps

may indicate which ramp characteristics are favorable and cause some ramps to return larger benefits when metered.

All three ramps experience a moderate level of demand and have a relatively large demand variance. The moderate demand characteristic is logical since metering is not practical on entrance ramps with very low volumes or very high volumes. On a low volume ramp metering is unnecessary since the arrival of cars is infrequent and does not greatly affect the mainline. Metering an entrance ramp with a very high demand can also cause problems if an excessive queue builds. Therefore, it is logical that ramps with a moderate level of demand are most conducive to ramp metering.

The high demand variance characteristic is also logical. Ramps with a high demand variance will experience short periods of very high volume that can stress the mainline. Ramp metering will smooth out these flows and greatly reduce the shock of cars entering the freeway.

This analysis indicates that ramps with a moderate demand and high demand variance should be considered as excellent candidates for metering. Unfortunately, the analysis was unable to identify additional common characteristics of the three ramps. Future research in this area would be a worthwhile venture with the potential of generating practical results.

Chapter 7: Conclusions

This research has identified a deficiency in the Intelligent Transportation Systems knowledge base and has developed and implemented a methodology to explore this unknown area. This final chapter contains concluding remarks about both the previously unknown benefit-cost relationship within ramp metering systems and the methodology developed to explore this relationship. The chapter closes with a discussion of potential areas of future research.

7.1 Performance-Cost Relationship

The performance-cost tradeoff curve revealed by this research contains the shape of a step function. Significant benefits can be generated from even the least expensive deployment scheme. As the budget increases, however, diminishing returns are experienced and additional benefits are generated at a very low rate. Additionally, above $700,000 the return of benefits actually decreases as spending increases. Figure 7.1 displays the shape of the performance-cost tradeoff curve.
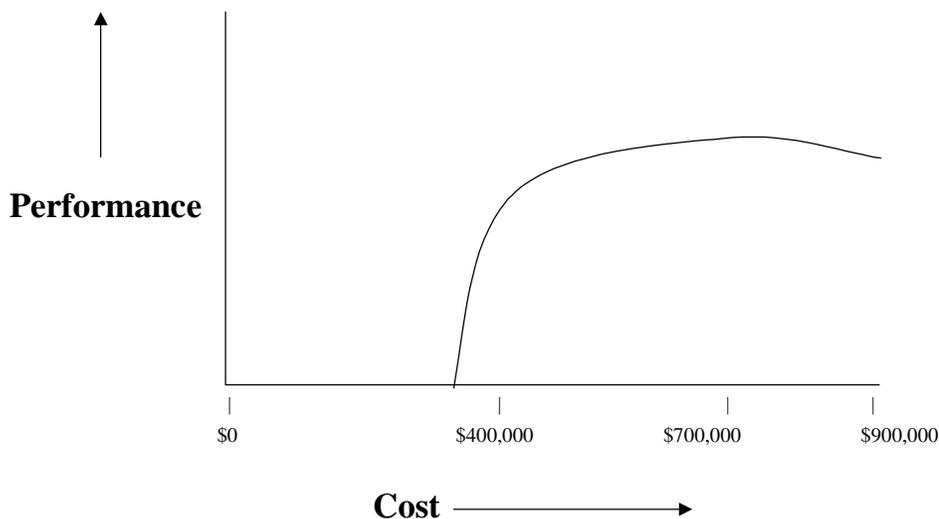
Figure 7.1: Performance-Cost Tradeoff Curve

It is important to remember that the methodology used to create this curve did not allow for ramp metering systems controlled by a single system-wide algorithm. If system-wide algorithms had been included, the performance-cost curve would likely experience another large increase in benefit at the budget level where system-wide control becomes economically feasible.

The performance-cost curve contains important information regarding ramp metering implementation. The data indicate that local area ramp metering only makes sense as a low cost solution. If a transportation agency has a small budget and wants to get a high return on its investment, local area meters are a good option. Implementing only a couple local meters at critical entrance ramps should improve traffic flow and increase safety. If the budget is moderate or large, however, installing many local ramp meters is not a wise investment. The results indicate that the benefits generated by installing a large number of local ramp meters is only slightly larger than the benefits derived from installing a few strategically placed meters. Agencies with larger budgets should implement a centrally controlled ramp metering system in which all meters are controlled by a system-wide algorithm. A smaller number of central control meters will likely provide greater benefits than a larger number of locally controlled meters.

7.2 Simulation-Optimization Methodology

The simulation-optimization methodology developed to explore the performance-cost relationship is another significant contribution of this research. Displayed in Figure 7.2, the methodology is useful for ramp metering systems or can be slightly adapted for use with other forms of ITS equipment.
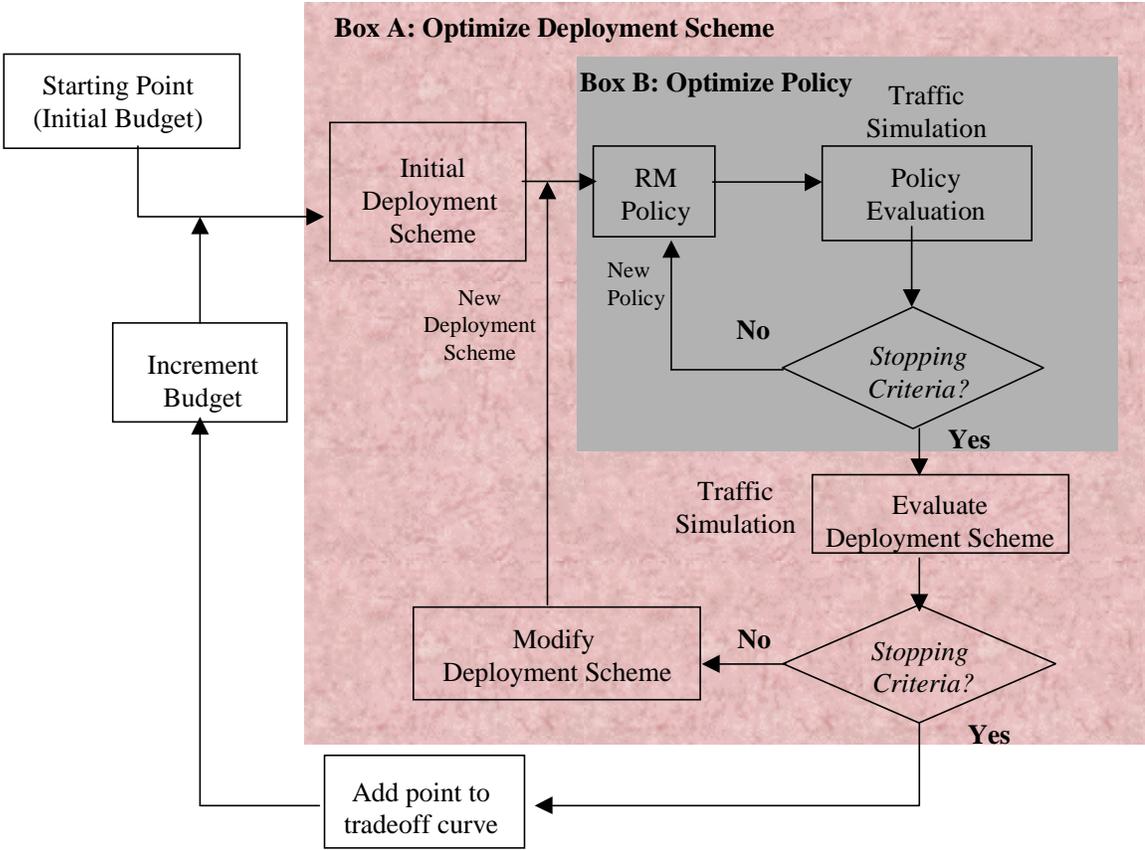
Figure 7.2: Simulation-Optimization Methodology

During the implementation and execution of this methodology, both strengths and weaknesses inherent to the methodology became apparent. The two principal strengths of the methodology are its abilities to objectively evaluate deployment schemes and to generate less obvious deployment schemes that may not have been otherwise considered. Locating an optimal control policy and then simulating the deployment scheme over a range of traffic flows is most accurate way to estimate the benefits of a potential improvement. This method is far superior to simply performing simulations with only one data set or employing available algebra based techniques. Additionally, the pseudo-random search embedded in the methodology increases the number and diversity of

alternatives considered. This search reduces the probability of not considering a deployment scheme that would return the greatest benefits.

The simulation-optimization methodology's largest weakness is the required execution time. Determining the optimal control policy for each ramp meter requires a considerable amount of time. One solution for reducing the required time without sacrificing a great deal of performance is to limit the number of ramp metering algorithms being considered. For example, this research examined and implemented four different traffic responsive, local ramp metering algorithms. Similar results likely would have been generated if only one proven technique, such as occupancy control, were considered.

The high number of replications required to accurately evaluate each deployment scheme also makes the methodology quite time consuming. Unfortunately, the benefits produced by ramp metering are only moderately larger than the variance in performance due to the randomness in the simulation. Therefore, at least twenty replications were required to accurate evaluate the performance of each deployment scheme.

Despite this large execution time, with slight modifications this methodology can be efficiently used in the future. The user should first design several deployment schemes that he or she feels would perform well. These deployment schemes should be listed on a "high potential list." Next, the simulation-optimization methodology should be executed, but only a small number of replications, such as five, should be performed to evaluate the deployment schemes. The highest performing deployment schemes generated by the methodology should be added to the high potential list. Lastly, all the

deployment schemes on the high potential list should be evaluated using an appropriately high number of replications.

This process will minimize the major weakness of the methodology without detracting from either of its major strengths. The user still objectively evaluates a large number of diverse deployment schemes. The timesaving results from not evaluating poorly performing deployment schemes as rigorously as schemes that perform well.

7.3 Future Research

This research provides a solid foundation for the exploration of the relationship between the benefits and cost of ramp metering. However, four areas of future research have been identified. These four areas are:

1. **System-wide Metering Algorithms:** The next logical step is to build a performance-cost curve using system-wide and other complex algorithms. This research would investigate levels of spending and benefit generation that were beyond the scope of this project. The research should also be able to locate the crossover point where centrally controlled ramp meters become more cost effective than multiple local meters.

2. **Additional Performance Metrics Analysis:** Two stochastic measures of performance used by this study were the average system-wide speed and the accident rate. Further research could locate better predictors for accident rate or a better metric for measuring the reduction in delay. Furthermore, additional MOP's may need to be added to measure benefits that were not captured in this study.

3. **Diversion:** When implemented, ramp metering systems often provide greater benefits than observed during simulation testing. Bruce Hellinga and Michel Van Aerde feel this is partially due to a traffic phenomenon that is often not included in simulations called diversion [17]. Diversion occurs when the queue at a ramp meter influences travelers to use arterial streets rather than the freeway for short trips. For many short trips the improved flow on the mainline does not make up for the wait experienced on the entrance ramp. If diversion occurs, flow on the freeway is further improved due to the lower demand. A simulation model that includes both the freeway and parallel arterials could incorporate diversion and possibly measure benefits that were not captured by this research.

4. **Field Implementation Data:** The introduction states that the costs and benefits derived from ramp metering implementation are difficult to estimate since ramp meter installation often occurs at the same time as other improvements. Although it may be difficult to obtain, analyzing the costs and benefit of actual ramp metering systems would provide a large insight into the nature of this relationship. In addition to searching for past data, processes or even legislative policies should be implemented that mandate the collection of detailed design information, cost estimates, and benefit estimates from all future ramp metering installations.

References

1. "400 Ramp Meters in Twin Cities Face Suspension." <u>The Urban Transportation Monitor</u> 17 March 2000: p. 1.

2. Alexiadis, Vassili. E-mail to the author. 30 September 1999.

3. Alexiadis, Vassilios, James W. Schmidt, and JHK & Associates. <u>Ramp Metering: A System Concept Design Methodology.</u> Proceedings of the Fourth Annual IVHS Meeting. Atlanta, Georgia. April 1994.

4. Apogee/Hagler Bailly. <u>Intelligent Transportation Systems: Real World Benefits.</u> USDOT Report FHWA-JPO-98-018. Arlington, Virginia: January 1998.

5. Arnold, E.D. <u>Ramp Metering: A Review of the Literature.</u> Virginia Transportation Research Council Technical Report VTRC99-TAR5. December 1998.

6. Au, Tung and Thomas P. Au. <u>Engineering Economics for Capital Investment Analysis.</u> Boston: Allyn and Bacon, Inc., 1983.

7. Beaulieu, Matthew. E-mail to the author. 11 October 1999.

8. Benekohal, R.F., and G. Abu-Lebdeh. <u>Variability Analysis of Traffic Simulation Outputs: Practical Approach for TRAF-NETSIM.</u> Proceedings of the Transportation Research Board 73rd Annual Meeting., Washington D.C., January 1994.

9. Carvell, James D., et al. <u>Freeway Management Handbook.</u> USDOT Report FHWA-SA-97-064. August 1997.

10. Cheslow, Melvyn. "The ITS Cost Data Repository at Mitretek Systems". Mitretek Systems. Inc. Working Paper. November 1998.

11. Cheu, Jin, et. al. "Calibration of FRESIM for Singapore Expressway Using Genetic Algorithm." <u>Journal of Transportation Engineering</u> Vol 124 no 6 (Nov/Dec 98): 526.

12. Devore, Jay L. <u>Probability and Statistics for Engineering and the Sciences: Fourth Edition.</u> Washington, DC: Brooks/Cole Publishing Company, 1995.

13. Dong, Haikun and Bin Ran. <u>Benefit/Cost Analysis on Ramp Metering: Optimization of TravelTime and Cost Factors</u>. Compendium of Technical Papers for the 66th ITE Annual Meeting. Minneapolis, Minnesota. September 1996.

14. Ehrhart, Angela A. "The Effect of Speed, Flow, and Geometric Characteristics on Crash Frequency on Different Types of Highways." Master's Thesis Presented to the University of Virginia, May 1999.

15. Garber , N.J. and Ravi Gadiraju. <u>Factors Affecting Speed Variance and Its Influence on Accidents.</u> *AAA Foundation for Traffic Safety.* Falls Church, VA. July 1988.

16. Gordon, Robert L. "Algorithm for Controlling Spillback from Ramp Meters." <u>Transportation Research Record</u> No. 1554. 1996: 162-171.

17. Hellinga, Bruce and Michel Van Aerde. "Examining the Potential of Using Ramp Metering as a Component of an ATMS". <u>Transportation Research Record</u> No. 1494. 1995: 75-83.

18. ITT Industries, Inc. <u>TSIS User's Guide: Version 4.3</u>. May 1999.

19. Keller, Gerald, Brian Warrack, and Henry Bartel. <u>Statistics for Management and Economics</u>. Belmont, Ca: Wadsworth, 1990.

20. Kelton, W. David, Randall P. Sadowski, andDeborah A. Sadowski. <u>Simulation With Arena.</u> New York: McGraw-Hill, 1998.

21. May, Adolf D. <u>Traffic Flow Fundamentals</u>. Englewood Cliffs, NJ: Prentice Hall, 1990.

22. Michalewicz, Z. and D.B. Fogel. <u>How to Solve It: Modern Heuristics</u>. New York: Springer, 1999.

23. Papageorgiou, M, H. Hadj-Salem, and F. Middleham. "ALINEA Local Ramp Metering: Summary of Field Results." <u>Transportation Research Record</u> No. 1603. 1997: 90-98.

24. Payne, Harold J., Stephen Thompson, and Gang-Len Chang. <u>Calibration of FRESIM to Achieve Desired Capacities.</u> Proceedings of the Transportation Research Board 76th Annual Meeting., Washington D.C., January 1997.

25. Piotrowicz, Gary and James Robinson. <u>Ramp Metering Status in North America: 1995 Update.</u> USDOT Report DOT-T-95-17. Washington, DC: June 1995.

26. Proper, Allen T. <u>Intelligent Transportation System Benefits: 1999 Update.</u> USDOT Report FHWA-OP-99-012. Washington, DC: May 1999.

27. Proper, Allen T, and Melvyn D. Cheslow. <u>ITS Benefits: Continuing Successes and Operational Test Results.</u> USDOT Report FHWA-JPO-98-002. Washington, DC: October 1997.

28. Roess, Roger P., and Jose M. Ulerio. <u>Comparison of the 1994 Highway Capacity Manual's Ramp Analysis Procedures and the FRESIM Model.</u> National Cooperative Highway Research Program Report 385. Washington, DC: 1997.

29. Ruthi, Ajay K. "Models for Assessing the Impact and the Potential Benefits of Intelligent Transportation System". Paper for IVHS Research Division of Joint Program Service. August 1995.

30. Salwin, Arthur E. Key Findings from the Intelligent Transportation Systems Program: What Have we Learned? Mitretek Systems, Inc. Publication MP 96W0000195. McLean, Virginia: October 1996.

31. Sisiopiku, Virginia P. and Mousa F. Abbasi. Performance Assessment of FRESIM Model in Simulating Ramp Metering Schemes. Compendium of Technical Papers for the 66th ITE Annual Meeting. Minneapolis, Minnesota, September 1996.

32. Thesen, Arne and Laurel E. Travis. Simulation for Decision Making. St. Paul, Mn: West Publishing Company, 1992.

33. Thompson, Nick and Selvin Greene. Ramp Metering for the 21st Century: Minnesota's Experience. Proceedings of the Seventh Annual ITS America Meeting and Exposition. Washington, DC, April 1997.

34. Turner, Shawn., et al. ITS Benefits: Review of Evaluation Methods and Reported Benefits. USDOT Report FHWA/TX-99/1790-1. Austin, Texas: October 1998.

35. Washington State Department of Transportation. Surveillance, Control, and Driver Information Implementation and Operations Plan. Seattle, Washington: September 1998.

36. Wang, Yuhao, and Panos D. Prevedouros. "Comparison of INTEGRATION, TSIS/CORSIM, and WATSim in Replicating Volumes and Speeds on Three Small Networks." Transportation Research Record No.1644. 1998: 80-92.

Appendix A: Cost Rules

      This appendix describes the rules used to calculate the cost of each deployment scheme. The costs are broken into three major categories: initial costs, installation costs, and maintenance and operation costs.

Initial Costs

The initial cost is the same regardless of the complexity of the deployment scheme.

- $10,000 needed for engineering staff to perform traffic analyses and generate ramp metering system design.
- $3000 required for publicity needed to educate the public about the meters
- Total initial cost is $13,000.   [13]

Installation Costs

Installation costs can be broken into two types. Fixed installation costs are the same regardless of the complexity of the deployment scheme. Variable installation costs depend on the amount of equipment deployed.

Fixed Installation Costs

- Communication (twisted pair wire) from detectors to meters: $79,040
- Communication Hub: $100,000
- Total fixed installation cost is $179,040

Variable Installation Costs

- Vehicle Detectors: $2,200 per loop (1 loop needed in each lane)
- 170 series processor: $6,250 each (1 needed per detector group)
- Ramp Meter: $35,000 per interchange (regardless of pre-timed or traffic responsive)
                                                                               [10]

Maintenance and Operating Costs

Annual maintenance and operating costs are estimated to be 10% of the total installation cost [10]. These costs were calculated using a five-year planning horizon.

Appendix B: Traffic Arrival Distributions

CORSIM Distribution Options

When configuring a CORSIM run, the user must specify both the flow rate (in vehicles per hour) at all arrival nodes as well as the statistical distribution that the vehicle inter-arrival times follow. The flow rates are easily calculated for any time and location using historical data from the test area stored in the Smart Travel Laboratory. CORSIM allows the user to choose the uniform distribution, normal distribution, or Erlang distribution to model the inter-arrival times. The traffic data stored in the Smart Travel Laboratory contains only vehicle counts and does not include any information about the distribution of the vehicle arrivals. More research needed to be performed to determine which distribution accurately models traffic arrivals in the test area.

Traffic Flow Research

Chapter 2 of <u>Traffic Flow Fundamentals</u> by Adolph D. May discusses distributions that can be used to model traffic arrivals. May states that the Normal Distribution can accurately model vehicle headways at the highest flow level. At this flow level, the freeway is being used near capacity and there are very few large gaps between vehicles. Each vehicle is simply following the vehicle in front of it. Some drivers will follow closely and other will follow further away, but the majority will follow at some intermediate distance. This phenomenon gives the headways a bell-shaped curve that can be approximated by the normal distribution [21].

The intermediate headway state occurs when traffic flows are still high, but not close enough to capacity to fit the normal distribution. Most vehicles are following the vehicle in front of it, but there are some gaps in the traffic flow. These gaps are larger

headways and cause the headway distribution to have a very pronounced and significant right tail. May states that this intermediate headway state can be accurately modeled with the Pearson Type III Distribution Model Family. May also notes that when shift parameter (α) equals zero and shape parameter (K) is a positive integer, the Pearson Type III distribution model becomes the simpler Erlang distribution model.

Data Collection and Analysis

In order to determine the most appropriate inter-arrival distribution, data from the test area was collected and analyzed. Test area data were manually collected by observing traffic cameras in the Smart Travel Laboratory and recording the vehicle headways. Data were collected on Thursday, February 10, 2000 and Friday, February 11, 2000 from 6:15AM to 7:15AM. Additionally, data were collected from all three through lanes at three different cameras locations near the upstream entrance of the test area (North Hampton, Military, and North – Military). Histograms of each day's data are provided below.
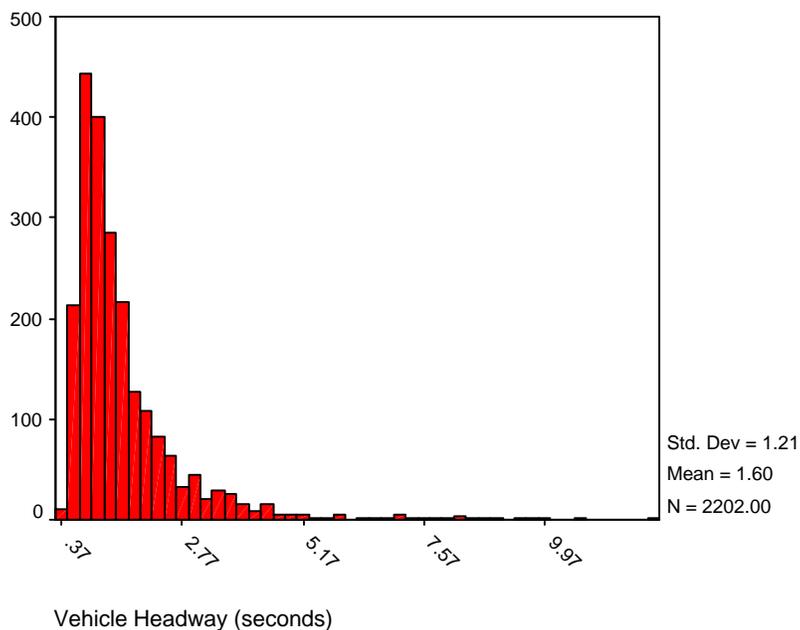
Figure B.1: Histogram of Inter-arrival Times (6:15AM – 7:15AM Thursday,

February 10, 2000)
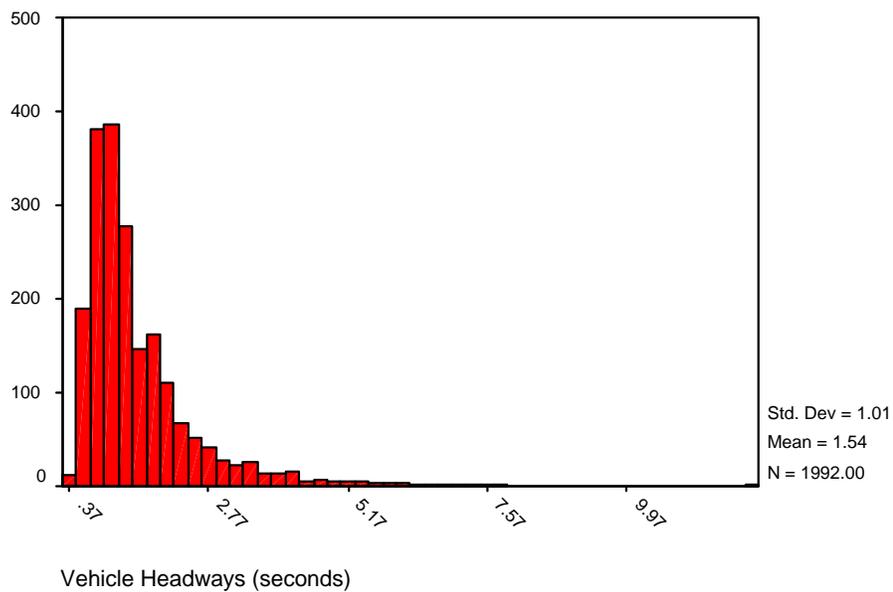
Figure B.2: Histogram of Inter-arrival Times. 6:15AM – 7:15AM Friday,

February 11, 2000

The histograms show that the frequency of inter-arrival times rises quickly from zero seconds to a peak near 0.9 seconds. After the peak the frequency drops quickly (although not nearly as quickly as between zero and the peak), but then has a long right tail. The distribution is not symmetrical and it appears that a normal distribution would not be appropriate. It appears that an Erlang distribution will provide the closest fit. CORSIM allows the user to specify an Erlang shape parameter (K) with an integer value between 1 and 4. Traffic Flow Fundamentals states that K can be estimated by dividing the sample mean by the sample standard deviation. The tables below contain estimates of K for many different data sets. The data were partitioned into different groups to analyze what effect variables have on the estimate of K.

Description	# Observations	Mean headway	Std. Dev.	K
All Data (2/10 & 2/11)	4196	1.57	1.12	1.40
All 2/10 Data	2203	1.60	1.21	1.32
All 2/11 Data	1993	1.54	1.01	1.53
Heavier Traffic	1197	1.47	0.77	1.92
Normal & Lighter Traffic	2999	1.61	1.23	1.31
Lane 1	399	1.83	1.27	1.45
Lane 2	1407	1.60	1.09	1.47
Lane 3	2390	1.51	1.10	1.37

Table B.1: Arrival Data Partitioned into Logical Groupings

Description	# Observations	Mean headway	Std. Dev.	K
North-Military (Lane 3)	400	1.56	1.34	1.16
North Hampton (Lane 2)	210	1.78	1.49	1.19
Military (Lane 3)	399	1.49	0.99	1.51
North Hampton (Lane 3)	399	1.58	1.40	1.13
North-Military (Lane 3)	399	1.41	0.64	2.21
Military (Lane 1)	399	1.83	1.27	1.45

Table B.2: February 10, 2000 – Raw Data from each Measurement

Description	# Observations	Mean headway	Std. Dev.	K
North Hampton (Lane 3)	399	1.55	1.25	1.24
North – Military (Lane 3)	399	1.52	0.94	1.63
North – Military (Lane 2)	399	1.55	0.92	1.69
North Hampton (Lane 2)	399	1.62	1.13	1.43
Military (Lane 3)	399	1.44	0.71	2.04

Table B.3: February 11, 2000 – Raw Data from each Measurement

When using all collected data, the estimate of K is 1.40. The value of K seems pretty consistent except for one factor – the level of congestion on the roadway during the data collection. When the road contained heavier traffic (e.g. cars travelling 40 – 50 mph due to congestion), the estimate for K is 1.92. When the road contains lighter traffic (e.g. Cars travelling close to their desired freeflow speed (60 mph or greater), the estimate of K is 1.32. Since CORSIM specifies the shape parameter must be an integer, the shape parameter must be set equal to 1 or 2.

The shape parameter was set to 2 for two major reasons. Primarily, an Erlang distribution with shape parameter 1 is known as the negative exponential distribution. The negative exponential distribution makes two assumptions:

1. Any point in time is as likely to have a vehicle arriving as is any other point in time.

2. The arrival of one vehicle at a point in time does not affect the arrival time of any other vehicle.

On freeways, individual headway times are rarely less than 0.5 seconds (on the order of 1 to 2 percent) and are almost never less than 0.2 seconds. This fact violates the second assumption which states that one vehicle arrival will not affect other vehicle arrivals.

The second reason for choosing a shape parameter of 2 involves the observations with heavy traffic flow. During these periods the estimate of K was close to 2 and two individual collection periods actually had K estimates greater than 2. It appears that a shape parameter of 2 accurately models the distributions of headways during heavier traffic. Since these heavy traffic situations occurs at the test area during the test time (morning rush hour), it is reasonable to model the headway distribution with an Erlang distribution and a shape parameter of 2.

Appendix C: Simulation Model Calibration

Much of CORSIM's internal logic contains parameters that determine how individual vehicles in the simulation behave. Parameters include attributes such as how long it takes a vehicle to change lanes, how closely one vehicle will follow another, a driver's desired speed, a driver's willingness to yield to other vehicles, etc. When all vehicles are aggregated, these individual parameters can have a large effect on the flow of traffic and can affect import macroscopic parameters such as the capacity of the roadway. Past research has shown that the default values that CORSIM provides for these parameters often lead to simulations that do not closely model the actual traffic pattern of the roadway. In order to gain accurate simulation results the model must be calibrated by adjusting individual parameters so that the behavior of the vehicles in the model resembles the behavior of vehicles in real life [11, 36]

Historical traffic data stored by the Smart Travel Lab was used to calibrate the CORSIM model. Two twenty-minute data sets were selected for the calibration process. Data from 6:30AM to 6:50AM on August 11, 1999 (a Wednesday) was used as the training set to search for the most accurate parameters. Then data from 6:30AM to 6:50AM on August 23, 1999 (a Monday) were used to verify the accuracy of the parameters. Lastly, a sixty-minute data set from 6:40AM to 7:40AM on August 23, 1999 was evaluated to ensure that the simulation could accurately model a longer time period.

A total of thirty-two simulation runs were completed with the training test set. On each test the speeds and volumes on the mainline links in the simulation were compared to speed and volume data gathered from vehicle detectors on the mainline. Multiple

replications were performed to ensure that improvements in the output were due to the current parameter values and not randomness in the simulation.

On the first run, the freeflow speed of each link was set at 65 miles per hour and CORSIM's default parameters were used for all other parameters. This run produced an average speed error of –39.6% and an average absolute speed error of 41.4%. The average error for volume was –20.4% and the average absolute error for volume was 22.0%. Since both the speed and volume in the model is less than the actual data, it is obvious that the vehicles in the simulation are not behaving as aggressively as the actual drivers in the test area.

Previous research indicates that the parameters with the largest impact on the simulation output are the free flow speed of each link and the car-following array [11, 36]. The free flow speed is the speed a vehicle will travel if no slower vehicles are blocking the road in front of it. The car-following array matches a desired headway separation to each of the ten driver types. Smaller desired headways will increase the capacity of the road.

The next eight simulation runs were spent adjusting the free flow speed and driver aggressiveness parameters with the goal of making the average error as close to zero as possible and the average absolute error as little as possible. The best results were achieved when the car following array was set from 10 – 1 and the free flow speed was set at 70 miles per hour for all mainline links. However, even with these parameters the average error of the speed was –31.1% with an average absolute error was 31.9% and the average error for volume was –3.1% with an average absolute error of 11.2%. Although the results are improved the speed error is unacceptably high.

Close inspection of the simulation's animation showed that traffic was becoming extremely congested in one section of the model. This section contains a left exit and right exit in proximity and requires many cars to change lanes in order reach their desired exit. This weaving section does cause slowdowns on the actual section of I-64, but not nearly as severe as the congestion being experienced in the simulation. From watching the animation it was obvious that vehicles were having difficulty making the lane changes needed to reach their exits. In some cases vehicles were evening stopping in the travel lanes to wait for an opening to make a lane change.

Many parameters were changed in order to allow the vehicles to make lane changes similar to actual drivers. First, the position of the warning signs that alert drivers of the upcoming exits were changed. Additionally, the time required to make a lane change was reduced to one second. Lastly, the percentage of drivers willing to yield to another vehicle changing lanes was increased to forty percent.

After these changes were made, the driver aggressiveness levels were reexamined to ensure that they still provided the best results. After performing multiple replications, it was clear that the final parameter values allowed the simulation to most accurately model reality. When evaluating these parameters with the training data, the average error of the speed was –11.3% and the average absolute speed error was 19.8%. The average volume error was –0.3%, with an average absolute error of 7.7%

Using the final parameter values, the simulation was run with the test data (August 23, 1999) to ensure that the parameters were not tailored to the August 11 data. The errors with the test set were similar, but a little higher than the errors generated from the August 11 training data. The average speed error was 2.5% and the average absolute

speed error was 24.9%. The average volume error was 3.1% and the average absolute volume error was 11.6%. Interestingly, while the training set produced negative speed and volume errors, the test set produced positive speed and volume errors. While inspecting the test data set, an unusual slowdown that occurred for eight minutes on several of the links was noticed. It appears that on August 11, 1999 a slowdown occurred due to some minor incident. Whether a stalled car, debris on the road, or some other occurrence, this incident slowed traffic causing the model to produce speeds and volumes greater than experienced on August 11, 1999. This explains the higher than expected error on the test set and indicates that the parameter values are valid for both the training data set and the test data set.

A final calibration study was performed using a sixty-minute data set that ranged from 6:40AM to 7:40AM on August 23, 1999. This test used the same parameters and was performed to ensure that the simulation could accurately portray traffic for longer time periods. The results of this study were quite favorable and indicate that sixty-minute time periods can also be accurately modeled. The average speed error was -2.7% with an average absolute speed error of 12.4%. The average volume error was 5.7% and the average absolute volume error was 10.9%.

Appendix D: Traffic Responsive Algorithm Evaluation Results

This Appendix contains the results of the traffic responsive algorithm evaluation tests. The results of this test were used to determine which ramp metering algorithm would control the ramp meter in each state. The degree of confidence field describes how confident we can be that the best performing algorithm has an average speed higher than the other algorithms. The value of the "n" field is how many times the algorithm and parameter value set were tested.

Ramp 1

	Mean Speed	**Standard Dev.**	**Accident Rate**	**Standard Dev.**	**n**	**Z**	**Confidence**
			Immediate Upstream Sensor				
Occupancy - Table 1	42.32	1.84	0.01282	0.00280	42		
Speed - Table 6	42.12	1.52	0.01167	0.00259	42	0.538	70.5%
Demand - Table 4	42.10	1.80	0.01328	0.00335	40	0.535	70.3%
Speed - Table 4	42.09	1.86	0.01245	0.00280	42	0.574	71.6%
Speed - Table 5	42.06	1.88	0.01209	0.00191	42	0.650	74.2%
Demand - Table 1	41.95	1.82	0.01307	0.00293	40	0.917	82.0%

	Mean Speed	**Standard Dev.**	**Accident Rate**	**Standard Dev.**	**n**	**Z**	**Confidence**
			Immediate Downstream Sensor				
Occupancy - Table 3	42.18	2.29	0.01323	0.00417	35		
Alinea - Parameter 4	42.10	1.97	0.01346	0.00304	40	0.170	56.7%
Occupancy - Table 4	42.07	1.49	0.01307	0.00242	35	0.236	59.3%
Speed - Table 6	41.99	1.77	0.01290	0.00292	35	0.403	65.6%

	Mean Speed	**Standard Dev.**	**Accident Rate**	**Standard Dev.**	**n**	**Z**	**Confidence**
			Second Downstream Sensor				
Occupancy - Table 4	42.10	1.66	0.01279	0.00233	35		
Speed - Table 6	42.08	1.84	0.01334	0.00384	35	0.052	52.0%
Speed - Table 4	42.03	2.78	0.01368	0.00333	35	0.116	54.6%
Alinea - Parameter 4	41.46	1.70	0.01417	0.00332	40	1.652	95.5%

Table D.1: Traffic Responsive Algorithm Evaluation Results – Ramp 1

Ramp 2

Immediate Upstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Occupancy - Table 1	42.63	1.93	0.01160	0.00317	28		
Speed - Table 5	42.39	3.39	0.01263	0.00313	28	0.325	62.7%
Demand - Table 2	42.21	1.87	0.01288	0.00306	40	0.891	81.3%
Speed - Table 6	42.01	2.03	0.01401	0.00281	28	1.164	85.6%
Demand - Table 4	41.87	1.67	0.01281	0.00196	40	1.694	95.5%

Immediate Downstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Speed - Table 6	42.42	2.05	0.01262	0.00232	35		
Speed - Table 4	42.13	2.01	0.01328	0.00343	35	0.591	72.3%
Occupancy - Table 1	41.90	1.75	0.01334	0.00302	35	1.144	87.3%
Alinea - Parameter 3	41.07	2.05	0.01396	0.00317	40	2.852	99.8%

Table D.2: Traffic Responsive Algorithm Evaluation Results – Ramp 2

Ramp 3

Immediate Upstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Occupancy - Table 4	42.25	1.98	0.01422	0.00440	28		
Speed - Table 2	42.20	1.64	0.01257	0.00209	28	0.093	53.6%
Demand - Table 2	42.16	2.16	0.01385	0.00454	40	0.177	57.0%
Demand - Table 5	41.74	2.13	0.01437	0.00410	40	1.005	84.1%

Immediate Downstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Occupancy - Table 2	42.45	1.96	0.01372	0.00342	35		
Occupancy - Table 5	42.37	1.79	0.01304	0.00277	35	0.179	57.1%
Speed - Table 5	41.87	1.86	0.01313	0.00321	35	1.269	89.7%
Alinea - Parameter 5	41.53	1.87	0.01367	0.00341	40	2.073	98.1%

Table D.3: Traffic Responsive Algorithm Evaluation Results – Ramp 3

Ramp 4

Immediate Upstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Speed - Rate 5	42.28	1.94	0.01205	0.00196	28		
Occupancy - Rate 5	42.11	1.60	0.01298	0.00190	28	0.344	63.5%
Demand - Rate 3	41.93	2.05	0.01371	0.00333	40	0.717	76.3%
Demand - Rate 1	41.82	2.47	0.01382	0.00372	40	0.846	80.1%

Immediate Downstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Occupancy - Table 2	42.53	2.25	0.01283	0.00338	35		
Occupancy - Table 3	42.48	1.41	0.01307	0.00254	35	0.102	54.0%
Occupancy - Table 1	42.27	1.56	0.01393	0.00275	35	0.570	71.6%
Speed - Table 5	41.84	1.99	0.01339	0.00285	35	1.353	90.9%

Table D.4: Traffic Responsive Algorithm Evaluation Results – Ramp 4

Ramp 5

Immediate Upstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Speed - Rate 3	42.45	2.12	0.01374	0.00343	28		
Speed - Rate 2	42.30	1.81	0.01347	0.00310	28	0.285	61.2%
Occupancy - Rate 5	42.16	1.98	0.01410	0.00382	28	0.515	69.7%
Demand - Rate 1	42.11	2.01	0.01378	0.00286	40	0.656	74.4%
Occupancy - Rate 3	41.97	1.85	0.01205	0.00236	28	0.899	81.9%
Alinea - Rate 2	41.91	1.87	0.01290	0.00382	40	1.087	86.1%
Demand - Rate 2	41.73	2.00	0.01363	0.00301	40	1.405	92.0%
Alinea - Rate 3	41.68	2.20	0.01464	0.00481	40	1.451	92.7%

Table D.5: Traffic Responsive Algorithm Evaluation Results – Ramp 5

Ramp 6

Immediate Upstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Demand - Rate 5	41.07	1.91	0.01228	0.00251	40		
Demand - Rate 3	40.85	2.01	0.01339	0.00331	40	0.522	69.9%
Occupancy - Rate 5	40.67	2.13	0.01261	0.00304	28	0.811	79.1%
Speed - Rate 1	40.67	2.20	0.01310	0.00343	28	0.796	78.7%

Immediate Downstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Occupancy - Rate 3	41.05	1.86	0.01336	0.00343	35		
Alinea - Rate 5	40.96	1.88	0.01371	0.00333	40	0.195	57.7%
Speed - Rate 4	40.71	1.97	0.01338	0.00350	35	0.728	76.6%

Table D.6: Traffic Responsive Algorithm Evaluation Results – Ramp 6

Ramp 7

Immediate Upstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Speed - Rate 6	42.88	1.79	0.01239	0.00346	28		
Speed - Rate 2	42.16	2.04	0.01323	0.00262	28	1.422	92.2%
Demand - Rate 3	42.08	2.10	0.01414	0.00329	40	1.699	95.5%
Demand - Rate 5	41.97	2.19	0.01328	0.00300	40	1.888	97.0%
Occupancy - Rate 3	41.74	1.97	0.01424	0.00360	28	2.274	98.9%

Immediate Downstream Sensor

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	n	Z	Confidence
Speed - Rate 4	42.73	1.87	0.01323	0.00363	35		
Occupancy - Rate 3	42.02	2.36	0.01354	0.00344	35	1.383	91.6%
Alinea - Rate 5	41.98	1.92	0.01376	0.00324	40	1.699	95.5%
Alinea - Rate 4	41.94	1.91	0.01382	0.00338	40	1.806	96.5%
Occupancy - Rate 5	41.86	2.08	0.01347	0.00347	35	1.829	96.6%
Alinea - Rate 3	41.68	1.89	0.01421	0.00324	40	2.407	99.2%

Table D.7: Traffic Responsive Algorithm Evaluation Results – Ramp 7

Appendix E: Clock-Time Algorithm Evaluation Results

All rates were tested using seventy replications. The degree of confidence field describes how confident we can be that the best performing rate has an average speed higher than the second best performing rate.

Ramp 1

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	Z Score	Degree of Confidence
Rate 8	47.37	2.55	0.00809	0.00177	0.359	64.1%
Rate 6	47.22	2.42	0.00816	0.00161		

Ramp 2

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	Z Score	Degree of Confidence
Rate 6	47.92	2.09	0.00784	0.00138	0.925	82.3%
Rate 5	47.59	2.14	0.00799	0.00126		

Ramp 3

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	Z Score	Degree of Confidence
Rate 10	47.72	2.14	0.00783	0.00126	0.070	52.8%
Rate 4	47.69	2.57	0.00796	0.00153		

Ramp 4

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	Z Score	Degree of Confidence
Rate 5	47.56	2.41	0.00805	0.00181	0.288	61.4%
Rate 8	47.45	2.10	0.00795	0.00123		

Ramp 5

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	Z Score	Degree of Confidence
Rate 6	47.05	2.13	0.00813	0.00157	0.170	56.8%
Rate 8	46.99	2.46	0.00818	0.00214		

Ramp 6

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	Z Score	Degree of Confidence
Rate 4	46.84	2.25	0.00811	0.00203	0.223	58.7%
Rate 5	46.75	2.24	0.00795	0.00167		

Ramp 7

	Mean Speed	Standard Dev.	Accident Rate	Standard Dev.	Z Score	Degree of Confidence
Rate 6	47.53	2.06	0.00802	0.00165	1.597	94.5%
Rate 10	46.92	2.47	0.00852	0.00197		

Table E.1: Clock-Time Algorithm Evaluation Results

Appendix F: Higher Replication Test Results

The following pages display the results of deployment schemes re-tested using between eight and ten replications. These results, termed "Evaluation 2", are compared to the previous results, termed "Evaluation 1", which were generating using only three to five replications.

Deployment Information			Evaluation 1				Evaluation 2				Comparison	
Deployment Scheme	Cost	Data Set Used	Total Score	Rank	Variance	N	Total Score	Ranking	Variance	N	Absolute Score Difference	Absolute Rank Difference
1000010001000000000000000111	$375,910	11-Aug-99	56.37	4	34.4	3	64.86	1	133.7	10	8.49	3
0000000000000000000000000111	$369,310	11-Aug-99	43.62	12	161.4	3	64.13	2	365.7	10	20.50	10
0100000010000000000000000000	$386,560	11-Aug-99	41.07	13	302.2	3	62.99	3	144.9	8	21.92	10
1000000100000000000000000000	$372,610	11-Aug-99	18.28	21	1429.2	3	61.90	4	102.5	8	43.63	17
1000010010100000000000000011	$345,985	11-Aug-99	55.10	5	713.1	3	60.04	5	481.1	10	4.94	0
0100000000010000000000000000	$372,610	11-Aug-99	65.06	2	101.1	5	58.70	6	48.1	8	6.35	4
0100010000000000000000000000	$386,560	11-Aug-99	38.48	15	2883.5	3	56.91	7	122.4	8	18.42	8
1100000101000000000000000000	$386,560	11-Aug-99	66.51	1	76.5	5	56.05	8	399.0	8	10.46	7
0000001011000000000000000000	$375,910	11-Aug-99	40.23	14	67.0	3	54.58	9	65.4	8	14.35	5
0000011100010000000000000000	$379,210	11-Aug-99	38.43	16	1544.1	3	53.23	10	662.5	8	14.80	6
0000000001000100000000000000	$372,610	11-Aug-99	50.65	9	52.4	3	52.61	11	467.9	8	1.96	2
0000000100000010000000000010	$386,560	11-Aug-99	51.31	8	833.5	4	50.80	12	465.6	8	0.52	4
0000000000010010000000000000	$356,635	11-Aug-99	52.87	7	959.4	5	50.73	13	172.4	8	2.14	6
0000100010100000000000000000	$386,560	11-Aug-99	57.53	3	46.0	5	49.81	14	287.9	8	7.71	11
0000000000100010010000000000	$386,560	11-Aug-99	45.13	11	144.2	5	48.54	15	116.3	10	3.41	4
0000000101101000000000000000	$386,560	11-Aug-99	53.18	6	855.1	3	47.85	16	323.4	8	5.33	10
0000001101001000000000000000	$386,560	11-Aug-99	29.25	17	188.4	3	44.76	17	1165.0	8	15.51	0
0000000010000100000000000000	$375,910	11-Aug-99	24.96	19	908.7	3	43.92	18	879.5	8	18.96	1
0000000000000110101000000000	$386,560	11-Aug-99	26.16	18	269.7	3	43.63	19	1003.8	8	17.47	1
0000000000000110101	$385,285	11-Aug-99	50.04	10	212.7	3	40.94	20	353.3	8	9.10	10
0000010000010000000000000000	$386,560	11-Aug-99	18.54	20	1002.6	5	40.46	21	610.0	8	21.92	1
0101000000000000000000000000	$372,610	23-Aug-99	46.23	16	779.5	3	55.44	1	75.3	10	9.21	15
0000001000011000000000000000	$372,610	23-Aug-99	54.92	7	125.5	5	54.13	2	221.1	10	0.79	5
0000000000001010000000000000	$386,560	23-Aug-99	52.14	9	167.7	3	54.08	3	226.3	8	1.93	6
1000011000001000000000000000	$375,910	23-Aug-99	54.88	8	765.8	4	53.24	4	150.2	10	1.64	4
0110000001000000010000000000	$375,910	23-Aug-99	55.18	5	137.6	5	52.80	5	352.6	10	2.38	0
0000000100000001000000000000	$386,560	23-Aug-99	46.26	15	75.1	5	52.68	6	80.5	8	6.42	9
0100000001000000001000000000	$386,560	23-Aug-99	43.65	17	48.9	3	52.02	7	208.5	8	8.37	10
1000000010000000000000000000	$386,560	23-Aug-99	30.72	20	209.8	3	51.73	8	425.5	8	21.01	12
0100000001000000010000000000	$375,910	23-Aug-99	58.84	2	112.6	5	51.43	9	505.0	10	7.41	7
1000011000001000000000000000	$372,610	23-Aug-99	55.14	6	455.3	3	50.63	10	75.9	8	4.51	4
0000010000001000001000000000	$372,610	23-Aug-99	46.97	14	29.8	3	47.56	11	124.0	8	0.59	3
0000000010000001000000000000	$386,560	23-Aug-99	51.74	11	123.7	3	46.78	12	277.7	8	4.96	1
0010000001100000000000000000	$386,560	23-Aug-99	32.88	19	509.7	3	46.32	13	169.8	8	13.44	6
1000000010101000000000000000	$386,560	23-Aug-99	49.28	13	56.6	3	45.99	14	404.0	8	3.29	1
0000000100000000011100000000	$375,910	23-Aug-99	52.14	10	316.1	3	45.77	15	279.0	10	6.37	5
0000000001000000010000000000	$375,910	23-Aug-99	39.81	18	2687.2	3	45.28	16	224.8	10	5.47	2
0000000000000000000000011100	$372,610	23-Aug-99	57.66	3	67.3	3	43.69	17	180.6	8	13.97	14
0000000000100000000000000010	$386,560	23-Aug-99	63.43	1	327.8	3	42.98	18	209.4	8	20.44	17
0000010000010000000000000000	$386,560	23-Aug-99	51.74	12	154.6	5	41.76	19	259.0	7	9.99	7
0000000101011000000000000000	$375,910	23-Aug-99	56.13	4	22.9	5	41.69	20	807.4	8	14.44	16
										average	10.35	6.4

Table F.1: Comparison of Scores of All Deployment Schemes in Budget One

Deployment Information			Evaluation 1				Evaluation 2				Comparison	
Deployment Scheme	Cost	Data Set Used	Total Score	Rank	Variance	N	Total Score	Ranking	Variance	N	Score Difference	Absolute Rank Difference
1111101111111111000000	$817,210	11-Aug-99	68.87	10	188.2	5	63.88	1	206.5	10	-4.99	9
1111111111111101000000	$817,210	11-Aug-99	68.69	14	199.4	5	63.08	2	363.4	8	-5.62	12
0101111101111111100000	$722,860	11-Aug-99	68.72	12	139.1	3	61.63	3	294.2	8	-7.09	9
0011011111101111100000	$710,185	11-Aug-99	74.86	1	42.9	5	61.40	4	370.7	10	-13.46	3
1011100111110101000000	$615,835	11-Aug-99	68.72	13	55.2	5	57.87	5	433.8	8	-10.85	8
0000000001001100101000	$444,385	11-Aug-99	73.39	3	85.8	5	57.79	6	510.4	8	-15.60	3
0110100110101111100000	$665,035	11-Aug-99	72.06	6	186.5	3	57.37	7	286.1	10	-14.69	1
0001100000110010000000	$476,335	11-Aug-99	69.02	9	139.6	5	56.98	8	425.8	9	-12.04	1
1010001011101110000000	$619,135	11-Aug-99	68.83	11	10.4	5	56.21	9	404.4	8	-12.62	2
1010110101101010000000	$649,060	11-Aug-99	68.59	15	101.4	3	56.21	10	375.9	8	-12.38	5
0001010101011011000000	$573,985	11-Aug-99	69.66	7	394.6	5	54.64	11	897.4	10	-15.02	4
0010100010000101000000	$542,035	11-Aug-99	67.93	17	133.8	5	54.61	12	309.6	8	-13.32	5
0100111000101111000000	$645,760	11-Aug-99	69.22	8	52.6	3	53.76	13	365.2	8	-15.45	5
0101011111001100000000	$717,535	11-Aug-99	72.65	4	46.7	5	51.86	14	90.6	8	-20.80	10
1110101001101010000000	$658,435	11-Aug-99	72.24	5	183.1	5	47.73	15	870.5	10	-24.51	10
0111111111111111100000	$781,960	11-Aug-99	74.25	2	38.8	4	45.17	16	2326.0	8	-29.08	14
0000010110000000011000	$480,910	11-Aug-99	68.38	16	43.6	3	41.19	17	262.9	8	-27.18	1
										average	-14.98	6.0

Table F.2: Comparison of Scores for Seventeen Highest Scoring Deployment Schemes

Appendix G: Linear Regression Models

Several regression models were built using data from previous simulation runs. Data from both the preliminary results (3-5 replications) and the later results (20 replications) were modeled. The first three models use the preliminary results as data. The last three models use the data with twenty replications.

Response Variables
Total Score, Speed Score, Accident Score

Predictor Variables

Dataset: Binary categorical variable indicating which data set was used (Aug 11, 1999 = 0; August 23, 1999 = 1).

Cost: The cost of the deployment scheme.

NumMet: The total number of ramp meters included in this deployment scheme.

NumSq: The total number of ramp meters squared.

M1 – M7: Binary categorical variables that indicate if a meter is located on each entrance ramp. For example, if there is a meter located on the third entrance ramp then $M3 = 1$.

MXY: Interaction variables between the individual ramp meters. An interaction variable was created for all pairs of meters that were no more than two exits away. For example M34 is equal to M3 * M4.

T1 – T7: Binary categorical variables that indicate whether a traffic responsive algorithm is in use at each entrance ramp. For example, if there is a meter on ramp 3 and it is being controlled by a traffic responsive algorithm then $T3 = 1$.

Model 1: Response = Total Score; Data = Preliminary Results (3-5 Replications)

```
The regression equation is
Total_Score = 42.2 - 2.43 DataSet +0.000021 Cost - 1.93 M5 - 2.28 M6 -
        3.89 M13- 2.25 M24 + 2.37 M56 + 3.79 M57 - 7.35 M67 - 0.98
        TR1 + 1.38 TR3 + 1.54 TR5 + 4.12 TR7


Predictor         Coef       StDev           T          P
Constant        42.184       3.840       10.98      0.000
DataSet        -2.4334       0.7538      -3.23      0.001
Cost        0.00002083  0.00000907        2.30      0.022
M5              -1.932       1.152       -1.68      0.094
M6              -2.278       1.550       -1.47      0.142
M13             -3.888       1.433       -2.71      0.007
M24             -2.249       1.099       -2.05      0.041
M56              2.367       2.673        0.89      0.376
M57              3.792       4.794        0.79      0.429
M67             -7.346       4.326       -1.70      0.090
TR1             -0.980       1.186       -0.83      0.409
TR3              1.385       1.294        1.07      0.285
TR5              1.537       1.105        1.39      0.164
TR7              4.115       5.237        0.79      0.432


S = 10.74      R-Sq = 4.1%      R-Sq(adj) = 2.5%

Analysis of Variance

Source            DF           SS          MS          F          P
Regression        13       3975.5       305.8       2.65      0.001
Residual Error   811      93542.6       115.3
Total            824      97518.0
```

Model 2: Response = Speed Score; Data = Preliminary Results (3-5 Replications)

```
The regression equation is
Avg_Speed = 58.7 - 2.87 DataSet -0.000027 Cost - 4.61 M1 - 5.98 M6 +
        2.53 TR5


Predictor         Coef       StDev           T          P
Constant        58.742       2.615       22.47      0.000
DataSet        -2.8727       0.8140      -3.53      0.000
Cost       -0.00002700  0.00000533       -5.07      0.000
M1              -4.609       1.080       -4.27      0.000
M6              -5.976       1.408       -4.24      0.000
TR5             2.5283       0.9722        2.60      0.009


S = 11.66      R-Sq = 12.2%      R-Sq(adj) = 11.7%

Analysis of Variance

Source            DF           SS          MS          F          P
Regression         5      15460.0      3092.0      22.76      0.000
Residual Error   819     111285.5       135.9
Total            824     126745.5
```

Model 3: Response = Accident Score; Data = Preliminary Results (3-5 Replications)

```
The regression equation is
Accident = 48.4 - 2.27 DataSet +0.000020 Cost + 5.06 TR6

Predictor         Coef        StDev          T         P
Constant        48.447        2.833      17.10     0.000
DataSet         -2.267        1.002      -2.26     0.024
Cost       0.00002030   0.00000487       4.17     0.000
TR6              5.063        1.806       2.80     0.005

S = 14.35       R-Sq = 3.3%      R-Sq(adj) = 3.0%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         3      5801.0      1933.7      9.38     0.000
Residual Error   821    169157.8       206.0
Total            824    174958.8
```

Model 4: Response = Total Score; Predictors = All Predictors; Data = Preliminary
Results 20 Replications)

```
The regression equation is
TotalScore = 52.9 - 4.55 DataSet + 2.42 M1 - 4.41 M2 + 8.15 M3 + 1.56M5
           - 14.2 M6 - 5.26 M7 - 4.55 M13 - 2.71 M35 + 11.2 M56 +
           4.83T2 - 3.04 T3 + 1.19 T4 - 1.59 T5

Predictor         Coef        StDev          T         P
Constant        52.919        1.493      35.46     0.000
DataSet        -4.5506       0.9057      -5.02     0.000
M1               2.422        1.360       1.78     0.080
M2              -4.414        1.882      -2.35     0.022
M3               8.155        3.251       2.51     0.015
M5               1.562        1.519       1.03     0.308
M6             -14.214        4.475      -3.18     0.002
M7              -5.255        3.096      -1.70     0.094
M13             -4.554        1.906      -2.39     0.020
M35             -2.706        2.053      -1.32     0.192
M56             11.200        5.427       2.06     0.043
T2               4.833        1.887       2.56     0.013
T3              -3.035        3.295      -0.92     0.360
T4               1.187        1.002       1.19     0.240
T5              -1.595        1.242      -1.28     0.204

S = 3.917       R-Sq = 50.6%      R-Sq(adj) = 39.9%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression        14     1019.56       72.83      4.75     0.000
Residual Error    65      997.35       15.34
Total             79     2016.92
```

Model 5: Response = Speed Score; Predictors = All Predictors; Data = Preliminary Results 20 Replications)

```
The regression equation is
Speed = 53.1 - 4.91 DataSet + 4.06 M1 - 4.11 M2 + 4.65 M3 - 18.6 M6 -
     5.25 M13 + 8.80 M56 - 2.72 T1 + 4.09 T2 - 1.75 T5 - 0.846 NumMet


Predictor        Coef        StDev          T          P
Constant       53.053       1.416       37.47      0.000
DataSet       -4.9087      0.8877       -5.53      0.000
M1              4.056       2.056        1.97      0.053
M2             -4.107       1.840       -2.23      0.029
M3              4.648       1.607        2.89      0.005
M6            -18.570       4.288       -4.33      0.000
M13            -5.252       1.842       -2.85      0.006
M56             8.803       5.246        1.68      0.098
T1             -2.718       2.031       -1.34      0.185
T2              4.092       1.826        2.24      0.028
T5             -1.749       1.061       -1.65      0.104
NumMet        -0.8461      0.7959       -1.06      0.291


S = 3.872      R-Sq = 57.3%      R-Sq(adj) = 50.3%


Analysis of Variance

Source            DF          SS          MS          F          P
Regression        11     1365.95      124.18       8.28      0.000
Residual Error    68     1019.63       14.99
Total             79     2385.59
```

Model 6: Response = Accident Score; Predictors = All Predictors; Data = Preliminary Results 20 Replications)

```
The regression equation is
Accident = 54.0 - 4.97 DataSet + 8.72 M1 + 11.9 M3 + 9.39 M4 + 10.1 M5
           - 5.73 M13 - 5.60 M35 - 3.16 M45 + 8.44 M56 + 4.63 T2 - 2.49
           T5 - 4.44 NumMet

Predictor         Coef       StDev           T          P
Constant        54.043       2.078       26.01      0.000
DataSet         -4.972       1.161       -4.28      0.000
M1               8.720       2.438        3.58      0.001
M3              11.909       3.106        3.83      0.000
M4               9.385       2.293        4.09      0.000
M5              10.094       2.863        3.53      0.001
M13             -5.726       2.333       -2.45      0.017
M35             -5.603       2.493       -2.25      0.028
M45             -3.162       2.595       -1.22      0.227
M56              8.438       4.159        2.03      0.046
T2               4.635       1.925        2.41      0.019
T5              -2.491       1.570       -1.59      0.117
NumMet          -4.441       1.809       -2.45      0.017

S = 5.011      R-Sq = 45.5%      R-Sq(adj) = 35.8%

Analysis of Variance

Source            DF            SS          MS          F          P
Regression        12       1405.94      117.16       4.67      0.000
Residual Error    67       1682.08       25.11
Total             79       3088.02
```