**Research Report**
Agreement T2695, Task 24
Automata Predictions

# Automata Model for Congestion Prediction

by

Zachary Wall and Daniel J. Dailey
ITS Research Program
College of Engineering, Box 352500
University of Washington
Seattle, Washington 98195-2500

**Washington State Transportation Center (TRAC)**
University of Washington, Box 354802
University District Building, Suite 535
1107 N.E. 45th Street
Seattle, Washington 98105-4631

Washington State Department of Transportation
Technical Monitor
Pete Briglia
Manager, Advanced Technology Branch

A report prepared for

**Washington State Transportation Commission**
Department of Transportation
and in cooperation with
**U.S. Department of Transportation**
Federal Highway Administration

December 2003

# TECHNICAL REPORT STANDARD TITLE PAGE

| 1. REPORT NO. | 2. GOVERNMENT ACCESSION NO. | 3. RECIPIENT'S CATALOG NO. |
|---|---|---|
| WA-RD 577.1 | | |

| 4. TITLE AND SUBTITLE | 5. REPORT DATE |
|---|---|
| AUTOMATA MODEL FOR CONGESTION PREDICTION | December 2003 |
| | 6. PERFORMING ORGANIZATION CODE |

| 7. AUTHOR(S) | 8. PERFORMING ORGANIZATION REPORT NO. |
|---|---|
| Zachary Wall and Daniel J. Dailey | |

| 9. PERFORMING ORGANIZATION NAME AND ADDRESS | 10. WORK UNIT NO. |
|---|---|
| Washington State Transportation Center (TRAC)<br>University of Washington, Box 354802<br>University District Building; 1107 NE 45th Street, Suite 535<br>Seattle, Washington 98105-4631 | 11. CONTRACT OR GRANT NO.<br>Agreement T2695, Task 24 |

| 12. SPONSORING AGENCY NAME AND ADDRESS | 13. TYPE OF REPORT AND PERIOD COVERED |
|---|---|
| Research Office<br>Washington State Department of Transportation<br>Transportation Building, MS 47372<br>Olympia, Washington 98504-7372<br>Doug Brodin, Project Manager, 360-705-7972 | Research Report<br>14. SPONSORING AGENCY CODE |

15. SUPPLEMENTARY NOTES

This study was conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. ABSTRACT

The overall goal of this project is to create a method for predicting traffic congestion on freeway corridors. When implemented, it will provide a traffic service like that of "pin-point Doppler" weather radar that can predict growing or dissipating congestion. Preliminary versions of the model used real-time loop data to successfully reproduce traffic behavior under moderately congested conditions. To improve the model for heavily congested conditions, the model had to be accurately calibrated. The process of calibrating the model revealed that inductance loop errors were preventing accurate results. An algorithm to correct the data from improperly functioning loops was developed and published at the 2003 Transportation Research Board Annual Meeting. The corrected loop data from the TDAD (Traffic Data Acquisition and Distribution) data mine are now being used to calibrate the model. The algorithm created in support of this effort can be used with malfunctioning loops to improve the freeway management system performance monitoring effort.

| 17. KEY WORDS | 18. DISTRIBUTION STATEMENT |
|---|---|
| General automata, loop data, real-time data, traffic management, congestion prediction, recurring congestion, non-recurring congestion, car following | No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22616 |

| 19. SECURITY CLASSIF. (of this report) | 20. SECURITY CLASSIF. (of this page) | 21. NO. OF PAGES | 22. PRICE |
|---|---|---|---|
| None | None | | |

## DISCLAIMER

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

## Table of Contents

## List of Figures



## List of Tables

## EXECUTIVE SUMMARY

The overall goal of this project is to create a method for predicting traffic congestion on freeway corridors. When implemented, it will provide a traffic service like that of "pin point Doppler" weather radar that can predict growing or dissipating congestion. Preliminary versions of the model used real-time loop data to successfully reproduce traffic behavior under moderately congested conditions. In order to improve the model for heavily congested conditions, the model had to be accurately calibrated. The process of calibrating the model revealed that inductance loop errors were preventing accurate results. An algorithm to correct the data from improperly functioning loops was developed and published at the 2003 Transportation Research Board Annual Meeting. The corrected loop data from the TDAD (Traffic Data Acquisition and Distribution) data mine are now being used to calibrate the model. The algorithm created in support of this effort can be used with malfunctioning loops to improve the freeway management system performance monitoring effort.

The first part of this report details the efforts and algorithms developed to correct the freeway loop data so that they are a viable input to any model. This was the primary accomplishment of this phase of the project.

The first section presents an off-line method for detecting and correcting errors in inductance loop traffic data. The method compares volume counts at a station level and balances the counts by creating a correction factor that represents the number of cars over/under counted by the station over a period of time. This correction factor is a property of the station itself and is detectable in a number of data sets.

After a correction to the original data was applied, the new time series presented a more consistent estimate of traffic counts on the freeway. The application of this method is useful for situations in which a self-consistent estimate of traffic volumes is necessary, such as in using the volume as input for a modeling effort.

The second part of this report details the efforts to date to build a robust model. It describes past work and sets a context for creating a viable model. An optimal control approach is described for the prediction model. This model has not yet been realized and is the work product for the next phase of the modeling effort.

When completely implemented, the model will be able to predict recurring congestion, non-recurring congestion once an incident has been identified and located, and dissipation of congestion. It will also be able to estimate the effects of lane and road closures on freeway congestion.

When completely implemented, the model will predict 20-30 minutes into the future, allowing for improvements in traffic control strategies such as ramp metering. It will also provide support for lane closure decisions, as well as predictions useful for traveler information akin to weather report predictions.

## 1. INTRODUCTION

The overall goal of this project is to create a method for predicting traffic congestion on freeway corridors. When implemented, it will provide a traffic service like that of "pin point Doppler" weather radar that can predict growing or dissipating congestion. Preliminary versions of the model used real-time loop data to successfully reproduce traffic behavior under moderately congested conditions. In order to improve the model for heavily congested conditions, the model had to be accurately calibrated. The process of calibrating the model revealed that inductance loop errors were preventing accurate results. An algorithm to correct the data from improperly functioning loops was developed and published at the 2003 Transportation Research Board Annual Meeting. The corrected loop data from the TDAD (Traffic Data Acquisition and Distribution) data mine are now being used to calibrate the model. The algorithm created in support of this effort can be used with malfunctioning loops to improve the freeway management system performance monitoring effort.

The first part of this report details the efforts and algorithms developed to correct the freeway loop data so that they are a viable input to any model. This was the primary accomplishment of this phase of the project.

The second part of this report details the efforts to date to build a robust model. It describes past work and sets a context for creating a viable model. An optimal control approach is described for the prediction model. This model has not yet been realized and is the work product for the next phase of the modeling effort.

When completely implemented, the model will be able to predict recurring congestion, non-recurring congestion once an incident has been identified and located, and dissipation of congestion. It will also be able to estimate the effects of lane and road closures on freeway congestion.

When completely implemented, the model will predict 20-30 minutes into the future, allowing for improvements in traffic control strategies such as ramp metering. It will also provide support for lane closure decisions, as well as predictions useful for traveler information akin to weather report predictions.

## 2. ALGORITHM FOR THE DETECTION AND CORRECTION OF ERRORS

### *2.1 Introduction*

This section presents an algorithm for correcting errors in archived loop data from freeway traffic management system that are the result of poorly calibrated sensors. These errors pose a significant difficulty when trying to use archived data in off-line analysis because the calibration errors are difficult to detect with traditional methods. In the work presented here, consistency of vehicle counts is used to judge the validity of the data; if vehicle counts are balanced, the data are valid; if vehicle counts are not balanced, the data are not valid. The method also can determine a correction factor. This correction factor is used to create a time series that can be combined with the original data to adjust the volume to create a consistent data set. To illustrate the methodology, an example case is presented that details the process of identifying a pair of reference stations that are properly calibrated. After the reference stations have been identified, a poorly calibrated station is identified, and the data from this station are corrected. Finally, the result of the correction process is discussed.

Inductance loop sensors are the primary method for freeway surveillance in use today. Freeway surveillance systems are used for both real-time traffic management applications and for off-line traffic planning applications. Traffic agencies have begun to archive the sensor data in an effort to better understand traffic phenomena over longer periods of time.

A basic problem with archived data is that invalid data affect the conclusions that are drawn from them [1]. Off-line processing of archived data is a necessary step in any analysis that uses historical records of traffic data. While a number of authors have proposed methods for determining the validity of sensor data [1, 2, 3], these methods are only able to answer the Boolean question of whether the data are good or bad. A significant problem in the use of loop inductance sensor data is the detection and recovery of errors that are the result of poorly calibrated sensors. Such sensors generate data that fall within the hard limits used for error detection but do not accurately represent the road conditions, and as an important segment of the sensor population, they need to be explicitly handled.

The problem of adjusting the output of poorly calibrated sensors has previously been discussed in situations involving dual loop detectors. In [4], duel loop “on-time” is

compared to identify failed detectors, and in [5], an estimate of the vehicle length from paired detectors is used to idenify a set of detector errors.

However, single loop detectors provide the most coverage of the freeway network and are also prone to calibration errors. Recent work has also used single loops to correct errors in freeway data [6, 7]. However, these methods are designed to work with a single loop sensor, while the method presented here works with pairs of single loop sensors. Poor calibration can result in a loop sensor that chronically under-counts the traffic passing over it or that chronically over-counts by counting traffic in neighboring lanes [8].

### *2.2 Problem Statement*

During the analysis of archival loop sensor data, it is desirable to post-process the data to detect and recover errors that were not detected when the data were originally collected. Specifically, the goal of the work presented here is to recover errors that are the result of improperly calibrated loops that under/over-count traffic volumes. During the development of a simulation model using traffic data [9], the lack of consistency due to over/under-counting at various stations posed a significant problem in the development process. This report presents a method to detect stations (groups of loop sensors that span all the lanes of the roadway) that are under/over-counting cars and then a companion method for recovering the data, essentially balancing the vehicle count at each station. This method is suitable for creating a valid dataset where vehicle counts are consistent over a series of stations.

The algorithm consists of three basic steps. The first step is to find a reference station that is correctly calibrated; all loops are properly counting cars. Next, stations adjacent to the reference station are compared to the reference station. The comparison process determines the validity of the adjacent station. Either the station is operating correctly, at which point it can be used as a reference station for another comparison, or errors are present. If errors are present, the bias in the station is calculated, and a correction time series is applied to the station data. After correction, the station is then used as the reference station for the next adjacent station calibration.

### 2.3 Algorithm Description

#### 2.3.1 Time Series Alignment

To compare the target station with the reference station, the time lag between the two stations must be determined. This time lag is required to account for the time a vehicle takes to travel between the two stations. Unlike the dual loop sensor configurations, where the loop sensors are typically less than 15 meters apart, the distance between two single loop sensors is much larger. In this experiment, the stations are .72km to .90km apart. Figure 1 shows the basic configuration of the sensors used in the experiment. The data used in the experiment were collected from SR 167 in Washington State over a series of Wednesdays in 2000 and 2001 when there were no incidents. The time lag ($\tau$) is calculated by finding the value that results in the largest correlation coefficient when the output of the two stations is compared.

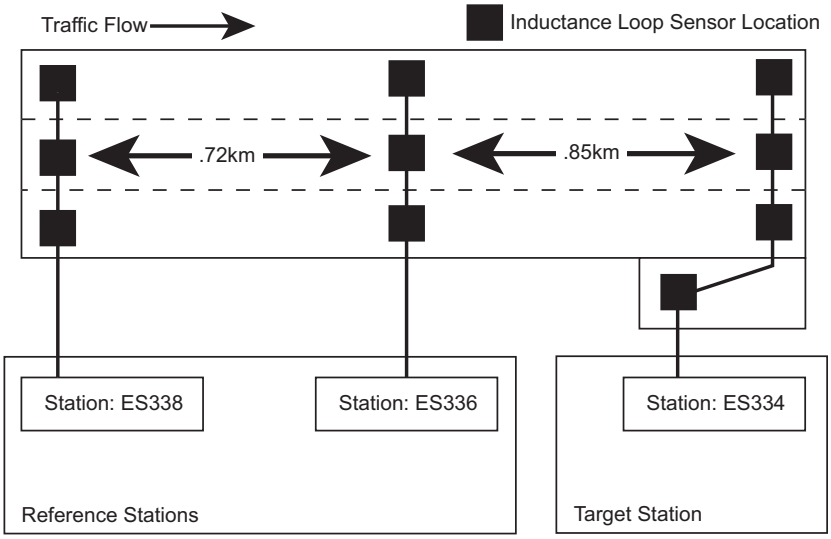


**Figure 1. Experiment setup.**

The correlation coefficient between the two stations is calculated in three steps. The first step is to aggregate the loop data. To do this, the total volume at each time step across all lanes for both the target and reference stations is computed using

$$v_j(k) = \sum_{i=1}^{N} v_{ij}(k) \quad k = 1...K \ , \tag{1}$$

where $K$ is the total number of 20-second intervals in a 24-hour period, $v_j$ is the volume at station $j$, $v_{ij}$ is the volume for lane $i$ at station $j$ , and $N$ is the total number of lanes at station $j$.

In addition, the inductance loops measure vehicle presence every 60th of a second, and the traffic management system aggregates these measurements over 20 seconds to create a "scan count." The mean of these scan counts is

$$s_j(k) = \frac{1}{N}\sum_{i=1}^{N} s_{ij}(k) \quad k = 1...K \ , \tag{2}$$

where $K$ is the total number of time steps, $s_j$ is the scan count at station $j$, $s_{ij}$ is the scan count at lane $i$ at station $j$ , and $N$ is the total number of lanes at station $j$.

This results in four separate time series: the upstream total volume ($v_u$), the upstream mean scan count ($s_u$), the downstream total volume ($v_d$), and the downstream mean scan count ($s_d$).

The second step is to calculate the correlation coefficient used for each discrete lag value (e.g., 0, 1, 2, 3). In this experiment, 20-second traffic counts are used, so each lag value represents 20 seconds. This calculation results in a 2-by-2 matrix of correlation coefficients. The values of the matrix elements are shown in Table 1. The final correlation value is calculated by summing all elements in the matrix and normalizing by dividing by four. The plot in the top of Figure 2 shows the correlation at various discrete time lags for a pair of loop sensors that are .72km apart.

**Table 1. Correlation matrix values.**

| | |
|---|---|
| $\mathrm{cor}(v_u(k), v_d k+\tau))$ | $\mathrm{cor}(s_u(k), v_d k+\tau))$ |
| $\mathrm{cor}(v_u(k), s_d k+\tau))$ | $\mathrm{cor}(s_u(k), s_d k+\tau))$ |

The final step is to estimate the actual lag coefficient by fitting a second order polynomial near the maximum of the correlation values and using the analytical derivative

to estimate the time lag . A plot of this polynomial curve is shown at the bottom of Figure 2. In this case, the maximum corresponded to a lag value of 1.48 (or approximately 30 seconds).

### *2.4 Finding a Reference Station*

The next step in the error correction process is to find a station that is properly calibrated. Since the error correction algorithm assumes that one station is outputting the correct values, a station must be designated as the 'reference' station before the algorithm can be applied to other stations. To accomplish this, the algorithm searches for two adjacent stations that are calibrated correctly.

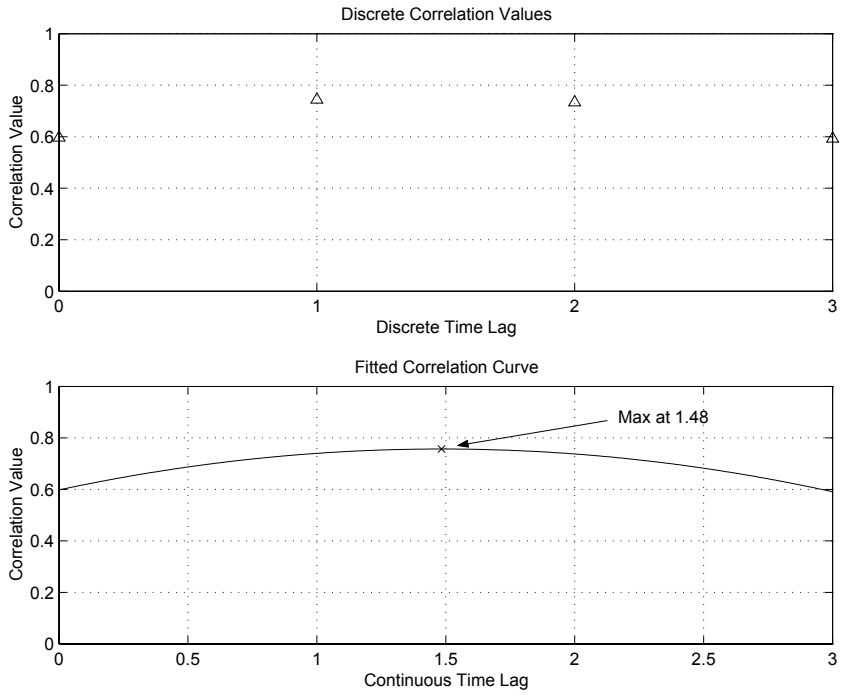


**Figure 2. Time lag calculation.**

The basic assumption used by this search is that it is possible to reconcile the vehicle counts in the station data so that the volumes of vehicles entering and exiting the freeway are consistent with the number of vehicles remaining on the freeway. The same number of cars counted by the upstream sensor should be counted by the downstream sensor at some future point in time.

The first step in the process of choosing a reference station is to determine which stations are potential candidates. The reference station candidates are limited to pairs of stations that have no on- or off-ramps between the mainline sensor locations.

Given a set of candidates, the next step is to evaluate the historical data quality of the sensors using the error algorithms built into the sensors. The number of samples recorded over a 24-hour period is divided by the total number of 20-second intervals in that period to verify that the loop detector is functioning continuously, and this fraction is used as a quantitative measure of data completeness. The ratio of the number of samples identified as valid, by the error algorithms built into the sensors, to the total number of observed samples is used as quantitative measure of the data validity. The multiple of these two measures is used as a likelihood measure, and sensors with less than a 95 percent likelihood are removed from consideration as a reference station.

Finally, the station pairs that meet the first two criteria are evaluated to determine the cumulative deviation in volume count over a 24-hour period for 30 different days. Using the earlier calculation of the lag between the two sensors, the total volume of cars counted by the upstream station is compared to the total volume of cars counted by the downstream station. The total volume of cars at a station *j* ($T_j$) is calculated as

$$T_j(k) = \sum_{l=1}^{k} v_{lj} \quad k = 1 \ldots K \tag{3}$$

where vj(t) is the 20-second volume count at station j at time t and K is the total number of time steps in the data sample.

To properly represent the continuous time lag using a time series with a discrete index, the actual data value is approximated by interpolation using the points surrounding the continuous lag. For example, to approximate k+1.5, the points k+1 and k+2 are used. In addition, these points are weighted using two parameters (α, β) that describe the contribution of each point to the total. β is set to the non-integer part of the time lag and α is set to 1-β. The time series describing the difference in total volume from time step zero to time step k, *D*(k), is computed as

$$D(k) = T_u(k) - \alpha T_d(k + \tau_l) - \beta T_d(k + \tau_u) \quad k = 1 \ldots K - \tau_u \tag{4}$$

where Tu is the total volume upstream calculated with Equation (3), Td is the total volume downstream calculated with Equation (3), α, β are interpolation coefficients to smooth τ, τl is the largest integer less than or equal to τ , and τu is the smallest integer greater than or equal to τ.

The time series of the difference between the total volume counted at each station can have three basic trends. The difference can increase over time – implying that vehicles are counted upstream but not counted downstream. The difference can decrease over time – implying that vehicles are not being counted upstream but are being counted downstream. Finally, the difference can vary around zero with a zero or near-zero mean. In this situation, both stations are counting cars at the same rate, and the errors can be attributed to other unknown sources of noise in the system. Figure 3 shows the three different types of time series.

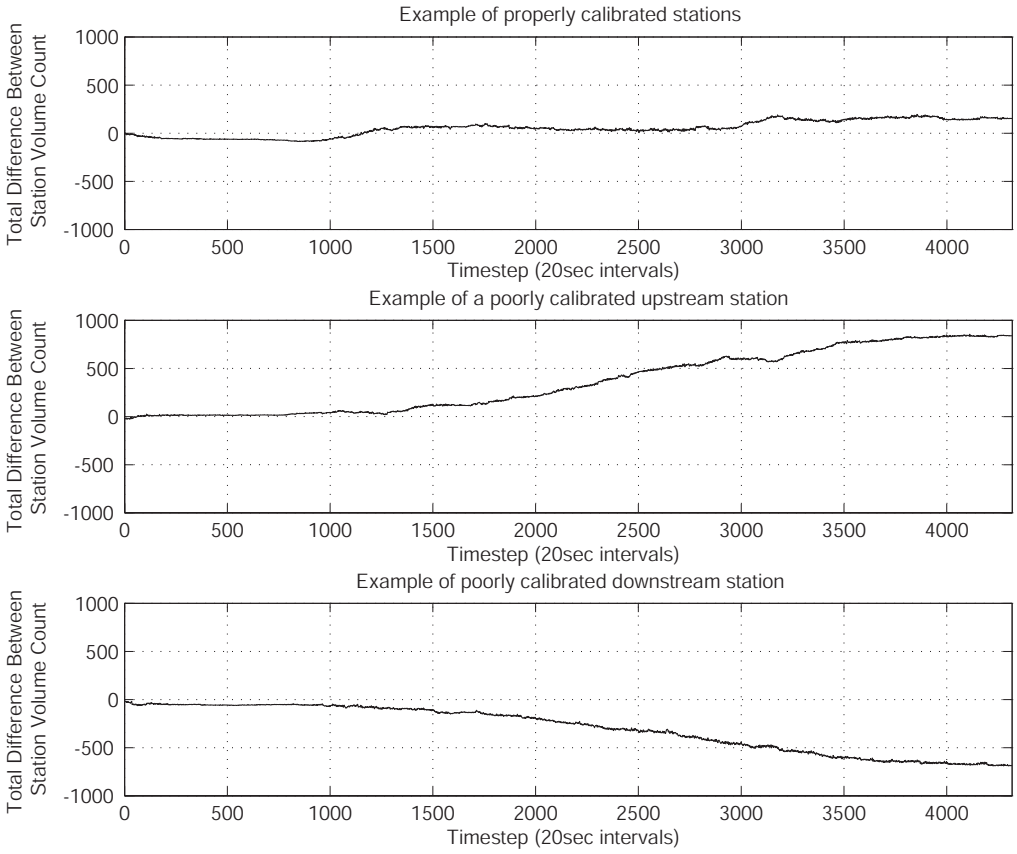


**Figure 3. Basic sensor errors due to calibration issues.**

Next, the difference between the 20-second data for the two stations (*d(k)*) is calculated as

$$d(k) = v_u(k) - \alpha v_d(k + \tau_l) - \beta v_d(k + \tau_u) \quad k = 1 \ldots K - \tau_u \,, \tag{5}$$

where *vu* is the 20-second volume upstream calculated as Equation (1), *vd* is the 20-second volume downstream calculated with Equation (1), and α, β, τl, *τu* are defined as before.

When the difference in the volume counts between two stations (adjusted for the time lag) has a zero mean with a small variance over the course of an entire day, those stations can be used as reference stations for comparison with other stations immediately upstream/ downstream of the pair. In general, the candidate pair with the smallest mean and variance is selected as the reference pair. Figure 4 shows the histogram of the errors between two reference stations. These two stations are designated as the initial reference stations because the mean difference in volume counts is near zero. Having designated the reference station, the next step is to iteratively traverse up or down the freeway and correct poorly calibrated stations as they are detected.

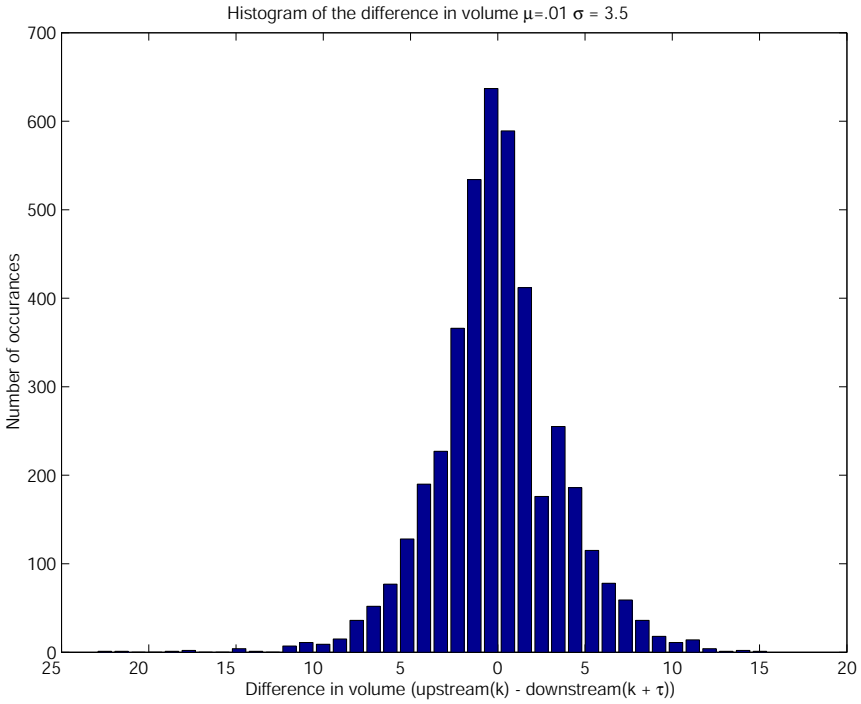


**Figure 4. Histogram of the error in volume counts between two reference stations.**

### *2.5 Detecting Stations With a Poorly Calibrated Loop Sensor*

Using the reference station previously identified, the total volume comparison takes on a new meaning. If the comparison results in a difference that increases over time (target station is downstream) or decreases over time (target station is upstream), it follows that the target station is poorly calibrated. While there could be other more serious reasons why the station volume counts do not match, a completely broken sensor for example, this work focuses only on the sensors that are reporting data flagged as normal, and it is assumed that data are being reported by all loops that make up a station.

Much like the process used to determine the reference station, Equation (5) is used to calculate the difference between the two stations adjusted for lag. Figure 5 shows the histogram of the difference between a reference station and a station with a poorly calibrated loop. Given a total volume of around 60,000 vehicles per day, the mean represents a loss of around 1000 vehicles per day or 1.67 percent. By comparison, the loss between two reference stations is less than 100 cars (.16 percent). Not only are such errors present in a single day of data, this bias exists in multiple days over a long time period. Figure 6 is a graph of the total number of uncounted cars between two adjacent stations for 30 different days collected over a period of two years. As shown by this figure, this particular station consistently does not count 1000 vehicles each day.

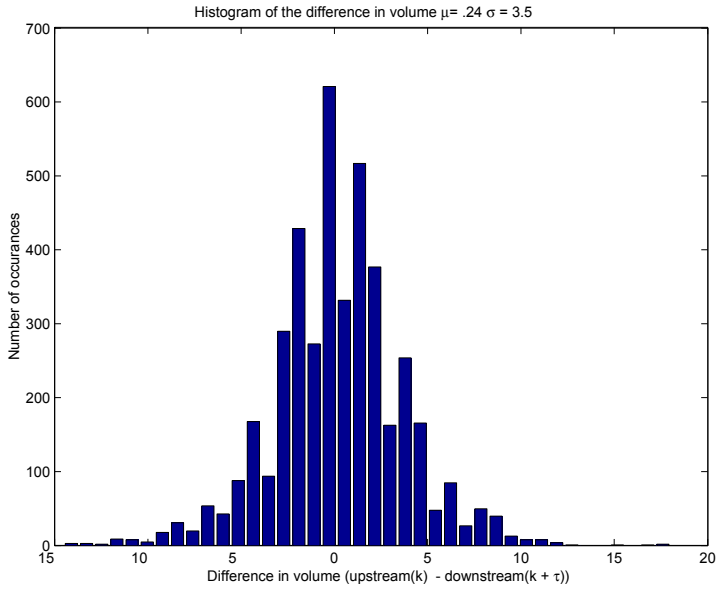


**Figure 5. Histogram of the error in volume counts between a reference station and a station with a poorly calibrated loop.**

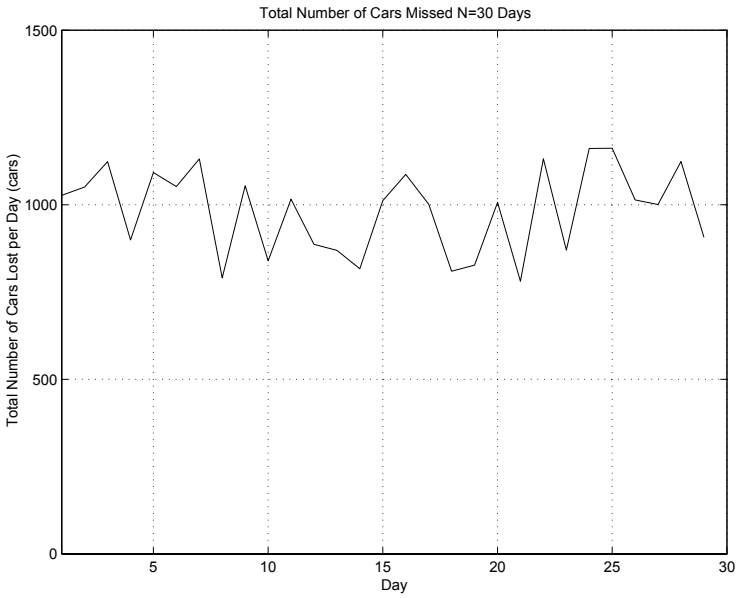


**Figure 6. Total number of uncounted cars over 30 different days.**

### *2.6 Correcting the Station Data*

Given the basic assumption that the over-counting or under-counting of cars is proportionate to the total number of cars that pass through the station, the first step in the correction process is to use Equation (3) to create a time series that describes the total number of cars that are counted by the station ($T_v$). When the total volume is calculated, on- and off-ramps are added in as appropriate. For example, if an off-ramp is immediately upstream of the downstream sensors, the volume count of the offramp is included in the downstream total. Conversely, if an on-ramp is immediately downstream of the upstream sensors, the volume count of the on-ramp is included in the upstream total. This procedure differs from "ramp balancing," as described in the FHWA *Traffic Monitoring Guide* [10], where the difference is that the adjustment is made to the freeway loop volume rather than the ramp loop volume.

The next step is to compute the ratio between the total number of cars miscounted in a day and the number of cars miscounted every 20 seconds. Since statistically only a fraction of a car is miscounted every 20 seconds, this step provides a correction factor that can be applied to the time series to determine the proper time step to add or subtract exactly one car. The number of cars that must be counted by the station before the algorithm adds or subtracts an additional car, $f$, is

$$f = \frac{T_v(K)}{\mu K} \ , \tag{6}$$

where μ is the expected value of E{d(k)} (shown in Figure 5). Using this factor, $T_v$ is converted to the total number of cars that were miscounted ($T_\varepsilon$) as

$$T_\varepsilon(k) = fix\left(\frac{T_v(k)}{f}\right) \quad k = 1 \ldots K - \tau_u \ , \tag{7}$$

where fix $(\bullet)$ is a numerical operation that rounds the number $(\bullet)$ to the nearest integer towards zero.

The process of creating $T_\varepsilon$ results in an integer valued function. Since volume data are quantized into integer values of whole cars, this algorithm preserves the integer values. Finally, the time series describing the total number of cars that were miscounted is converted into a time series describing the instantaneous number of cars ($I_\varepsilon$) that are miscounted by taking the difference of $T_\varepsilon$ as

$$I_\varepsilon(k) = T_\varepsilon(k+1) - T_\varepsilon(k) \quad k = 1 \ldots K - 1 \, . \tag{8}$$

While $I_\varepsilon$ is a time series of length $K$, its internal structure is a binary sequence of 0s and 1s. At a majority of the time steps $k$, $I_\varepsilon$ is zero, but in a small percentage of instances it adds cars to the data stream. The $I_\varepsilon$ sequence removes the bias from the poorly calibrated sensor, resulting in a time series that maintains consistency throughout the day. Using the time series $I_\varepsilon$, the original station volume counts can be corrected by adding the $I_\varepsilon$ to the original time series, resulting in a new corrected time series.

### *2.7 Verifying the Station Data*

Now that a method for determining the correction factor and correcting the data for a one-day data set is available, the next step is to determine a suitable correction factor for a sequence of days. As shown before in Figure 6, the effects of the poorly calibrated sensor are visible over a number of days. In the same way, Figure 7 shows the time series of calibration factors for these days. A general correction factor can be computed by using the mean of the correction factor time series that applies to a number of days.

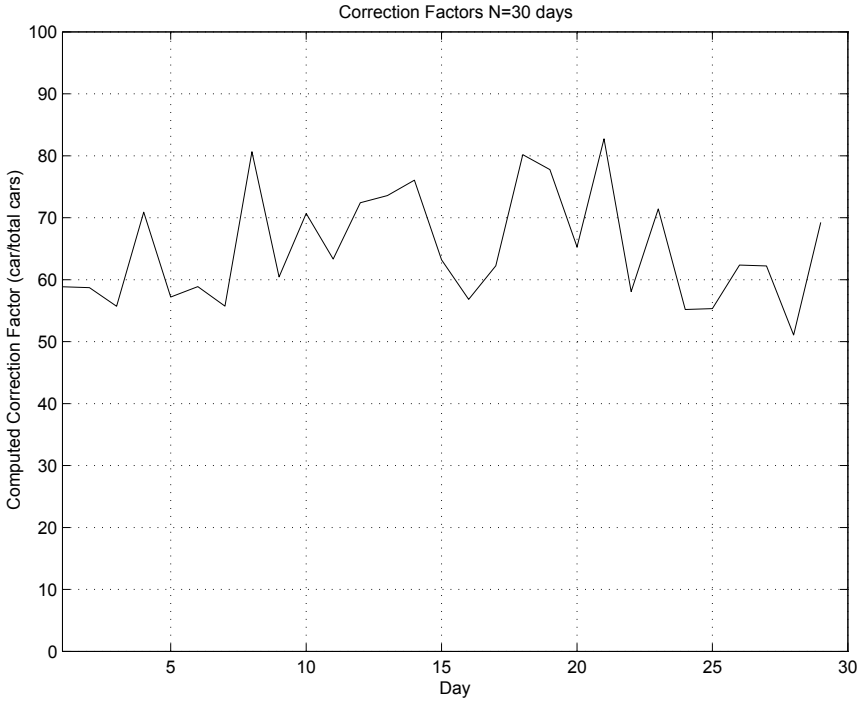


**Figure 7. Correction factor for 30 different days.**

For a particular day, by using equations (3), (7), (8) and the general correction factor, $I_\varepsilon$ can be calculated. The corrected station data,

$$v_c(k) = v_t(k) + I_\varepsilon(k) \quad k = 1 \ldots K \ , \tag{9}$$

can be compared to the reference station by using equations (3) and (4). The result of this comparison is shown in Figure 8. In addition, Figure 9 shows the comparison of the corrected cumulative sum of uncounted vehicles with the uncorrected cumulative sum. In this case, less than 0.5 percent of the total vehicle count is lost. From this, we conclude that the adjustment process successfully restores the consistency of vehicle counts to the station with a poorly calibrated loop because the time series comparison of the reference station to the corrected station shows the removal of the trend. After the station volumes have been properly adjusted, the adjusted station can be used to correct stations that are farther downstream. In this way, a whole freeway network can adjusted iteratively without either using extremely large lag factors for a single reference station or incorporating large numbers of loops into the calibration process.

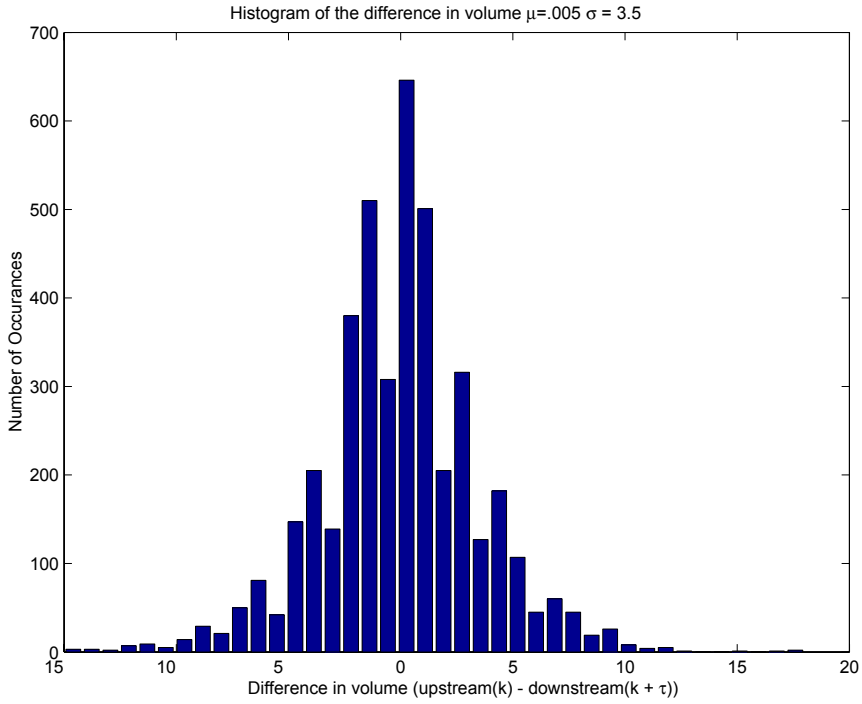


**Figure 8. Histogram of a reference station compared to a corrected station.**

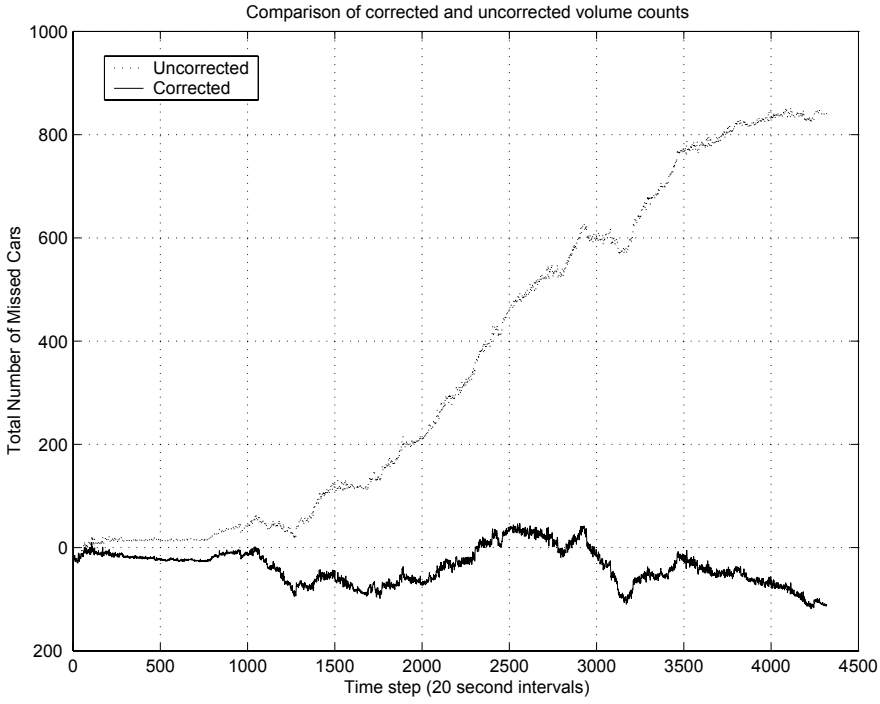


**Figure 9. Time series comparison of a corrected station to the uncorrected time series.**

### *2.8 Findings*

The algorithm consists of three basic steps. The first step is to find a reference station that is correctly calibrated; all loops are properly counting cars. Next, stations adjacent to the reference station are compared to the reference station. The comparison process determines the validity of the adjacent station. Either the station is operating correctly, at which point it can be used as a reference station for another comparison, or errors are present. If errors are present, the bias in the station is calculated, and a correction time series is applied to the station data. After correction, the station is then used as the reference station for the next adjacent station calibration.

### *2.9 Conclusion*

This section presented an off-line method for detecting and correcting errors in inductance loop traffic data. The method compares volume counts at a station level and balances the counts by creating a correction factor that represents the number of cars over/under counted by the station over a period of time. This correction factor is a property of the station itself and is detectable in a number of data sets.

After a correction to the original data was applied, the new time series presented a more consistent estimate of traffic counts on the freeway. The application of this method is useful for situations in which a self-consistent estimate of traffic volumes is necessary, such as in using the volume as input for a modeling effort.

# 3. A HIERARCHICAL RAMP ACCESS CONTROL SYSTEM

## *3.1 Introduction*

This section offers a framework for controlling the ramp access system of a freeway corridor. The framework is original in that traffic dynamics are modeled with a microscopic model of traffic flow. A microscopic model is used because it models the traffic more realistically than standard macroscopic models and provides a significant improvement in the ability to represent unobservable quantities. To use the microscopic model in the controller, a control system is proposed that incorporates three different data sources: real-time data, historical data, and simulated data. In addition, the proposed controller uses a hierarchical approach to accomplish both local and system-wide control of the freeway.

While other ramp access controllers have employed a hierarchical or multi-layer approach, our controller is different in that a microscopic simulator is applied in the calculation of the steady-state set points used by the local controllers. The controller incorporates an optimization algorithm that utilizes the microscopic simulator to create a desired steady-state control input that accounts for variable traffic conditions across the entire network. In addition to the system-level controller, an $H_\infty$ controller is presented that carries out the high-level control strategy at the local level. The $H_\infty$ controller is an optimal control concept developed in the 1980's, and the term $H_\infty$ is derived from the cost metric that is minimized to obtain the control input. This approach provides a guarantee that the control input will provide the "best" result possible for the system being controlled. While $H_\infty$ control structures have been previously used for traffic control, the $H_\infty$ controller is significantly improved by the ability to use the set points provided by the high-level controller.

Throughout history urbanization has spurred development of efficient, high-volume transportation networks. By the 21$^{st}$ century, the limited ability of urban centers to increase the physical size of their transportation networks has necessitated research focused on improving the performance of existing traffic networks by creating intelligent transportation systems (ITS). To this end, many transportation authorities have invested large amounts of money to create regional centers to monitor and control the transportation network.

An area of particular interest involves improving the performance of the highest capacity links in the network—the highway corridors. Wattleworth was one of the first researchers to suggest that congestion on the freeway could be reduced by metering ramps to control the rate at which cars enter the freeway [11]. Soon after, Payne and Ikasen proposed a closed-loop controller that dynamically adjusted local metering rates [12]. While a local control scheme effectively reduces recurring congestion, its ability to reduce the effects of non-recurring congestion remains limited.

Payne also created an early system-wide controller by aggregating a number of local controllers together, but later work overshadowed his methods. Papageorgiou proposed a model of traffic flow to create a discrete-time optimal controller [13]. Developing a method for achieving the discrete time optimal solution for the controller has been a popular area of study with many researchers [14, 15, 16, 17, 18]. More recent developments in system-wide control include fuzzy logic controllers [19, 20, 21, 22, 23, 24, 25] and neural network controllers [26, 27, 28].

While a great deal of work has been done on models for the traffic ramp metering problem, continuum models that use a basic model of traffic flow derived from (or utilizing) the Lighthill and Whitham model [29] are still the basis for many of the most recent controllers [17, 18, 30]; however, this type of model is limited by the unobservability of key states of the roadway, as well as the presence of bad data. In this section, an original controller is introduced that utilizes a microscopic model to accomplish tasks that were previously performed with macroscopic models. In addition to providing basic functionality, the model provides new methods that significantly increase the ability of the model to estimate unobserved states and mitigate the problems caused by bad data.

The objective of this section is to describe an original and improved solution to the ramp access control problem. The proposed architecture assumes the presence of an existing freeway surveillance system that is centrally located and accessible. The proposed controller itself would be integrated with other ITS initiatives as part of a complete, dynamic traffic management system.

The design of the controller takes into account a number of key goals. The first consideration is that the controller must be able to manage congestion. Another

consideration is that the controller must be traffic responsive in that it detects and responds to changing traffic conditions. A new model of traffic flow is presented that takes advantage of existing data sources and is also able to estimate conditions when data are lacking for a simulation. This model is a significant improvement over existing models currently used for ramp access controllers.

## 3.2 Background

The development of the controller draws from two primary bodies of knowledge: general control theory literature describing large-scale complex systems and traffic control theory. The general control theory sets the context for evaluating and understanding how to best utilize the concepts from the transportation literature. The traffic control theory sets the baseline for understanding the properties of the system and existing solutions for this problem.

### 3.2.1 Control Theory

The definition of a large-scale system varies on the basis of the context in which the work is presented. In the realm of systems control, a system is large if it exceeds the capacity of a single control structure, while in the realm of model development, a system is large if its input-output behavior cannot be understood without partitioning the model into smaller modules [31]. While the control of large-scale systems dates back to the mid 1960s, a review of the literature reveals that large-scale system control techniques have been applied to a wide variety of systems: transportation, communication, power, and the economy [31, 32, 33].

Transportation systems are particularly suited for analysis and control with large-scale methods both because of the high level of complexity inherent to them and because of the loosely coupled nature of transportation networks. Furthermore, large-scale control methods have been applied with a good deal of success to the ramp access control problem [12, 32, 33, 34, 35, 36, 37]. Therefore, this section will utilize methods and techniques from the large-scale control theory to develop both the model and the controller.

A number of different methods have been proposed for dealing with large-scale systems. In general, the inability to control a large-scale system with a single control structure is handled by most large-scale control algorithms by using methods that involve simplifying

the system model in some way to allow for the application of traditional control algorithms. Common techniques to accomplish this include model reduction through aggregation of states, model reduction by decomposition, or the creation of a hierarchical or multi-layer control structure [31].

Hierarchical control schemes offer a number of benefits over other approaches. A primary benefit of a hierarchy is the ability to separate high-level abstract control from low-level physical control. Hierarchical control also has the ability to model complex hybrid systems that are composed of both discrete and continuous elements. Some current research in the area of large-system control focuses on controllers that use a hierarchical or multi-level approach to reduce a complex system into a number of less-complex sub-modules organized in a hierarchy [33, 39].

In the realm of transportation control, a number of authors have proposed hierarchical solutions for ramp access control [32, 36, 37, 38]. In a similar manner, our proposed controller will be developed in a two-level hierarchy: a system-wide controller responsible for optimizing the traffic flow on a global level, and a local-area controller responsible for carrying out the physical task of regulating the flow at a particular point.

Where other hierarchical systems have relied on macroscopic models as the basis for the high-level control and linearized local controllers for the low-level control, our proposed controller advances the concept in that it utilizes a microscopic model for traffic dynamics in the high-level controller and a non-linear $H_\infty$ controller for the local controller. While non-linear $H_\infty$ controllers have been proposed for local feedback control, use of an $H_\infty$ controller in a hierarchical control system designed for ramp access control is a new approach.

### *3.2.2 Traffic Control Theory*

In this section, the traffic theory used to develop a controller for a freeway corridor ramp metering system is described. The model of the system in Figure 10 shows both the physical freeway corridor to be controlled and the proposed controller model. The definitions of commonly used terms can be found in Appendix A. The controller uses a hierarchical control structure and completely relies on the microscopic model for state information. A hierarchical control structure has been applied to ramp metering control problems before and has been shown to be very effective [32, 40]. As shown in Figure 10, the controller consists

of three main blocks: the microscopic model, the system controller, and the local feedback controller.

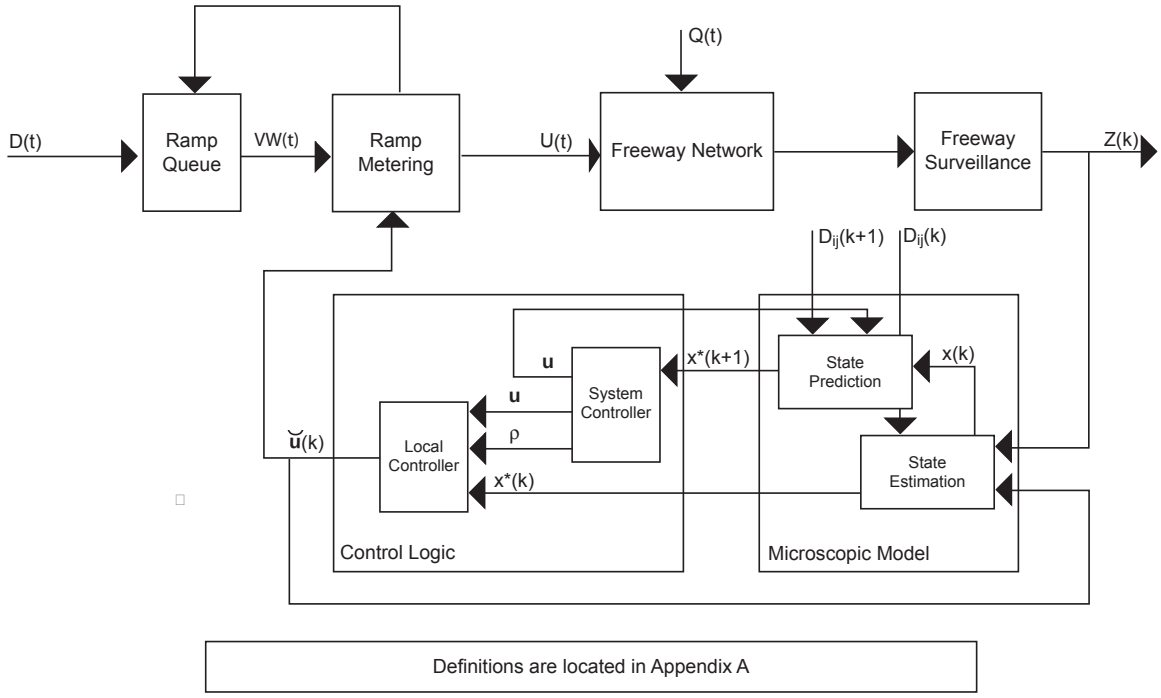


**Figure 10. Freeway corridor ramp metering system.**

The dynamic traffic flow modeling for the controller is provided by the microscopic model. The microscopic model uses origin-destination (O-D) rate information to determine when to create or remove representations of each individual vehicle on the freeway. The model 'moves' the vehicles by using an algorithmic description of driver behavior.

To completely describe the inputs to the microscopic model, three main data sources are needed. The first data source is the real-time data from the freeway surveillance system that measures the traffic rate on the on-ramps and off-ramps. The desired control input is needed to further describe the input information. As a supplement to the real-time surveillance and control data, the microscopic model also requires an estimate of the origin-destination (O-D) information for the cars entering the freeway. By using a prediction of the future inputs, the model is able to predict future states of the freeway.

The system controller is the first tier of the controller hierarchy. The function of the system controller is to minimize congestion on a global scale. This is accomplished by using

an optimization routine that tries to minimize the congestion in the model (as measured by the total delay in the system). The optimization scheme is bolstered by the fact that it can use the predictive output of the microscopic model. The system-wide controller sets the desired parameters (i.e., density, control input) that are used by the local feedback controllers.

The local feedback controller is a robust $H_\infty$ controller. By using the parameters passed to it by the system controller, the local feedback controller generates the ramp signaling time series with the provided state information. The local controller only utilizes the state information from nearby locations to determine the control input. The use of automatic control techniques to control flow on the freeway has been shown in the literature to be very effective at regulating traffic density in a local region [41, 42, 43]. While earlier works used the linear quadratic gaussian (LQG) formulation, recent work has also proposed using $H_\infty$ control for local ramp control [44].

Our controller improves on existing controllers in three main ways. First, by using a microscopic model for state estimation, bad data can be recovered in a manner that is not possible through normal time series methods. This is a significant improvement over use of a macroscopic model that must rely completely on sensor measurements. The microscopic model can give an estimate of sensor information even when there may be no data from the physical sensor.

Second, the freeway is modeled at a finer granularity with the microscopic model . Previous controllers discretize space according to the placement of the on- and off-ramps [13, 17, 18, 21]. This is a significant advantage in that macroscopic models are unable to provide accurate control information when bottlenecks or other discontinuities are located outside of the sensor region. The microscopic model allows the controller to get an estimate of control volumes at any place in the network and develop control strategies that takes the effect of these areas into account.

Third, the microscopic model is able to incorporate dynamic O-D information to predict the future state of the system. This provides a more accurate prediction than one that is based on historical data. O-D information a non-stationary [32]. Therefore, historical data cannot completely characterize its behavior.

## 3.3 Approach

### 3.3.1 Overview

While previous models used for traffic control have utilized macroscopic flow techniques, it is also possible to model traffic flow as a granular flow process by using microscopic techniques. While there have been many successful macroscopic models (e.g., [45, 46, 47, 48, 49]), the desire to represent the diversity of the 'realistic' contents of a road using a short time scale has spurred the development of a number of microscopic models in the last few years. Theoretically, every aspect of the environment, driver, and car can be modeled with a microscopic model, but in practice, most models are simplistic because of difficulties associated with having a large parameter space.

To set the context for the development of the model, the first step is to describe the physical properties of the freeway network, as well as the existing surveillance and control systems. Next, a microscopic model is proposed to replace the macroscopic model of traffic dynamics commonly used in traffic control problems to describe the observed physical properties. The proposed microscopic model represents the entire contents and models the dynamics of the physical freeway system in a realistic and accurate manner. To conclude this section, a method for determining the information necessary to properly start and operate the model is proposed.

After the functions of the microscopic traffic dynamics simulator are described, a method for incorporating the simulator into a ramp access control system is proposed. The proposed method breaks the control problem into two main parts: a high-level, system-wide controller and a low-level, local controller. Using the definition of the microscopic simulator created earlier, a logical description of the microscopic model is formulated that incorporates a simulator directly into the control architecture. Next, the optimization routine for use by the system-wide controller to coordinate the operation of the local-controllers is described. The proposed method is original in the fact that it uses a microscopic simulator of traffic flow to model the mainline traffic dynamics in the optimization routine.

Next, the structure of a local feedback controller to control the ramps in a local region is developed with $H_\infty$ control techniques. While previous work has developed local ramp controllers by using $H_\infty$ methods, our proposed controller is original in that it is designed to

work in coordination with a system-wide controller. This is significant in that system-wide control mitigates some of the problems in the way a local controller prioritizes incoming traffic. In addition, the controller utilizes the simulated data for control rather than real-time data sampled from the freeway itself.

### *3.3.2 Freeway Traffic Flow Surveillance and Control Systems*

Mainline traffic flow is a highly non-linear, stochastic process that is affected by many completely unobservable variables. While models of traffic flow based on concepts of fluid dynamics can model traffic flow on the macroscopic level, models of traffic flow that operate on the microscopic level with concepts of individual driver behavior give a better representation of the actual state of the traffic network. In addition, the current freeway surveillance system, while adequate for some tasks, often does not deliver a complete or accurate picture of freeway conditions. Current solutions to freeway observation have two main problems: 1) the sensors provide a very coarse discretization of the highway in space and 2) broken sensors can lead to unrecoverable errors. We first provide an overview of the existing freeway surveillance system and its limitations and then propose a microscopic model for estimating the mainline traffic dynamics.

The primary sensors used today for freeway corridor surveillance are inductive loop detectors. Physically, inductive loop detectors are coils of wire embedded just below the surface of the highway at regular intervals. The position of the sensors is predetermined by the regional traffic authority and are typically placed near points of interest such as ramps or interchanges. At each geographic location a cabinet is dedicated to collecting the sensor information, which is usually in the form of occupancy and volume counts per lane. Some freeway locations also have paired loop detectors. By grouping two loop detectors a known distance apart, it is possible to calculate the average speed of cars on the roadway. As part of this calculation, the mean length of the vehicles that passed over the detector during one sample period is also generated. In general, current freeway surveillance systems return volume and occupancy counts for sections of the freeway network and speed and length information for a small subset of these.

Inductive loop detectors are often prone to erroneous or noisy data. The effect of these errors can range from a complete loss of information for that region to an increase in signal

noise. The sheer volume of sensors required to monitor the freeway would suggest that at any particular time, some number of sensors will be broken. This becomes an issue in the development of the controller because some of the errors (e.g., a sensor that is stuck in the "on" position) cannot be recovered in real time or by using time series methods.

While the development of most current controllers assumes the presence of valid data, few controllers include a strategy to mitigate the problems caused by broken sensors. Although some people have proposed mitigation solutions for dealing with these situations [22], the presence of broken sensors generally has a negative effect on the performance of the controller.

The approach to calculating mainline traffic dynamics is an original one. The proposed method is to use a microscopic simulation to model the mainline freeway dynamics and to run the controller using information generated by the model. The proposed microscopic model improves on existing macroscopic models both in its ability to match the empirical data collected from the freeway and its ability to function in situations where the real data are bad.

### *3.3.3 Microscopic Model of Traffic Flow*

Unlike macroscopic models that are derived from a few basic principles (e.g., conservation of cars), microscopic models can be constructed from a number of different methods. On a basic level, microscopic models require three main components: a framework for describing the freeway network, a model of how entities behave, and a method of loading entities into the model. Each of these components is important to the overall success of the model, but the framework is unique in that it sets the constraints for the other components in terms of what is possible and what is not possible.

The proposed microscopic model uses a general automaton (GA) framework. Automaton networks have been used in many different scientific disciplines to model continuum dynamics on a microscopic level. A general automaton comprises a set of states, the interconnections between them, and a rule set that specifies the progression of states. Automaton frameworks are beneficial in that they reduce model complexity while maintaining accuracy. The GA framework for traffic flow is a generalization of the cellular automaton (CA) traffic flow model first proposed by Nagel and Scheckenberg [51]. While

the CA framework is often used because it is suitable for a parallel processing platform, the symmetry restrictions of the CA framework limit its ability to precisely model traffic flow. The GA framework frees the model from the symmetry restrictions, allowing it to provide a much more precise model of traffic at the cost of the ability to use a parallel processing platform.

The first step in the approach to developing a microscopic model is to create a set of cells and interconnections that model the topology of the roadway. To do this, the physical geography of the road is partitioned into 4-foot-long regions (with a region width equal to the lane width), and these regions are defined as "cells." In addition, interconnections are defined such that the cells are linked together. This process is done in such a manner that the direction of traffic flow and the number of lanes on the freeway are represented appropriately. During this process, the cells are sequentially numbered from the left lane to right lane, from the entrance of the model down to the exit of the model. The index of a cell allows it to be referenced by physical locations on the freeway.

Next, each cell in the GA structure is assigned a state that represents whether or not it is occupied. If the cell is unoccupied, then the state of the cell is simply zero. However, if the cell is occupied, then the state of the cell is the percentage of the cell that is occupied by a vehicle (the model allows for partial occupancies) and an index describing the specific car that is occupying it. The size of the cell and the rule set are such that only one car can occupy a cell at any particular time step, even if it only partially occupies the cell.

Even though the cell itself should be considered a dimensionless quantity (it only has a 'state' comprising whether or not it is occupied by a particular car), the index of a cell can be mapped to a location or distance. The process of converting the index to a physical location varies on the basis of the actual freeway being modeled and the parameters of the model itself. As a simple example of this, in a three-lane highway that is 10 miles long, cell number 1500 corresponds to a 4-foot patch of roadway in the left-most lane 2,000 feet from the start of the model. From this representation, this model has the ability to accurately model the position and length information of all the vehicles on the roadway. Assuming that this definition models the roadway and the vehicles in it, the next step is the development of a method to update the state of each cell.

A typical microscopic behavior model considers the current vehicle dynamics, driver behavior characteristics, road geometry, and the surrounding vehicles (those within the driver's perception) while determining the next 'state' or course of action. Each driver/vehicle combination in the traffic may be described by a set of parameters such as position, actual speed, route choice, and driving style. While the simplest microscopic models can be described with analytic equations [50], as model complexity increases, it is difficult to precisely derive the equations necessary to describe microscopic traffic phenomena because of the complexity, randomness, and interdependency of the process. Because of this, computer simulations, rather than analytic models, are the primary means of implementing microscopic models (e.g., [51, 52, 53, 54, 55]). Figure 11 shows the flow diagram for the proposed method of updating the state of each cell. The method updates the model sequentially, beginning with the car closest to the exit of the model and moving upstream.

The process of evaluating the traffic conditions and determining an appropriate course of action (lane changing, braking, acceleration) is handled entirely by the 'car and driver model.' Car and driver models are typically composed of four main sub-models: free-flow driving models, lane changing models, merging models, and car-following models. A robust model incorporates the functionality of all four sub-models to determine the behavior of the drivers on the road. One of the key advantages of using a microscopic model is the ability to use a stochastic driver model to create a heterogeneous traffic flow. In this case, a distribution describes the desired speed of the current driver. The ability to model heterogeneous traffic flow enables the microscopic model to describe traffic phenomena that are not easily captured by macroscopic models.

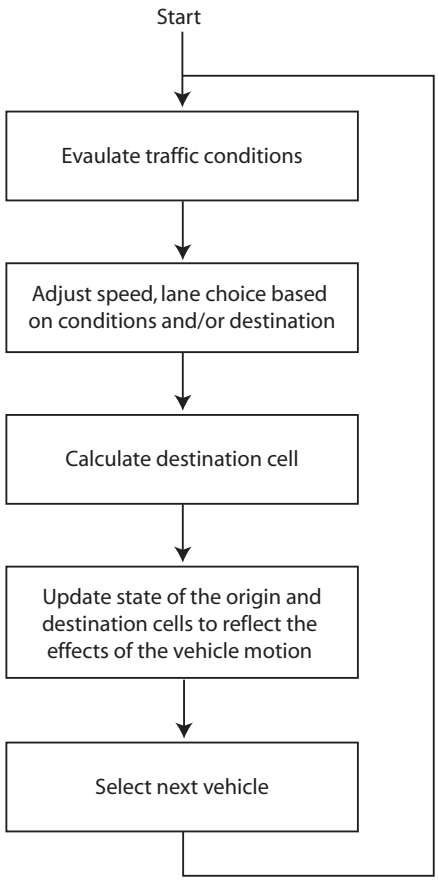


**Figure 11. Flow diagram for the cell update process.**

In the microscopic model, driver behavior is divided into two main parts. The first part determines the acceleration that the driver will use during the next time step. Figure 12 shows a flow diagram for this process. The sections of the process that are related to free-flow driving models and car following models are marked accordingly.

In the model, the driver evaluates the road conditions using the time headway, which is defined as

$$headway = \frac{\Delta X}{v_c} \ , \tag{10}$$

where $\Delta X$ is the distance between the current vehicle and the vehicle in front of it, and $v_c$ is the speed of the current vehicle.

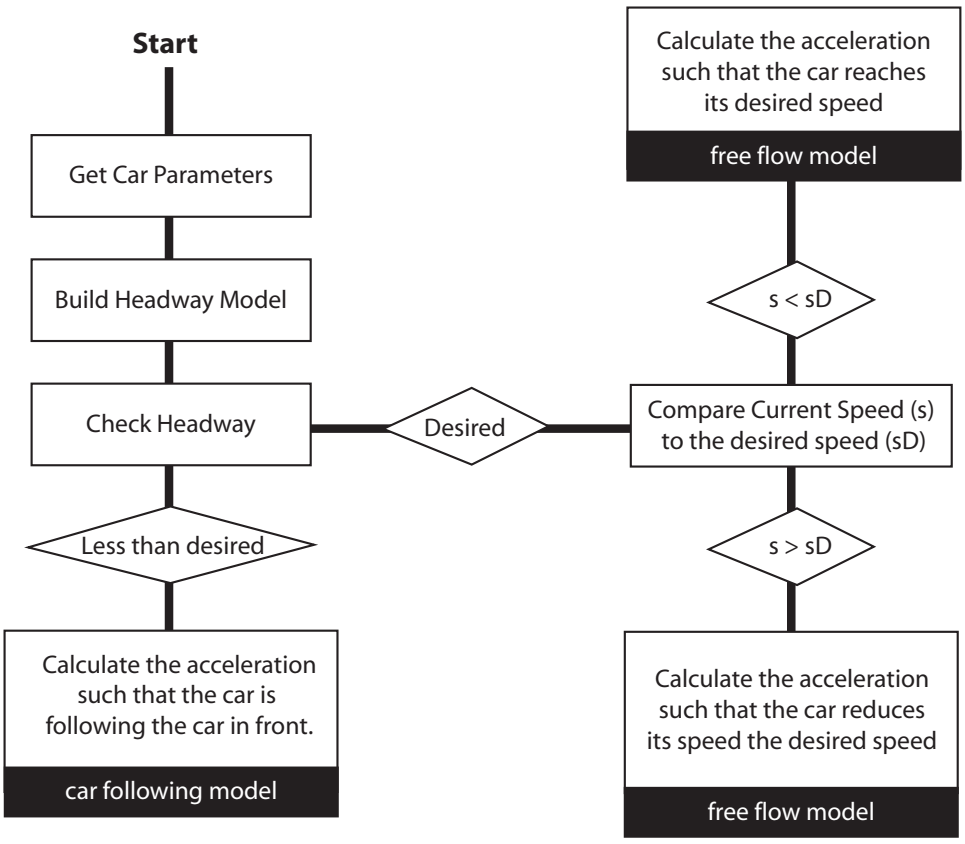


**Figure 12. Flow chart for the acceleration model.**

If the car in front is too close to the current vehicle, or the current vehicle begins to accelerate, the headway will shrink and the car will be forced to act on the changing conditions. By adjusting the driver's desired headway, the user is able to change the way that vehicles interact with each other in the model.

The second part of the model is the lane changing model shown in Figure 13. The lane changing model is responsible for two important driver behaviors. The first behavior is how the driver responds when the driving conditions are not desired, and the second behavior is how the driver navigates the freeway.

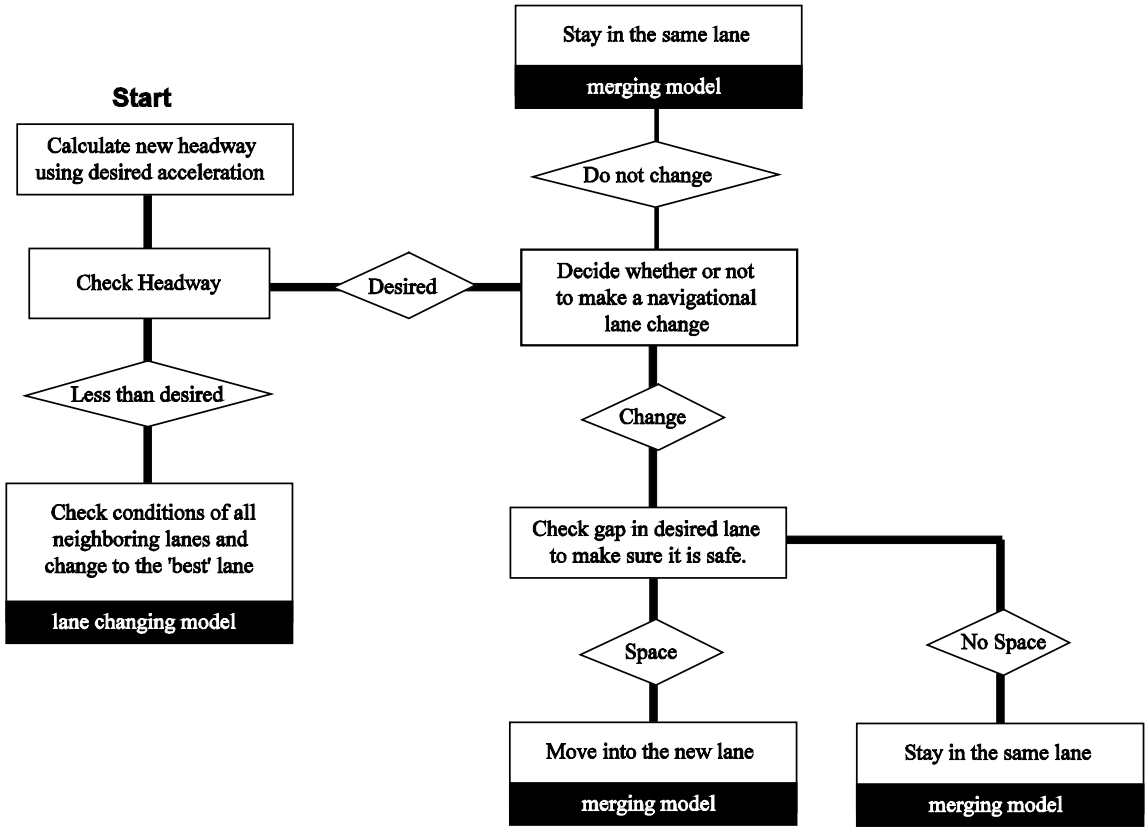


**Figure 13. Flow chart for the lane changing model.**

As shown by Figure 13, the headway is once again used to evaluate the conditions on the roadway. If the headway is less than desired, then the driver attempts to change lanes to improve the driving conditions. If the headway is okay, then the driver continues in the same lane until he/she has to make a navigational lane change. The navigational lane change can occur because of many conditions: the end of the current lane, the desire to exit the freeway, or an obstacle in the road to be avoided. The combination of the acceleration model and the lane changing model completely describe the behavior of the driver. , The acceleration and next lane choice for the current vehicle are calculated with these two models together, and then the vehicle is sent down the road.

The destination cell is calculated by using the standard laws of physics. The change in speed (calculated from the desired acceleration) is added to the current speed, and the location of the destination cell is calculated via the following formula:

$$C_1 = T\left(v_k + Ta_k\right) + C_o \ , \tag{11}$$

where $T$ is the time step, $v_k$ is the current speed of the current vehicle, $a_k$ is the desired acceleration of the current vehicle, $C_o$ is the location (in feet) of the origin cell, and $C_1$ is the location (in feet) of the destination cell.

Once the location of the destination cell is calculated, it is converted to an index value. Then, the cell specified by the index and the cells immediately upstream of the destination cell are marked as occupied on the basis of the length of the car by using the following conditional logic

$$C_1 - \frac{car\_length}{cell\_length} > C_j \rightarrow C_j \;\; is \;\; occupied\,, \tag{12}$$

where $C_j$ is the upstream cell under consideration and $j$ is the index number of the upstream cells.

Next, the cells that were previously occupied by the vehicle are emptied so that upstream cars have cells to move into. Finally, cells are sequentially checked by index number (in reverse) until the next occupied cell is found. Then the car occupying this cell is set as the next car, and the process repeats itself until every cell in the model has been checked. At that point, the model has completed one time step.

The next step is determining how to add or remove vehicles. In general, the only way to get vehicles on and off the freeway is via the ramps and the entrance and exit for the model. For these purposes, the vehicles entering at the beginning of the model are treated the same way as cars entering via the on-ramps in order to simplify the operation. The rate at which vehicles exit the road is based on the function of the entire model dynamics. The rate of vehicles exiting the road at the end is not directly controlled, but the vehicles will exit the model at the same rate as the physical system if the rest of the model is functioning correctly. In order to calculate the flow of vehicles out of the model via the off-ramps, a destination is assigned to each vehicle as it is added to the model. This destination, either an off-ramp or the end of the model, is checked every time the vehicle moves. When the vehicle reaches its 'destination,' it is removed from the model.

The information corresponding to where each vehicle enters and leaves the freeway is called origin-destination (O-D) information. The proposed method of calculating the number

of cars entering and leaving the freeway at any specific time is to use O-D information. It is important to note that this method does not just calculate the O-D percentages but the O-D rate at each ramp. In other words, it calculates a time varying rate that represents the arrival rate of cars that enter at ramp *i* and exit at ramp *j* for all possible pairs of *i* and *j* at each time *t*. By using these data, the model will be able to properly add and subtract cars such that the number of cars in the model at any particular time matches the number of cars actually on the freeway.

The proposed method of estimating the O-D information uses a combination of O-D and time-of-day vehicle arrival rate information. Even though traffic arrival rates on a day-to-day basis seem to be stationary, the traffic O-D rate itself is a non-stationary process that cannot be measured directly by turning percentages [56]. A solution to this problem is the generation of dynamic O-D data. This is a difficult problem because of the inability of freeway surveillance systems to detect the destination of a car when it first enters the roadway. However, a number of authors have proposed systems for modeling this phenomenon [30, 32, 57]. Using the definitions proposed by Daganzo [57], O-D information is incorporated into the model by first simplifying the desired departure rate (demand, *D*) equation to imply that there is no pre-existing ($t < 0$) demand to use the freeway,

$$D_{ij}(k) = A_{ij}(k). \tag{13}$$

Now the arrival rate, $A_{ij}(k)$, is simplified into the product of the aggregate rates and time varying function that describes the O-D fractions for the ramp,

$$D_{ij}(k) = A_i(k)\alpha_{ij}(k). \tag{14}$$

The O-D is dynamically predicted by first estimating the arrival event of cars and then by assigning a destination tag to each car as it is placed in the model using O-D fractions. With this information, the simulation will be able to calculate the on-ramp flow rate using the aggregate rate function and the off-ramp flow rate from the destination tags.

To predict the arrival event of cars, historical rate and real-time information are combined with the short-term prediction method of Akahane and Koshi [58] to calculate the $A_i(k)$ in Equation (14). This method predicts the ramp inflow volume by using an appropriate historical record of volumes, an adaptive parameter model, and a measurement

of current ramp volumes. Next, a method is proposed to calculate $\alpha_{ij}$ to complete the equation.

The second part of dynamically determining the O-D is to assign the destination of the car when it enters the road. Several techniques for estimating O-D information have been proposed in the literature [30, 59]. This project adopted the method by Chang and Li [30] to dynamically predict O-D fractions - $\alpha_{ij}$. While a primary concern in the literature is calculating the off-ramp flow, the off-ramp flow in the model is determined during the execution of the model and is not required initially. Now that $D_{ij}$ is available, it can be used to calculate the inflow rate at all points. In addition, when a new vehicle is added to the model, the model sets the destination flag of the car such that in the future, the rate of cars exiting the road is correct.

To incorporate this functionality into the model, a small group of cells corresponding to the physical location of the on-ramps (corresponding to the index value $i$) is designated as entry cells, and another small group of cells corresponding to the physical location of the off-ramps (corresponding to the index value $j$) is designated as exit cells. The state of these cells is updated externally to represent the effect of vehicles entering the freeway using on-ramps. Since the microscopic model is developed for control purposes, a control input, $u_i$, is defined as the number of cars added to the road at point $i$ in one time step. In general, the function $D_{ij}$ adds cars into a ramp queue, $q_i$. At every update step, the state of the cell is checked to see if the control input specifies the addition of a car to the model. If the controller specifies adding a car and the freeway has room, then the car is removed from the queue and added to the roadway. Otherwise, it remains in the queue of cars waiting to enter the roadway. If a vehicle flagged with destination $j$ is moved beyond point $j$, it is removed from the model. In this way, the effect of the functions $D_{ij}$ and the control input $u_i$ in the model is recreated.

At this point, the necessary components for the microscopic simulation model have been completely defined: the framework, the method of propagating vehicles inside the model, and the method of adding and subtracting cars from the model. However, as the information in this model is intended to be used directly for control, it is also important that the model is not too far removed from the real-time sensor data. Real-time data are incorporated into the

operation of the model in two ways. As stated before, the real-time data are used during the process of calculating dynamic O-D rate information. Therefore, the aggregate rate at the current time depends on the real-time sensor information.

Second, the accuracy of the model is verified by comparing it to the real-time mainline data. In order to do this, the time series output of a selected set of loop detectors is compared with the state of the model at the same location. The set of detectors will be chosen such that all of them are known to be operating properly, and the entire freeway system being modeled is covered. If the correlation between the model and the sensors drops below a certain point, then the model will conclude that it is no longer accurate in the particular region (where the correlation drop occurred).

Now that all of the pieces necessary to describe the microscopic model are assembled, the next step is to fully create it in a realizable form. To this end, a computer simulation that is able to update each state using the prescribed driver model is created. This simulator displays an accurate picture of the freeway corridor that can be analyzed and used in any number of applications; however, in this case, the first step is to verify that the model accurately represents the traffic dynamics.

In the development of macroscopic models, the basic equations used are statements of conservation, speed, and flow. The verification process for the microscopic model will rely on the same variables used to verify the macroscopic model. Therefore, even though the design of the model is specified in terms of how drivers behave, how detailed the topological description of the road is, and how cars enter the roadway, the performance of the model is evaluated by how well it matches the speed, flow, and conservation properties of the data. In general, if the microscopic model is able to reproduce the same speed, flow, and conservation information as the real data under the same conditions, then the model is assumed to be valid for those conditions. The model will be complete once it is validated for all normal operating conditions.

Of the many driver model parameters available, three parameters were chosen as primary: the desired headway, asymmetrical lane changing probabilities, and min/max breaking probabilities. The headway model is used in the first step of the driver behavior tree. A driver's desired headway is the basic measure of the way that person drives; a large

headway means that the person is a cautious driver, whereas a small headway means that the person is more aggressive. Depending on the difference between the desired headway and actual headway, the driver will perform different actions (change lane, change speed, etc.). These parameters directly affect the limits of the flow of vehicles through the model. The asymmetrical lane changing probabilities govern the tendency of a driver to change lanes in order to increase speed. The min/max breaking probabilities govern the tendency of a driver to brake when the headway is smaller than a desired value. In these situations, the driver will brake according to these probabilities.

To test the suitability of the microscopic model in predicting the volume, speed, and density of the freeway using a set of known inputs, a test was run to check the output during times when the freeway is not congested. Using the simulator, a line-search algorithm was used to optimize the error between the model output and the historical data as a function of a number of model parameters. Parameters that did not have a large effect on model performance or that were known constants (legal speed limit) were not optimized in order to reduce the time of the computation. Figures 14 and 15 show the output of the simulation after optimization.

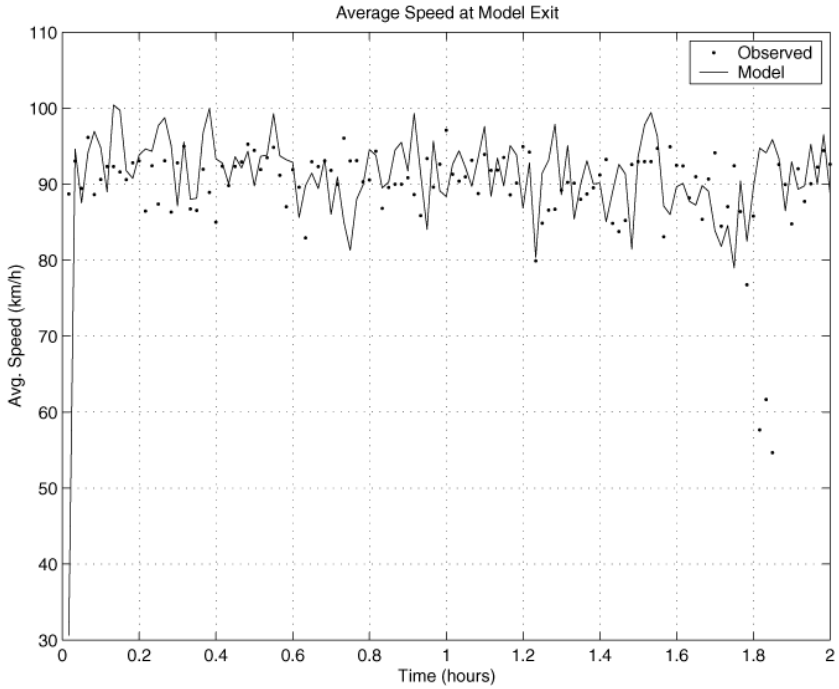


**Figure 14. Comparison of average speed of cars exiting the model with observed data.**

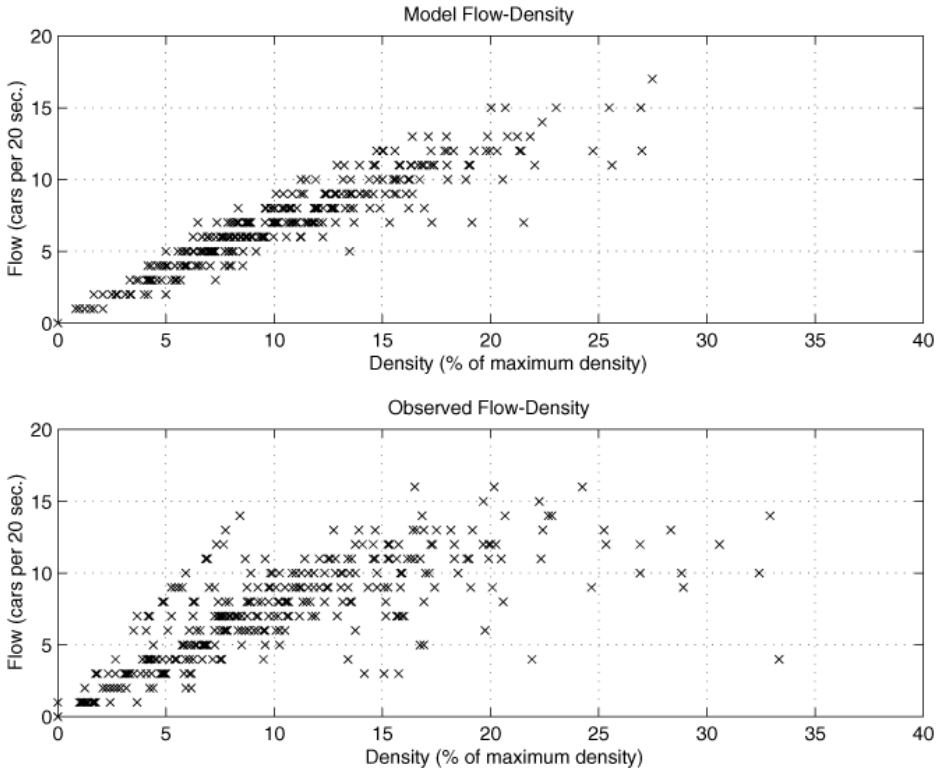


**Figure 15. Flow-density comparison between the model and observed data.**

These graphs demonstrate the ability of the simulator to model traffic with the microscopic model when it is not congested up to the point at which traffic becomes congested. While the simulator is currently not able to detect the incidents that cause non-recurring congestion, with additional effort it should be possible to create a microscopic model that is able to represent the effects of non-recurring congestion. The ability of the microscopic traffic model to represent the effects of non-recurring congestion is a significant improvement over the existing macroscopic models. Moreover, these graphs demonstrate the ability of the microscopic model to return data that closely match the data returned by the surveillance system

### *3.3.4 Ramp Access Control*

#### 3.3.4.1 Problem Description

With the microscopic simulator able to calculate the progression of the traffic flows and densities, the next step is to formulate a method to improve the operation of the freeway. Traffic theory suggests that if the traffic density on the freeway increases beyond a certain

point, the traffic flow out of that area decreases. When the flow of traffic is reduced, the resulting delay in travel time is defined as congestion. Congestion causes two primary negative effects: the travel time of vehicles on the roadway increased and the number of vehicles able to use the freeway decreases. These primary problems also cause a number of secondary problems, such as increased road wear, increased pollution, and overcrowding of alternative routes. Accordingly, there is a great benefit to reducing congestion.

Having proposed a microscopic model of traffic flow and chosen the goal of eliminating or preventing congestion, the next task is to develop a method of doing this. An effective method of reducing congestion is to limit the flow of vehicles that are allowed to enter the freeway. In the simplest case, if all additional vehicles were prevented from entering the freeway for a sufficiently long period, the freeway would completely empty and the congestion would be gone. Clearly, some cars must be let on the freeway, but by spreading out the vehicles wanting to use the freeway over time, it is possible to decrease congestion (or prevent the onset of congestion). Currently, the most prevalent method of limiting the volume of cars entering the roadway is ramp metering. Ramp metering with traffic signals allows an outside entity such as the police or a traffic authority to restrict access to the freeway.

There are two main categories of ramp metering policies: global and local. In a local metering system, the method used to limit the flow is solely a function of the local state (i.e., local density, flow, queue length), while in a global metering control system, the method considers the states of other sections of the freeway as well as local variables (i.e., downstream density, upstream speed). In general, global methods are superior to local methods because they are able to account for a larger variety of traffic conditions. In this work, the implementation of a hierarchical controller to globally control the ramp metering system to reduce or prevent congestion is proposed.

#### 3.3.4.2 Control Objective

In order to specify an objective for control, a number of statistics describing the vehicles on the roadway must first be defined. These statistics are sufficient to get a measure of the total delay on the roadway at any time.

$vw_i(k)$ is the number of vehicles in queue at on ramp $i$ at time $k$.

$N$ is the total number of ramps on the road.

$NV$ is the total number of vehicles on the road.

$vv_m(k)$ is the current speed of vehicle m at time $k$.

$vd_m(k)$ is the desired speed of vehicle m at time $k$.

$vx_m(k)$ is the distance vehicle $m$ has until it reaches its destination at time $k$.

The total delay, $Z$, experienced by drivers during the discrete time period $[0,T]$ is

$$Z = \sum_{k=o}^{T} \left( \frac{\Delta T}{T} \sum_{m}^{NV} \left( \frac{vd_m(k) - vv_m(k)}{vv_m(k)\, vd_m(k)} \right) vx_m(k) + \Delta T \sum_{i}^{N} vw_i(k) \right), \tag{15}$$

where $\Delta T$ is the time step.

The ability of the proposed controller to use this metric to minimize the total delay in the system is a significant improvement over existing models. To calculate the total delay in the system using macroscopic models, a number of approximations must be made.

### 3.3.5 Ramp Controller

A hierarchical control system is proposed that utilizes the microscopic model of traffic dynamics from before, with the goal of minimizing the total delay experienced by drivers in the system. Since the microscopic model of traffic dynamics is contained in simulation form, the first step is to develop a framework for connecting a simulation into a controller. There have been a few recent examples of controllers that directly use a simulated model of a complex system in the feedback loop. One such controller uses a simulation of a glass tank to provide input to an adaptive neuro fuzzy controller [60]. Another controller uses a hybrid simulation model to control a complex environmental system based on a knowledge-based/expert system [61]. In the glass furnace application, the fuzzy controller sets the parameters of a reduced-order, model-based controller that is directly connected to the glass furnace. In tests, the controller was able to reduce the time to reach the desired effect by a factor of 2.5, demonstrating the viability of using the result of a simulation to drive a controller. The incorporation of a microscopic simulator into the feedback loop of a ramp

access controller should be a significant improvement over existing methods because of the superiority of microscopic models in representing accurate traffic dynamics.

The first step in the process of linking the simulator to a controller is to extract information from the simulator that can be used in the control algorithm. As stated before, the state of the microscopic model comprises 4-foot sections of the freeway and the percentage of that section that is occupied by a particular car. In addition, the model contains statistics about each car in terms of its destination, desired speed, current speed, and length. For the controller, the most important quantities are the average speed and density at different locations on the freeway. To get this information, a function $x^* = M(x)$ is defined that calculates a number of statistics on the internal state and returns them in the vector $x^*$.

The operation of $M$ is to divide the freeway into $N$ 300-foot segments and then calculate the density and average speed in each segment. In addition to this, the vehicle-specific statistics used to calculate the overall delay in the system are also calculated. The statistical state $x^*$ is then defined as

$$\begin{aligned} x^* = [&\rho_0, \rho_1, \rho_2, \rho_i, ..., \rho_N, v_0, v_1, v_2, v_i, ..., v_N, vv_0, vv_1, vv_2, vv_j, ..., vv_M, vd_o, vd_1, \\ &vd_2, vd_j, ..., vd_M, vx_0, vx_1, vx_2, vx_j, ..., vx_M, vw_0, vw_1, vw_2, vw_i, ..., vw_N] \end{aligned} \tag{16}$$

where $\rho_i$ is the density in segment $i$, $v_i$ is the average speed in segment $i$, and $vv_j$, $vd_j$, $vx_j$, and $vw_i$ are defined as before.

Treating the model for traffic dynamics as a 'black box,' a function $G$ that describes the input-output relationship of the model can be defined. For simplicity, the operation of the microscopic model and $M$ is combined into $G$. The inputs of $G$ are the inputs to the microscopic model, and the output of $G$ is the output of $M$. Figure 16 is the block diagram describing $G$. The inputs and outputs of $G$ are defined as follows:

u – Steady state ramp metering rate (in vehicles per time step) – [$u_0$, $u_1$, $u_i$, …, $u_N$]

$x^*$ – Time-varying statistical state of the freeway.

The functional representation of $G$ is

$$x^* = G(u). \tag{17}$$

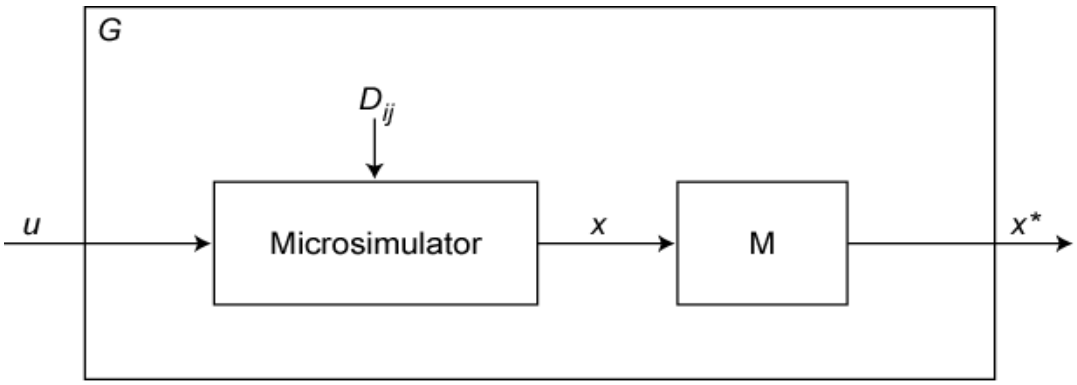


**Figure 16. Mainline traffic function block diagram.**

With this model, it is possible to construct a controller. Following previous work [36, 38], the ramp access controller is broken into two main parts: a single, high-level, system-wide controller and a group of low-level local controllers. The system-wide controller synchronizes the operation of the local controllers by setting the target density and control input, while the local controllers use standard automatic control principles to regulate the traffic using the set points.

#### 3.3.5.1 System-Wide Controller

The system-wide controller is responsible for dynamically setting the target control and density values required by the local controllers. Since the goal of the control is to minimize the congestion on the freeway, it is necessary to define a scalar valued function of the total delay in the system that is a function of the control input. To do this, it is important to note that the contents of $x^*$ are sufficient to calculate the total delay as calculated by Equation (6); therefore, the proposed method to calculate the total delay in the system is to use Equation (7) to calculate $x^*$ with a control input $u$ and then use that as the input to Equation (6). In addition, $T$ is set to 15 minutes. As a result, the controller will be trying to find a $u$ that best controls the traffic over a 15-minute period by solving the nonlinear programming problem

$$\min f\left((G(u)\right), \tag{18}$$

where $f(x^*)$ is the equation to calculate total delay from Equation (7)

In this formulation, $u$ is a matrix describing the control input at every ramp location $i$ at every time step during the 5-minute period. This method should be a significant improvement over existing methods because of the greater granularity that the microscopic model provides.

#### 3.3.5.2 Solution Algorithm

Because the controller is a non-linear programming problem that does not have a standard analytical form, the Hooke and Jeeves direct search algorithm is used to search for a solution [62]. This algorithm has the advantage of not requiring a derivative for success. Another advantage is that at every step an approximation of the final solution is given, so it is possible to limit the number of iterations of the search to meet time constraints (at the cost of accuracy). The disadvantage of a direct search method is that there is no optimality guarantee unless one searches all possible functional values. Applying the general algorithm from Haberman [63] directly to the problem, the following states are obtained:

- *Step 1:* Set the initial condition of the vector of control inputs **u**. Use the previous optimization result if it exists; otherwise set the control inputs to allow the maximum flow.
- *Step 2:* Set the outside loop to count through the index of ramps by iterating down the control vector corresponding to the time step from top to bottom.
- *Step 3:* For the desired input, make an exploratory move by varying input $u_i$. If the function value decreases, set the new value to the desired input and repeat step 3. If the function value does not decrease, go to step 4.
- *Step 4:* Check the step size. If the step size is not at the minimum, decrease the step size and go to step 3. If the step size is at the minimum, move to the next ramp and go to Step 3. If this is the last ramp, then stop.

The result of this algorithm is the control input u that is necessary for the local controller to operate. In addition, $G(u)$ can be evaluated to get the steady state value of ρ relating to this control input.

#### 3.3.5.3 Local Controller

The function of the local controller is to regulate the density of traffic in the region corresponding to each on-ramp (indexed by $i$). To accomplish this, a discrete-time non-linear $H$-infinity controller is implemented by expanding upon methods used by Kachroo and Rakha [44]. The local controller regulates traffic by adjusting the rate (flow) of cars

entering the roadway via the ramp metering system. The need to use a local controller arises from the fact that the system controller only updates the control parameters every 15 minutes. To be responsive to local variations in the traffic flow during the 15 minutes between updates, local control regulates the ramp metering.

The output of the local controller is a time series signal that can be converted by the ramp metering system into a series of green light pulses. Figure 17 shows a basic diagram describing the parameters of the local area control problem. See Appendix A for a definition of terms.

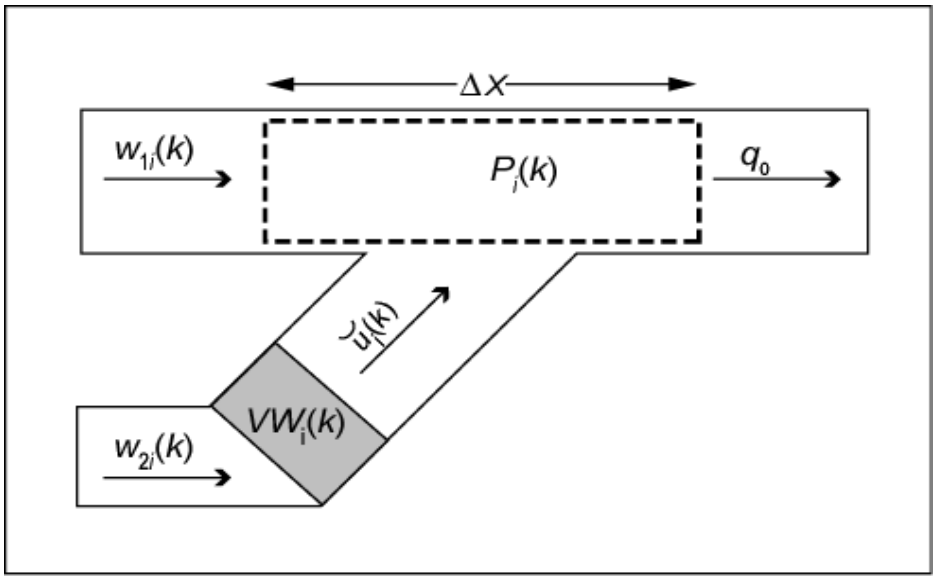


**Figure 17. Local controller structure.**

The basic formula for modeling the local system is derived from the macroscopic flow model and uses the principle of conservation. In general, cars are conserved, which requires that the flow of cars into the region either has a corresponding flow of cars out of the region or a corresponding increase in density. The discrete time formula for this (from Papageorgiou [13]) is shown below,

$$P_i(k+1) = P_i(k) + \frac{\Delta T}{\Delta X}\left[ w_{1i}(k) + \breve{u}_i(k) - q_o(k) \right] . \quad (19)$$

To correctly model the regulator function, an equation to model the growth of the number of vehicles waiting to use the on-ramp is added [44],

$$VW_i(k+1) = VW_i(k) + \Delta T\left[w_{2i}(k) - \breve{u}_i(k)\right] . \tag{20}$$

Because one control objective is to have the on-ramp flow match the output of the system-wide controller, the controlled on-ramp flow is defined as a state, and the desired deviation from the on-ramp flow (as defined by the system-wide controller) is defined as the control output with the following formula,

$$\breve{u}_i(k+1) = \breve{u}_i(k) + \tilde{u}_i(k) , \tag{21}$$

where $\tilde{u}_i$ is the control output of the local controller.

A linear function is now used to approximate the speed-density relationship (from Haberman [63]),

$$v(k) = v_f\left(1 - \frac{P_i(k)}{\rho_m}\right) , \tag{22}$$

where $v_f$ is the maximum speed in segment *i* (determined by the speed limit of the road) and $\rho_m$ is the maximum density in segment *i* (determined by the Highway Capacity Manual).

Next, Equation (12) is substituted into the formula for $q_o(k)$,

$$q_o(k) = v_f P_i(k)\left(1 - \frac{P_i(k)}{\rho_m}\right) , \tag{23}$$

Then $q_o(k)$ in Equation (9) is replaced with Equation (13), resulting in (from Haberman [44])

$$P_i(k+1) = P_i(k) + \frac{\Delta T}{\Delta X}\left[w_1(k) + \tilde{u}_i(k) - v_f P_i(k)\left(1 - \frac{P_i(k)}{\rho_m}\right)\right] . \tag{24}$$

Using equations (20), (21), and (24), the state vector $\breve{x}$ is defined as $\left[P_i, VW_i \breve{u}_i\right]$. Now, the state equations can be defined by

$$\breve{x}(k+1) = \begin{bmatrix} 1+\dfrac{\Delta T}{\Delta X}\left[\dfrac{v_f \breve{x}_1}{\rho_m} - v_f\right] & \dfrac{\Delta T}{\Delta X} & \dfrac{\Delta T}{\Delta X} \\ 0 & 1 & -\Delta T \\ 0 & 0 & 1 \end{bmatrix} \breve{x}(k) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tilde{u}_i(k) + \begin{bmatrix} \dfrac{\Delta T}{\Delta X} & 0 & 0 \\ 0 & \Delta T & 0 \\ 0 & 0 & 1 \end{bmatrix} w(k) \quad (25)$$

where $\breve{x}_1$ is the first element of the state vector $\breve{x}$.

Next an *H*-infinity controller for the above state space is constructed. To do this, the objective for control is first specified,

$$z(k) = \left[ w_1 \left(1 - \frac{\rho_i}{\breve{x}_1}\right) \quad w_2 \quad w_3 \left(1 - \frac{u_i}{\breve{x}_3}\right) \right] \breve{x}(k) + \left[ w_4 \right] \tilde{u}_i(k) \ , \qquad (26)$$

where $w_1, w_2, w_3, w_4$ are weighting coefficients, $\breve{x}_1$ is the first element of the state vector $\breve{x}$, and $\breve{x}_3$ is the third element of the state vector $\breve{x}$.

The set points $\rho_i$ and $u_i$ are both set by the system-wide controller ($u_i$ is the result of the optimization routine and $\rho_i$ is available from x* = $G(u)$). The weighting points are chosen such that the controller balances between the goal of matching the target density and the constraint of keeping queue length to a minimum.

The block diagram for the *H*-infinity controller is shown in Figure 18. In the figure, $w(k)$ is the external input, $y(k)$ is the output, $z(k)$ is the controlled output, and $\breve{u}(k)$ is the controlled input.

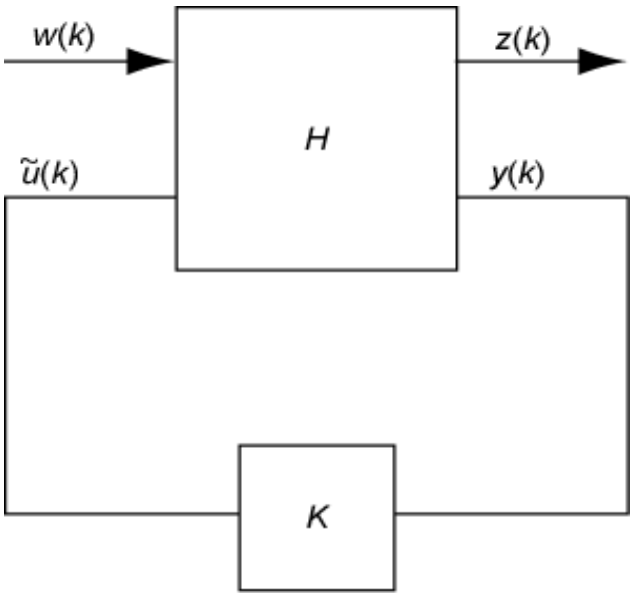


**Figure 18. Local controller block diagram.**

In order to calculate the control law *K*, a solution that can account for the nonlinear term in the state equation must be used. While it is possible to create a linear disturbance model, in this case, the solution is found directly by using state-dependent Riccati equation (SDRE) techniques [64]. To do this, the system must be put into a form in which the non-linear dynamics are represented by direct parameterization such that $\breve{x}_k$ is the state at which the dynamics are evaluated.

Thus, using the state equations of (25), define

$$\breve{x}(k+1) = A(\breve{x}_k)\breve{x}(k) + B_1(\breve{x}_k)w(k) + B_2(\breve{x}_k)\tilde{u}(k), \tag{27}$$

and from Equation (26) define

$$z(k) = C_1(\breve{x}_k)\breve{x}(k) + D_{12}(\breve{x}_k)\tilde{u}(k). \tag{28}$$

In addition, it is assumed that the measurements differ from the actual values by $N(0,\sigma)$, so the observer equation becomes

$$y(k) = C_2(\breve{x}_k)\breve{x}(k) + D_{21}(\breve{x})w(k)\ , \tag{29}$$

with $C_2 = I$ and $D_{21} = \text{diag}(w_5,w_6,w_7)$, where $w_5$,$w_6$, and $w_7$ are weighting coefficients. It is important to note that the measurements of the state come directly from the microscopic

simulation. In this way, the problems caused by missing data and by the conversion between occupancy and density are avoided. Accordingly, the elements of $D_{21}$ represent the uncertainty in the simulated data.

An important assumption at this point is that the parameterizations of ($A$,$B_1$), ($A$,$B_2$), and ($C_1$,$A$) are stabilizable for $x \in \Omega$, where $\Omega$ is the region of interest [64]. Now, the augmented state matrix is created for this system by evaluating the matrices at the control point $\breve{x}_k$,

$$H(k) = \begin{bmatrix} \breve{x}(k+1) \\ z(k) \\ y(k) \end{bmatrix} = \begin{bmatrix} A(\breve{x}_k) & B_1(\breve{x}_k) & B_2(\breve{x}_k) \\ C_1(\breve{x}_k) & D_{11}(\breve{x}_k) & D_{12}(\breve{x}_k) \\ C_2(\breve{x}_k) & D_{21}(\breve{x}_k) & D_{22}(\breve{x}_k) \end{bmatrix} \begin{bmatrix} \breve{x}(k) \\ w(k) \\ u(k) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{30}$$

and taking the $z$-transform of Equation (30),

$$H(z) \coloneqq C(zI - A)^{-1} B + D. \tag{31}$$

Now, the bilinear transformation,

$$z = \frac{1 + \frac{s}{c}}{1 - \frac{s}{c}}, \tag{32}$$

where $c$ is a positive constant, is used to transform $H(z)$ from the discrete domain to the continuous domain, with an augmented state matrix of

$$H(s) = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{bmatrix}. \tag{33}$$

A $\gamma$ is selected such that the solutions to the Riccati equation, $P(x_k) \geq 0$, $Q(x_k) \geq 0$, exist with $\lambda_{\max}\left[P(x_k)Q(x_k)\right] < \gamma^2$ [50]. Now the continuous time $H_\infty$ solution from Gawron et al. [50] is used

$$\begin{bmatrix} A - B_2 R_u^{-1} D_{12}^T C_1 & \gamma^{-2} B_1 B_1^T - B_2 R_u^{-1} B_2^T \\ -\hat{C}_1^T \hat{C}_1 & -\left(A - B_2 R_u^{-1} D_{12}^T C_1\right)^T \end{bmatrix} \tag{34}$$

$$\begin{bmatrix} \left(A - B_1 D_{21}^T R_w^{-1} C_2\right)^T & \gamma^{-2} C_1^T C_1 - C_2^T R_w^{-1} C_2 \\ -\hat{B}_1 \hat{B}_1^T & -\left(A - B_1 D_{21}^T R_w^{-1} C_2\right) \end{bmatrix}, \tag{35}$$

where $\hat{C}_1 = \left(I - D_{12} R_u^{-1} D_{12}^T\right) C_1 \quad \hat{B}_1 = B_1 \left(I - D_{21}^T R_w^{-1} D_{21}\right)$

$$R_u = D_{12}^T(\hat{x}) D_{12}(\hat{x}) \qquad R_w = D_{21}(x) D_{21}^T(x)$$

Next the observer equation is constructed,

$$\dot{\hat{x}} = A_o \hat{x} + B_o y \,, \tag{36}$$

and the related control law,

$$u = F\hat{x} \,, \tag{37}$$

where $F = -R_u^{-1}\left(B_2^T P + D_{12}^T C_1\right) \qquad L = -\left(QC_2^T + B_1 D_{21}^T\right) R_w^{-1}$

$$B_o = -ZL \qquad Z = \left(I - \gamma^{-2} QP\right)^{-1}$$

$$A_o = A + B_2 F + \gamma^{-2} B_1 B_1^T P + ZL\left(C_2 + \gamma^{-2} D_{21} B_1^T P\right) ,$$

again using the solution from Gawron et al. [50].

Next the inverse bilinear transform of equations (36) and (37) is taken, resulting in

$$\hat{x}(k+1) = A_{ok} \hat{x}\left(k\right) + B_{ok} y(k) \tag{38}$$

$$\tilde{u}(k) = F_k \hat{x}\left(k\right). \tag{39}$$

Using the result of this controller, the next control input describing the number of cars allowed to enter the freeway in the next time step in the local region *i* at time *k* is

$$\breve{u}_i(k) + \tilde{u}(k) \,. \tag{40}$$

For this controller to be successful, a few key implementation issues need consideration. First, the controller must be driven at a relatively high frequency to account for the time varying Hamiltonian. Fortunately, the period of the system itself is at minimum 20 seconds, so solving the SDRE every 10 seconds should be appropriate to ensure that the controller has the right dynamics for the current state. Second, while SDRE controllers are not optimal,

they have been shown to be the local optimal solution for the control problem [65]. Recent work also demonstrates that the SDRE solution is an effective control design technique [66] that can outperform local linearized models of control [67].

# 4. CONCLUSIONS AND RECOMMENDATIONS

## *4.1 Conclusions*

In this report, a controller for regulating access to the freeway using ramp metering is proposed. The proposed controller is original in that it utilizes a microscopic simulation of the freeway as a dynamic model by which control strategies can be determined. The microscopic simulator provides a significant advantage over existing models of traffic flow because it provides increased flexibility in the way that traffic flow dynamics are modeled and evaluated. To this end, the model extends existing dynamic O-D prediction routines in a new way that allows for efficient calculation of predicted exit flows by using the prediction of entrance flows. The microscopic simulator also has the ability to model the freeway and its bottlenecks with a significantly finer granularity than has possible with macroscopic models.

Beyond the advances possible from the microscopic model, the controller also incorporates $H_\infty$ control techniques in an original way to solve the ramp access control problem. The hierarchical controller breaks the control problem into two sub-problems: the calculation of a system-wide control scheme and the implementation of the control scheme at the local level. The $H_\infty$ controller takes set points provided to it by the system-wide controller and regulates the traffic density accordingly. This is a significant improvement over other $H_\infty$ controllers that only work by using pre-calculated values in a local region. In addition, the $H_\infty$ controller utilizes measurements from the microscopic model that were previously unavailable through established means.

In general, the controller presented here represents an improvement over existing controllers. It utilizes real-time, simulated, and historical data concurrently to achieve a detailed estimation of the state of the freeway. Using the estimation of freeway conditions, the controller dynamically sets operating points that consider the entire state of the freeway network and then executes control to reach target objectives at the local level.

## *4.2 Future Work*

The process of building a completed controller that can be implemented in a real-world environment involves many steps. The following issues need to be addressed to finalize the

implementation. After the resolution of key issues, the controller can be tested and its operation validated.

### 4.2.1 Microscopic Simulator Models Congestion States

In order for the microscopic simulator to provide the information necessary to control the freeway at all times, the simulator must be able to determine when an incident occurs. While the existing framework for the simulator is general enough to be able to show the effects of congestion, the driver model must be updated and calibrated to increase the verisimilitude of the simulator.

The existing performance measures are robust enough to determine whether the simulator is properly functioning, but a wide range of data must be used to ensure that the simulator is performing correctly. Once the simulator can model both the recurring and non-recurring congested states of the freeway given a variety of data sets, the rest of the controller can be built around it.

### 4.2.2 Incident Handling

Currently, when the model detects an incident (or becomes uncorrelated from the real data), the controller scheme is unable to provide a valid control input. Instead, the controller must wait until the model and real data become correlated again. It should be possible to retroactively adjust the model to correct for the discrepancies. Methods for doing this could include dynamically altering the driver model to take into account a change in the traffic flow, adding obstacles into the model (to simulate incidents or closed lanes) and then letting the normal driver model respond accordingly, or arbitrarily adding or subtracting cars to adjust the volumes in the model to match the real-time data.

Once a scheme for retroactive adjustment is in place, it will be possible to control ramp access during times immediately after an incident. The ability to reduce congestion associated with a non-recurring event (such as a stall or accident) will greatly improve the overall performance of the controller and increase the number of situations in which using the controller is advantageous.

### 4.2.3 Real World Implementation

Currently, the ramp access controller, as it is proposed here, has not been customized to work with an actual freeway system. To further analyze the significance of the controller, it would be beneficial to customize it to work with an actual section of the freeway and evaluate how well the controller operates in a real-world environment. By implementing the controller and testing it against real-world conditions, the ability of the controller to handle the intricacies of actual conditions could be evaluated.

## REFERENCES


1. Cleghorn, D., F. Hall, and D. Garbuio. Improved Data Screen Techniques for Freeway Traffic Management Systems. *Transportation Research Record 1320*, TRB National Research Council, Washington, D.C., 1991, pp.17-23.
2. Jacobson, L., N. Nihan, and J. Bender. (1990). Detecting Erroneous Loop Detector Data in a Freeway Traffic Management System. *Transportation Research Record 1287*, TRB National Research Council, Washington, D.C., pp. 151-166.
3. Nihan, N. Aid to Determining Freeway Metering Rates and Detecting Loop Errors. *Journal of Transportation Engineering*, Vol. 123, No. 6, 1997, pp. 454-458.
4. Coifman, B. Using Dual Loop Speed Traps to Identify Detector Errors. *Transportation Research Record 1683*, TRB National Research Council, Washington, D.C., 1999, pp. 47-58.
5. Ametha, J., S. Turner, and S. Darbha. Formulation of a New Methodology to Identify Erroneous Paired Loop Detectors. *2001 IEEE Intelligent Transportation Systems Proceedings*, 2001, pp. 591-596.
6. Coifman, B. Improved Velocity Estimation Using Single Loop Detectors. *Transportation Research A,* Vol. 35A, No. 10, 2001, pp. 863-880.
7. Zhanfeng, J., C. Chao, B. Coifman, and P. Varaiya. The PeMS Algorithms for Accurate, Real-Time Estimates of g-factors and Speeds From Single-Loop Detectors. *2001 IEEE Intelligent Transportation Systems Proceedings*, 2001, pp. 536-541.
8. Chen, L. and A. May. Traffic Detector Errors and Diagnostics. *Transportation Research Record 1132*, TRB, National Research Council, Washington, D.C., 1987, pp. 82-93.
9. Dailey, D. and N. Taiyab. *A Cellular Automata Model for Use with Real Freeway Data*. WA-RD 537.1. Technical Report. Washington State Department of Transportation. June 2002.
10. FHWA. *Traffic Monitoring Guide*. Section 3, Chapter 4, USDOT FHWA-PL-01-021, May 2001.

11. Wattleworth, J. and D. Berry. Peak-Period Control of a Freeway System - Some Theoretical Investigations. *Highway Research Record*, Vol. 89, 1965, pp. 1-25.

12. Isaksen, L. and H. Payne. Suboptimal Control of Linear Systems by Augmentation with Application to Freeway Traffic Regulation. *IEEE Transactions on Automatic Control*, Vol. 18, No. 3, 1973, pp. 210-219.

13. Papageorgiou, M. *Applications of Automatic Control Concepts Flow Modeling and Control*. New York: Springer-Verlag, 1983.

14. McDonnell, J., D. Fogel, L. Fogel, C. Rindt, and W. Recker. Evolving Optimal Ramp Control Rules. *International Journal of Expert Systems*, Vol. 8, No. 3, 1995, pp. 287-308.

15. Zhang, H., S. Ritchie, and W. Recker. Some General Results on the Optimal Ramp Control Problem. *Transportation Research C*, Vol. 4, No. 2, 1996, pp. 51-69.

16. Wu, J. and G. Chang. An Integrated Optimal Control and Algorithm for Commuting Corridors. *International Transactions in Operational Research*, Vol. 6, No. 11, 1999, pp. 39-55.

17. Alessandri, A., A. Di Febbraro, A. Ferrara, and E. Punta. Optimal control of freeways via speed signaling and ramp metering. *Control Engineering Practice*, Vol. 6, 1998, pp. 771-780.

18. Zhang, H., R. Jayakrishnan, and W. Recker. An Optimization Algorithm for Freeway Traffic Control. *ITSC 2001, 2001 IEEE Intelligent Transportation Systems Proceedings*, 2001, pp. 100-105.

19. Sasaki, T. and T. Akiyama. Traffic Control Process of Expressway by Fuzzy Logic. *Fuzzy Sets and Systems*, Vol. 26, No. 2, 1988, pp. 165-178.

20. Chen, L., A. May, and D. Auslander. Freeway Ramp Control Using Fuzzy Set Theory for Inexact Reasoning. *Transportation Research A*, Vol. 24A, No. 1,1990, pp. 15-25.

21. Taylor, C.E. and D.R. Meldrum. *Evaluation of a Fuzzy Logic Ramp Metering Algorithm: A Comparative Study Among Three Ramp Metering Algorithms Used in*

*the Greater Seattle Area.* Technical Report WA-RD 481.2. Washington State Department of Transportation. February 2000.

22. Taylor, C.E. and D.R. Meldrum. *Algorithm Design, User Interface, and Optimization Procedure for a Fuzzy Logic Ramp Metering Algorithm: A Training Manual for Freeway Operations Engineers.* WA-RD 481.1. Technical Report. Washington State Department of Transportation. February 2000.

23. Bogenberger, K., K. El-Araby, and H. Keller. (2000). Design of a Genetic Fuzzy Approach for Ramp Metering. *ITSC 2000, IEEE Intelligent Transportation Systems Proceedings*, 1-3 October 2000, Dearborn, MI, pp. 470-475.

24. Lee, S., Y. Teng, and W. Wang. Highway Ramp Control via Fuzzy Logic. *Ninth IEEE International Conference on Fuzzy Systems*, 7-10 May 2000, San Antonio, TX, Vol. 1, pp. 274-278.

25. Bogenberger, K., H. Keller, and S. Vukanovic. (2001). A Neuro-Fuzzy Algorithm for Coordinated Traffic Responsive Ramp Metering. *ITSC 2001, IEEE Intelligent Transportation Systems Proceedings*, 25-29 August 2001, Oakland, CA, pp. 94-99.

26. Zhang, H. and S. Ritchie. Freeway Ramp Metering Using Artificial Neural Networks. *Transportation Research C*, Vol. 5, No. 5, 1997, pp. 273-286.

27. Hiramatsu, A., N. Kazuo, H. Shimoura, and K. Tenmoku. Adaptive Genetic Algorithm for On-ramp Traffic Control. *2000 IEEE International Conference on Systems, Man and Cybernetics*, 8-11 October 2000, Nashville, TN, Vol. 5, Part 5, pp. 3672-3677.

28. Fallah-Tafti, M. The Application of Artificial Neural Networks to Anticipate the Average Journey Time of Traffic in The Vicinity of Merges. *Knowledge Based Systems*, Vol. 14, Nos. 3-4, 2001, pp. 203-211.

29. Lighthill, M. and G. Whitham. (1955). On kinematic waves: II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society*, Vol. 229, pp. 317-345.

30. Chang, T. and Z. Li. Optimization of Mainline Traffic Via An Adaptive Co-Ordinated Ramp-Metering Control Model with Dynamic OD Estimation. *Transportation Research C*, Vol. 10, No. 2, 2002, pp. 99-120.

31. Mahmoud, M., M. Hassan, and D. Mohamed, D. *Large-Scale Control Systems: Theories and Techniques*. New York: Marcel Dekker, 1985

32. Chen, O., A. Hotz, and M. Ben-Akiva. Development and Evaluation of a Dynamic Ramp Metering Control Model. *Proceedings from the 8th IFAC/IFIP/IFORS Symposium*, 16-18 June 1997, Chania, Greece, Vol. 3, Part 3, pp. 1089-1095.

33. Wan, B. and S. Qin. Recent developments in large-scale systems research in China. *Proceedings of the IEEE International Symposium on Industrial Electronics*, 25-29 May 1992, Xian, China, Vol. 1, pp. 22-28.

34. Greenlee, T.L. and H.J. Payne. Freeway Ramp Metering Strategies for Responding to Incidents. *IEEE Conference on Decision and Control*, New Orleans, LA, 1977.

35. Looze, D., P. Houpt, N. Sandell, and M. Athans. On Decentralized Estimation and Control With Application to Freeway Ramp Metering. *IEEE Transactions on Automatic Control*, Vol. 23, No. 22, 1978, pp. 268-275.

36. Papageorgiou, M. Multilayer Control of Freeway Traffic. *Proceedings of the IEEE Conference on Decision and Control*, 1982.

37. Mahmoud, M.S. and S.Z. Eid. Optimization of Freeway Traffic Control Problems. *Optimal Control Applications and Methods*, Vol. 9, No. 1, 1988, pp. 37-49.

38. Yang, L. Model and Technology of the On-Ramp Control Over Freeway Traffic Flow. *A Postprint Volume from the IFAC Symposium on Transportation Systems*, 24-26 August 1994, Tianjin, China, Vol. 2, Part 2, pp. 859-864.

39. Gu, J.C., and B.W. Wan. Steady State Hierarchical Optimizing Control for Large-Scale Industrial Processes with Fuzzy Parameters. *IEEE Transactions on Systems, Man, and Cybernetics- Part C: Applications and Reviews*, Vol. 31, No 3, August 2001, pp. 352-360

40. Papageorgiou, M. A Hierarchical Control System for Freeway Traffic. *Transportation Research B*, Vol. 17B, No. 3, 1983, pp. 251-261.

41. Papageorgiou, M., H.S. Habib, and J.M. Blosseville. ALINEA: A local feedback control law for on-ramp metering. *Transportation Research Record 1320*, 1991, pp. 58-64.

42. Payne, H., W. Thompson, and L. Isaksen. Design of a Traffic-Responsive Control System for a Los Angeles Freeway. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 3, No. 3, 1973, pp. 213-224.

43. Haj-Salem, H., P. Poirier, J.–F. Heylliard, and J.–P. Peynaud. ALINEA: a local Traffic Responsive Strategy for Ramp Metering: Field Results on A6 Motorway in Paris. *ITSC 2001, IEEE Intelligent Transportation Systems Proceedings*, 25-29 August 2001, Oakland, CA, pp. 106-111.

44. Kachroo, P. and H. Rakha. Isolated Ramp Metering Feedback Control Utilizing Mixed Sensitivity for Desired Mainline Density and the Ramp Queues. *Proceedings of the SPIE-The International Society for Optical Engineering. Proceedings of the 1998 Mobile-Robots XIII, and Intelligent Transportation Systems*, 3-5 November 1998, Boston, MA, Vol. 3525, pp. 458-466.

45. Daganzo, C. Requiem for Second-Order Fluid Approximations of Traffic Flow. *Transportation Research B*, Vol. 29B, August 1995, pp. 277-286.

46. Daganzo, C. A Continuum Theory of Traffic Dynamics For Freeways With Special Lanes. *Transportation Research B*, Vol. 31B, April 1997, pp. 83-102.

47. Klar, A. and R. Wegener. Vehicular Traffic: From Microscopic to Macroscopic Description. *Transport Theory and Statistical Physics*, Vol. 29, Nos. 3-5, 1997, pp. 479-493.

48. Kuhne, R.D. and M.B. Rodiger. Macroscopic Simulation Model for Freeway Traffic With Jams and Stop-Start Waves. *Winter Simulation Conference Proceedings*, 8-11 December 1991, Phoenix, AZ, pp.762-770.

49. Shvetsov, V. and D. Helbing. Macroscopic Dynamics of Multilane Traffic. *Physical Review E*, Vol. 59, June 1999, pp. 6328-6339.

50. Gawron, C., S. Krauss, and P. Wagner. Metastable states in a microscopic model of traffic flow. *Physical Review E*, Vol. 55, May 1997, pp. 5597- 5601.

51. Benz, T. The Microscopic Traffic Simulator AS (Autobahn Simulator). *Proceedings of the 1993 European Simulation Multiconference SCS*, 1999, San Diego, CA, pp. 486-489.

52. Emmerich, H. and E. Rank. An improved cellular automaton model for traffic flow simulation. *Physica A*, Vol. 234, 1997, pp. 676 – 686.

53. Nagel, K., P. Simon, P. Wagner, and D.E. Wolf. Two-lane Traffic Rules for Cellular Automata: A Systematic Approach. *Physical Review E*, Vol. 58, August 1998, pp. 1425-1437.

54. Saito, T., K. Yasui, S. Fujii, and S. Itakura. Development of Microscopic Simulation Model for Traffic Network (MICSTRAN-II) and Traffic Flow Simulator For Evaluation of Traffic Signal Control. *Proceedings of the Second World Congress on Intelligent Transport Systems*, 9-11 November 1995, Yokohama, Japan, pp. 1920-1925.

55. Yang, Q. and H.N. Koutsopoulos. A Microscopic Traffic Simulator for Evaluation of Dynamic Traffic Management Systems. *Transportation Research Part C*, Vol. 4C, June 1995, pp. 113-129

56. Lovell, D. and W. Levine. The Freeway Access Control Problem - A Survey of Success and Continuing Challenges. *ITSC 2001, IEEE Intelligent Transportation Systems Proceedings*, 25-29 August 2001, Oakland, CA, pp. 542-547.

57. Lovell, D. and C.F. Daganzo. Access control on networks with unique origin-destination paths. *Transportation Research B*, Vol. 34, 2000, pp. 185-202.

58. Akahane, H. and M. Koshi. Short-Term Predictions of Inflow Volumes of Urban Freeways. *Second International Conference on Road Traffic Control IEE*, London, UK, 1986, pp. 35-88.

59. Change, G.L. and J. Wu. Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors. *Transportation Research B*, Vol. 28, No. 2, 1994, pp. 141-160.

60. Muysenberg, H.P.H., R.A. Bauer, and E.G.J. Peters. Advanced Control of Glass Tanks Using Simulation Models and Fuzzy Control. *Ceramic Engineering and Science Proceedings*, Vol. 19, No. 1, 1998, pp. 127-135.

61. Stanciulescu, F. Simulation and control of environmental systems using a mathematical-heuristic model and algorithm. An application to the Danube Delta system. *Environmental Modelling & Software*, Vol. 12, Nos. 2-3, 1997, pp. 211-218.

62. Hooke, R. and T.A. Jeeves. 'Direct Search' Solution of numerical and statistical problems. *Journal of the ACM*, Vol. 8. 1961, pp. 212-229.

63. Haberman, R. *Mathematical Models: Mechanical Vibrations, Population Dynamics and Traffic Flow*. New Jersey, Prentice-Hall, 1977.

64. Cloutier, J. State-Dependant Riccati Equation Techniques: An Overview. *Proceedings of the American Control Conference*. 4-6 June 1997, Albuquerque, NM, Vol. 2, pp. 932-936.

65. Hayase, M., T. Yamazaki, and E. Rijanto. Nonlinear optimal control: principle of local optimality. *Proceedings of the IEEE International Conference on Industrial Technology*, 19-22 January 2000, Goa, India, Vol. 2, pp. 202-205.

66. Xin, M., S.N. Balakrishnan, and Z. Huang. Robust State Dependant Riccati Equation Based Robot Manipulator Control. *Proceedings of the 2001 IEEE International Conference on Control Applications*, 5-7 September 2001, Mexico City, Mexico, pp. 369-374.

67. Erdem, E.B. and A.G. Alleyne. Experimental real-time SDRE control of an underactuated robot. *Proceedings of the 40th IEEE Conference on Decision and Control*, 4-7 December 2001, Orlando, FL, pp. 2986-2991.

## APPENDIX A: NOTATION

### *Indices*

i – On ramp

j – Off ramp

k – Time step

*K* – total number of time steps in the data sample

m – Car

### *General Notation*

D(t) – Actual trip demand rate

Q(t) – Actual uncontrolled flow of vehicles into the freeway

U(t) – Actual controlled flow of vehicles into the freeway

VW(t) – Total number of vehicles waiting to use the freeway

Z(k) – Sensor information collected from the freeway

x(k) – Microscopic state information (general automata cells)

x*(k) – Statistical state information extracted from the microscopic model

$D_{ij}(k)$ – Estimated O-D demand rate

$u_i$ – Steady state ramp metering rate at ramp i

$u – [u_1, u_1, u_2, u_3, u_4, u_5, \ldots, u_N]$

$\rho_i$ – Steady state local density at ramp i

$\rho – [\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6, \ldots, \rho_N]$

$\check{u}_i(k)$ – Actual ramp metering rate at ramp i

$\check{u}(k) – [\check{u}_1(k), \check{u}_2(k), \check{u}_3(k), \check{u}_4(k), \check{u}_5(k), \check{u}_6(k), \ldots, \check{u}_N(k)]$

$A_{ij}(k)$ – Arrival rate of vehicles to the entrance of the freeway indexed by trip

$A_i(k)$ – Arrival rate of vehicles to the entrance of the freeway indexed by on-ramp

$\alpha_{ij}(k)$- Percentage of vehicles entering at location i that exit at location j

$S_j(k)$- Off-ramp flow at location j

$\Theta_{ij}(k\text{-}l,k)$ – The fraction of vehicles that exit via ramp j during time k that departed from ramp i at time k-l.

***Global Controller Notation***

G(u) – A function that calculates statistics on the freeway using the steady state control vector u.

M(x(k)) – A function that calculates statistics on the freeway over the last time step

N - The total number of ramps on the road

NV - The total number of vehicles on the road

$vv_m(k)$ - The current speed of vehicle m at time k

$vd_m(k)$ - The desired speed of vehicle m at time k

$vx_m(k)$ - The distance vehicle m has until it reaches its destination at time k

$vw_i(k)$ - The number of vehicles in queue at ramp i

$\Delta T$ – The time in-between calculations of the state of the microscopic model

T – The length of the optimization horizon

Z – Total delay experienced by drivers during time T

$f(x^*)$ – Objective function for the system-wide controller

***Local Controller Notation***

w(k) – external input to the controller

z(k) – objective function

y(k) – measured output

$\tilde{u}(k)$ – control input equal to the change in controlled flow

H – Plant model

K – Controller model

$w_{1i}(k)$ – uncontrolled flow into the control region from upstream

$w_{2i}(k)$ – uncontrolled flow into the on-ramp queue

$VW_i(k)$ – number of vehicles in the on-ramp queue

$P_i(k)$ – Density in the local region

$q_o$- flow out of the control region, later approximated using other terms

$\Delta X$ – length of the control region

$\Delta T$ – time step

$\rho_m$- maximum density in the control region

$v_f$ – free flow speed

$\breve{x}(k)$ - Internal state of the plant equal to $[P_i, VW_i, \breve{u}_i]$

$\hat{x}(k)$ - Internal state of the controller

$A_{ok}$, $B_{ok}$, $F_k$ - The solution to the $H_\infty$ control problem

$w_1$, $w_2$, $w_3$, $w_4$, $w_5$, $w_6$, $w_7$ – Weighting coefficients