**Title**

Optimization and Microsimulation of On-ramp Metering for Congested Freeways

**Permalink**

https://escholarship.org/uc/item/95k1q411

**Author**

Gomes, Gabriel C.

**Publication Date**

2004-11-01

# Optimization and Microsimulation of On-ramp Metering for Congested Freeways

**Gabriel C.Gomes**
*University of California, Berkeley*

CALIFORNIA PARTNERS FOR ADVANCED TRANSIT AND HIGHWAYS

# Optimization and Microsimulation of On-ramp Metering for Congested Freeways

Gabriel C. Gomes

# Abstract

This dissertation investigates various aspect of the design and testing of on-ramp metering control systems, including optimization-based control and microscopic freeway modeling. A new technique for generating optimal metering plans is developed. As with most predictive designs, the ramp metering rates are found as the solution to a nonlinear optimization problem. In contrast to previous designs, the new approach 1) produces a globally optimal solution to the nonlinear problem, 2) requires only to solve a single linear program, and 3) allows the enforcement of hard constraints on the on-ramp queue lengths. The price that is paid for these features is that the objective function being minimized is not Total Travel Time, but rather a member of a class of "TTT-like" objective functions. A TTT-like objective function is defined as a linear combination of mainline flows with weights that, similarly to the Total Travel Time cost weights, decrease linearly in time from some initial value to zero at the final time. An example application of the technique shows that the globally optimal metering plan with respect to a TTT-like objective function also performs well in terms of Total Travel Time. A macroscopic analysis of local traffic-responsive ramp metering on a short stretch of freeway, with a single on-ramp and no offramps, is also presented. The study compares the performance of two popular local traffic-responsive ramp metering algorithms: Alinea and Percent-Occupancy, and addresses issues pertaining to parameter tuning and loop-detector placement. The second half of the dissertation describes the construction of a detailed microsimulation model of a stretch of Interstate 210 in Pasadena, CA. The VISSIM microsimulation package was used to create this model. Descriptions of the data gathering and processing procedures, bottleneck identification, network coding, and model calibration are provided. The model is used to test the performance of candidate local traffic-responsive controllers. Questions concerning the relative merits of these controllers, parameter tuning, and loop-detector placement are addressed in the context of the large-scale microscopic model.

# Acknowledgements

*To my family: Ma, Dad, Kina, my Yet-Unnamed-Little-One and the Mother-to-Be*

# Contents

# Chapter 1

# Introduction

Recurring daily congestion is a common feature of many urban freeways. Residents of large metropolitan areas around the world can attest to the frustration of the ever-deteriorating daily commute. It is a complex problem without an obvious solution. The phenomenon of urban sprawl, which continues to draw more and more people away from the urban centers, has lead to longer commute distances, and to a steady increase in the demands on the freeway system. A dilemma faced by urban planners and freeway managers is whether to increase the capacity of the system by expansion, or to increase its efficiency with operational enhancements.

Over the years, traffic engineers have devised a multitude of techniques for improving freeway operations with minimal investment. Incentives such as priority lanes for high occupancy vehicles and toll reductions are used to promote carpooling. Increased tolls during peak congestion periods are meant to induce temporal shifts in demand. Traffic surveillance with closed-circuit television and driver "hot lines" allow for prompt detection and response to incidents. Periodic freeway patrols further reduce incident response times. A variety of methods of information dissemination, such as radio traffic updates, Internet-based tools such as PeMS, dynamic message signs, and onboard displays, keep drivers informed of their best options.

Yet another tactic, and the topic of this dissertation, is on-ramp metering. The goal of on-ramp metering is to reduce the delay caused to divers by recurrent traffic jams by restricting their access to the freeway with stoplights placed on the freeway entrances. This may seem at first like a futile approach, amounting to a simple transferal of the congested area from the mainline to the on-ramps, with no reduction in delay. Numerous field experiments have demonstrated, however, that delay can be reduced with a judicious metering strategy. The properties of traffic behavior that enable these savings are outlined in Section 2.2.1.

The first experiments with on-ramp metering took place in Detroit and Chicago in the early 1960's [40]. The Chicago Expressway Surveillance Project, established in 1961, was the first comprehensive and well-documented evaluation of an automatic metering system [61]. Since then, over 2,300 ramp meters have been deployed in North America [84]. Bogenberger and May [9] give a comprehensive account of the state of the art as of 1999. Their report describes operational systems in 21 metropolitan areas across the U.S., 5 of them involving over 100 ramps, and 12 in other countries. The report illustrates how much ramp metering has progressed since its inception in the 1960's. Some strategies now incorporate internal models of the freeway and advanced feedback control techniques. Sections 2.1 and 2.2 give an

overview of the evolution of freeway models and optimization-based ramp metering strategies.

The presentation if the thesis is divided into two parts. The first, comprising Chapters 3 through 5, focuses on the theoretical derivation and simulation testing of a new technique for computing optimal metering plans. Chapter 3 introduces the Asymmetric Cell Transmission Model – the ACTM – a variation on the original cell transmission model created especially for use in the ramp metering design of Chapter 5. The main advantage of this model is in the simple concave form of its nonlinear terms. This structural simplicity comes at the expense of its scope: the ACTM is only applicable to linear freeways with on-ramps and offramps.

Chapter 4 applies the ACTM to a stretch of freeway in the vicinity of a single on-ramp. The aim of this chapter is to derive specific guidelines concerning parameter selection for two well-known local traffic responsive metering strategies (Alinea and Percent-Occupancy).

Chapter 5 provides the theoretical derivation of the ACTM-based optimal ramp metering strategy. Similarly to previous predictive designs, the control plan is computed here as the solution to a nonlinear optimization problem. The distinguishing feature of the proposed design owes to the concave structure of the ACTM: the nonlinear optimization problem can be replaced with a linear program, and can therefore be solved globally and efficiently with any modern LP solver. This represents a tangible improvement over previous designs that attempted to solve their respective nonlinear programs with local methods. Also in contrast to many other designs, the proposed strategy includes explicit bounds on the lengths of the on-ramp queues. It is contended that in many cases a scarcity of on-ramp storage space in the system may severely limit the ability of any metering strategy to improve performance. In these situations, the amount and distribution of the scarce resource becomes an important factor, which should be considered in the control design.

A few limitations to the approach, aside from the possible imprecisions of the ACTM, should be mentioned. The method is limited in its ability to find a true global optimizer of Total Travel Time by three factors. The first is that the objective function used in the problem is not Total Travel Time, but a closely related Generalized Total Travel Time. These two objective functions are qualitatively similar, in the sense that they both strive to move vehicles down the freeway as quickly as possible. It is shown with an example that the optimal plan produces a substantial reduction in both measures. Second, the method is not applicable to scenarios that are so congested that the mainline queue backs up into some on-ramp, even when the freeway is optimally controlled. Fortunately, this situation is not common. Third, the optimal plan may require some on-ramps to be metered at rates below the minimum allowable rate (typically 180 vph to 240 vph). It may even call for some ramps to be temporarily closed. In this case, the sub-minimum values in the optimal plan must be replaced with the minimum allowed rate. The example in Section 5.5 demonstrates that this step produces only a small sacrifice of global optimality.

The second part of the presentation, Chapters 6 through 8, describes the use of a microscopic traffic simulator for testing local ramp metering strategies on an existing freeway. Such a study might be commissioned as a prerequisite for installing or altering a particular freeway control system. The microscopic model allows the designer to test several features that may be difficult (although perhaps not impossible) to test macroscopically. VISSIM was chosen as the microscopic simulator for this study.

Chapter 6 describes the stage of the evaluation common to both macroscopic and micro-

scopic modeling: data collection. The study focuses on the morning commute on a congested stretch of Interstate 210 in southern California. The geometric, traffic, and hardware characteristics of the site were gathered from several sources with the help of the Caltrans D07 staff. The calibration phase of the study is described in Chapter 7. It was found that the VISSIM software is capable of reproducing all of the features of the freeway that were considered important to the evaluation of ramp metering. These included an HOV priority lane, bypasses on some of the on-ramps for HOV vehicles, a variety of vehicle types with a significant percentage of trucks, and most importantly, the control hardware and the intricate control logic contained in the field controller chip. VISSIM's C-based programming module enabled us to encode the particular control and override methods used on I-210. Chapter 8 is devoted to simulation experiments using the VISSIM model and either fixed-rate, Percent-Occupancy, or Alinea control.

# Chapter 2

# Literature review

This chapter provides an overview of the two areas within the field of Traffic Engineering that are of greater relevance to the propositions of this dissertation. Section 2.1 focuses on the evolution of macroscopic models for freeways, with emphasis on the first-order LWR model, and the discrete cell transmission model. Section 2.2 contains discussions on the motivations for and limitations to on-ramp metering. These discussions are referred to later in the development of the optimal strategy in Chapter 5. To give context to the proposed optimal design, a number of previous predictive (static and dynamic) schemes are described in Section 2.2.3.

## 2.1  Macroscopic freeway modeling

In the mathematical study of traffic dynamics, two fundamentally different approaches can be taken. The *microscopic* approach seeks to reproduce the behavior of the individual driver/vehicle unit, as it responds to its environment by adjusting its speeds. Microscopic models typically involve variables such as vehicle position, speed, and headway. The *macroscopic* approach ignores the dynamics of the individual driver and instead attempts to replicate the aggregate response of a large number of vehicles. These models represent traffic as a compressible fluid, in terms of flow, density, and average speed. The field of Traffic Engineering has benefitted immensely from macroscopic models; they are widely used in the design of freeway facilities (e.g. FREQ), and they form the basis of nearly all model-based on-ramp metering designs. In this section we will define some fundamental quantities and review the evolution of macroscopic models of freeway traffic.

The Highway Capacity Manual 2000 [8] provides the following definitions. Symbols $x$ and $t$ represent position (measured in the direction of traffic flow) and time.

**Speed** $v(x,t)$ is defined as a rate of motion expressed as distance per unit of time. Depending on how it is measured, it is referred to as either *space mean speed* or *time mean speed*. Space mean speed is computed by dividing a length of freeway by the average time it takes for vehicles to traverse it. Time mean speed is the average speed of vehicles observed passing a given point. The latter is easier to measure in the field, as it can be obtained directly from conventional sensing devices. *Free-flow speed* is the average speed of traffic measured under conditions of low volume, when vehicles can move freely at their desired speed.

**Flow** $f(x,t)$ is the total number of vehicles that pass the point $x$ during a given time interval containing $t$, divided by the length of the time interval. It is usually expressed as an hourly rate, and is easily measured with road sensors.

**Density** $\rho(x,t)$ is the number of vehicles occupying a length of freeway about point $x$ at instant $t$. Its measurement is difficult because it requires the observation of a stretch of road. Instead, it is often approximated from measurements of flow and speed with the following formula:

$$\rho(x,t) = \frac{f(x,t)}{v(x,t)} \tag{2.1}$$

**Demand** is the number of vehicle occupants or drivers (usually expressed as the number of vehicles) who desire to use a given facility during a specified time period.

**Capacity** is the maximum hourly rate at which persons or vehicles reasonably can be expected to traverse a point or a uniform section of a lane or roadway during a given time period under prevailing roadway, traffic, and control conditions.

**Bottleneck** is defined as any road element where demand exceeds capacity. Freeway bottlenecks sometimes appear near heavy on-ramps, where a localized increase in demand is combined with a decrease in capacity due to lane changing.

One of the first, if not the first attempt to correlate freeway speed, density, and flow was by Greenshields in 1934 [33]. He used photographic images to estimate aggregate vehicular speeds and densities on a straight two-lane roadway, and found that they could be reasonably well approximated by a straight line. Using Eq. (2.1) he derived a parabolic relationship between flow and density, as shown in Figure 2.1.
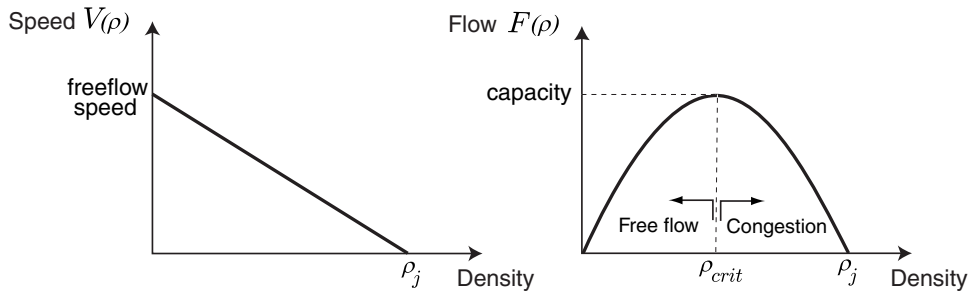


Figure 2.1: Greenshields' speed and flow relationships.

The function $f = F(\rho)$ is commonly known as the *fundamental diagram*. Many later researchers have suggested alternative shapes that provide a better fit to the measured data. Some examples are shown in Figure 2.2. These all share the following characteristics:

1. $F(0) = F(\rho_j) = 0$, where $\rho_j$ is the *jam density*.

2. The continuous portions of $F(\rho)$ are concave.

3. A critical density $\rho_{crit}$ can be define where the maximum flow is attained. Then, $F'(\rho) \geq 0$ for $\rho < \rho_{crit}$ and $F'(\rho) \leq 0$ for $\rho > \rho_{crit}$.[1]

---

[1]$F'(\cdot)$ denotes the slope of $F(\cdot)$.
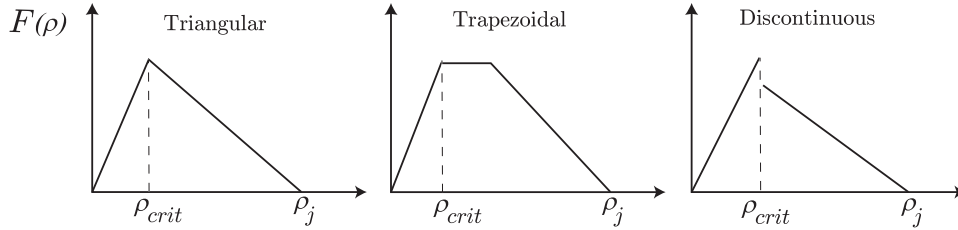
Figure 2.2: Alternative shapes for the $F(\rho)$ function.

The critical density $\rho_{crit}$ separates the fundamental diagram into two regimes: *free-flow* ($\rho < \rho_{crit}$) and *congestion* ($\rho > \rho_{crit}$) (see Figure 2.1). Measurements on the free-flow side are usually well represented by a straight line, whereas measurements in congestion tend to be more scattered. Some authors claim that there is a difference in the maximum measured flow $F(\rho_{crit})$, depending on whether the freeway is in free-flow or congestion, and contend that a *discontinuity* exists at $\rho = \rho_{crit}$ (as in Figure 2.2). This is described in [2, 13, 36] as a *capacity drop*, on the order of 4-10% in peak flow, as the freeway transitions into congestion. Hall et al. [37] warn however that, although a noticeable drop may exist, it can only be measured *at a bottleneck*, since it is masked at other locations by shock waves. The issue of capacity drop is important to on-ramp metering, and will be revisited in Section 2.2.1.

Lighthill and Whitham in 1955 [52] were the first to pose a macroscopic dynamic model of traffic using Greenshields' hypothesis of a static flow/density relationship. However they did not restrict their model to parabolas, but more generally to *continuous concave* functions with $F(\rho_j) = F(0) = 0$. The model consists of the following two equations:

$$\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial f(x,t)}{\partial x} = 0 \tag{2.2}$$

$$f(x,t) = F(\rho(x,t)) \tag{2.3}$$

Equation (2.2) is the principle of conservation of mass, or in this case conservation of vehicles, from fluid mechanics. These equations can be written more compactly as:

$$\frac{\partial \rho(x,t)}{\partial t} + F'(\rho(x,t))\frac{\partial \rho(x,t)}{\partial x} = 0 \tag{2.4}$$

This PDE had been used prior to 1955 to describe the movement of water in rivers. Lighthill and Whitham's main contribution was a complete characterization of its solutions, as well as a simple and illustrative graphical method. Richards [88] independently posed a similar model in 1956, although he assumed Greenshields' parabolic form, and devised a more limited graphical method. Equation (2.4) is nevertheless commonly known as the Lighthill-Whitham-Richards, or LWR, model.

## 2.1.1 The LWR model

The LWR model is an example of a nonlinear wave equation whose solutions can be described in terms of *characteristics*, or trajectories in the time/space plane along which the densities and flows are described by ordinary differential equations. Equation (2.4) is a relatively

simple wave equation: its characteristics are straight lines with slope $F'(\rho)$, and the densities and flows remain *constant* along the characteristics. The value of the density or flow at any point is therefore equal to a boundary or initial condition to which the point is connected by a characteristic. The boundary value problem is *well-posed* if the family of characteristics emanating from the initial and boundary conditions spans the entire time/space plane. The main complication arises, as described in [52], when characteristics intersect, leading to multiple values at the point of intersection. In this situation, the PDE admits only *weak solutions*, which necessarily contain discontinuities, or *shocks*. The discontinuities are such that the more general integral form of Eq. (2.2) between arbitrary points $x_1$ and $x_2$ applies:

$$\frac{d}{dt} \int_{x_1}^{x_2} \rho(x,t)\,dx \; + \; f(x_2,t) \; - \; f(x_1,t) = 0 \tag{2.5}$$

The speed of the shock ($c_s$) can be found by applying Eq. (2.5) to a small window $[x_1, x_2]$ containing the moving shock at $x = s(t)$ [99] (see Figure 2.3):

$$
\begin{aligned}
f(x_1,t) \; - \; f(x_2,t) \; &= \; \frac{d}{dt} \int_{x_1}^{x_2} \rho(x,t)\,dx \tag{2.6}\\
&= \; \frac{d}{dt} \int_{x_1}^{s^-(t)} \rho(x,t)\,dx \; + \; \frac{d}{dt} \int_{s^+(t)}^{x_2} \rho(x,t)\,dx \\
&= \; \dot{s}\,\rho(s^-,t) + \int_{x_1}^{s^-(t)} \frac{\partial \rho}{\partial t}\,dx \; - \; \dot{s}\,\rho(s^+,t) \; + \; \int_{s^+(t)}^{x_2} \frac{\partial \rho}{\partial t}\,dx
\end{aligned}
$$

Taking the limit $x_1 \to s^-$ and $x_2 \to s^+$:

$$c_s(t) \; = \; \dot{s}(t) \; = \; \frac{f(s^+,t) - f(s^-,t)}{\rho(s^+,t) - \rho(s^-,t)} \; = \; \frac{f^+ - f^-}{\rho^+ - \rho^-} \tag{2.7}$$

The speed of the shock is represented graphically on the fundamental diagram as the inclination of the chord joining the upstream and downstream states, as illustrated in Figure 2.3. It is shown in [26, 99] that the concavity of $F(\rho)$ implies that only shocks of increasing density, or *deceleration* shocks, can develop in the traffic stream. This is because conditions of heavy downstream traffic and light upstream traffic generate waves that coalesce, whereas conditions of smaller downstream density lead to diverging waves. Hence, all shocks that form naturally must satisfy:

$$\rho^- \leq \rho^+ \tag{2.8}$$

Denoting the speed of the characteristics that terminate in the shock with $c^- = F'(\rho^-)$ and $c^+ = F'(\rho^+)$, then Eqs. (2.7), (2.8), and the concavity of $F(\rho)$ imply:

$$c^- > c_s > c^+ \tag{2.9}$$

Ansorge [4] completed the LWR theory in 1990 with a condition of *increasing entropy*, analogous to the 2nd Law of Thermodynamics. In gas dynamics, the entropy criterion is used to select a physically relevant solution from a number of weak solutions. Ansorge found that the principle, when applied to traffic dynamics, is equivalent to Eq. (2.8); i.e. it is always

13
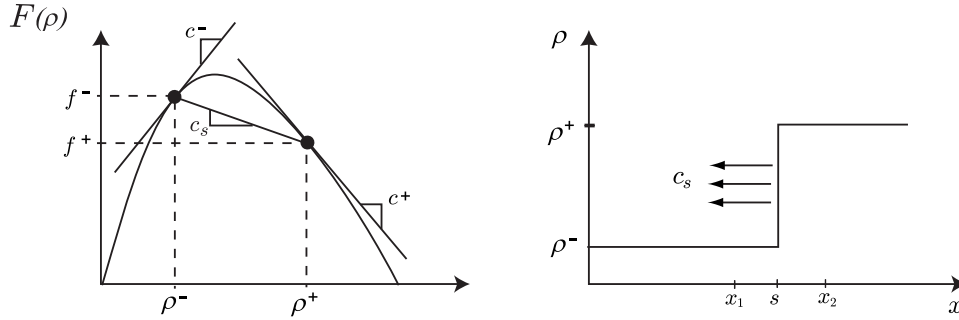
Figure 2.3: Speed of a shock.

satisfied on the continuous portions of the solution, and imposes a condition of increasing density across all discontinuities. It was found furthermore that this condition is sufficient to eliminate all of the weak solutions except for the physically relevant one. That is, given a number of weak solutions, Ansorge's criterion states that the sole realistic solution is the one with *no stable acceleration shocks*. It has already been noted that the strict concavity of the fundamental diagram precludes the *natural* formation of acceleration shocks in the traffic stream (Eq. (2.8)). These types of shocks can only be *imposed* by the boundary or initial conditions. Thus, Ansorge's criterion requires more simply that any acceleration shocks imposed by initial or boundary conditions must dissipate instantaneously[2]. This criterion is supported by observation: acceleration shocks, for example at the front of a stoplight queue when it turns green, tend to dissipate quickly.

Pipes [85] deduced an illustrative formula for the acceleration of a vehicle moving with the traffic stream in the LWR theory:

$$\frac{dv}{dt} = -\rho \left( V'(\rho) \right)^2 \frac{\partial \rho}{\partial x} \tag{2.10}$$

where $V(\rho) \triangleq F(\rho)/\rho$ is the equilibrium speed. Vehicle acceleration is seen in Eq. (2.10) to have the opposite sign of the density gradient; i.e. vehicles decelerate as they move into more dense regions, and vice-versa. Equation (2.10) also reveals one of the main criticisms of the theory, that vehicles undergo infinite decelerations as they traverse a shock.

Since 1955, many numerical schemes have been developed for solving the LWR equations. May [62] gives an account of the available software as of 1981. FREQ, one of the most widely used freeway simulation programs, automates the graphical technique developed by Lighthill and Whitham. It also includes several enhancements such as access ramps, weaving lanes, truck percentages, etc. Other applications have approximated the PDE with finite differences, however not always successfully. Ross notes in [90] that some of these computer programs have the undesirable tendency to "lock up", meaning that congestion persists after a blockage has been removed. In hindsight it seems that those algorithms may have been consistent with the theory, but they failed to select the entropy increasing weak solution (as in Ansorge's example [4]).

---

[2]Ansorge's proof assumes a *strictly* concave fundamental diagram; i.e. $F''(\rho) < 0$. It does not apply to fundamental diagrams, such as those in Figure 2.2, where $F''(\rho)$ sometimes equals zero. In this case, the acceleration shock may persist.

14

Daganzo's *cell transmission model* (CTM) resolved the "lock-up" problem in 1995 [21, 22]. The CTM is a finite difference approximation of LWR based on the intuitive concepts of *sending* and *receiving* flows. Under the CTM, the highway is divided into homogeneous, consecutively numbered sections of length $L_i$, where $i$ is a section index. Time is discretized into uniform intervals of duration $\Delta t$, such that

$$v\,\Delta t \;\leq\; \min_i L_i \tag{2.11}$$

where $v$ is the free-flow speed. The state variable for section $i$ is $n_i(k)$, the number of vehicles it contains at time $k\Delta t$. Equations (2.12) and (2.13) below are discrete forms of the conservation equation and fundamental diagram respectively:

$$n_i(k+1) \;=\; n_i(k) \;+\; y_i(k) \;-\; y_{i+1}(k) \tag{2.12}$$
$$y_i(k) \;=\; \min\{\,n_{i-1}(k)\;;\;Q_i\;;\;(w/v)[N_i - n_i(k)]\,\} \tag{2.13}$$

$v$ and $w$ are the slopes of the free-flow and congested portions of a triangular fundamental diagram. They are assumed to be uniform along the freeway in [21, 22]. $Q_i$ is the maximum number of vehicles that can move between sections $i-1$ and $i$ during one time interval (related to capacity); $N_i$ is the maximum number of vehicles that fit in section $i$ (related to jam density); $y_i(k)$ is the number of vehicles that actually move between sections $i-1$ and $i$ during time interval $k$. The flow variable $y_i(k)$ is computed with Eq. (2.13) as the minimum of what can be *sent* by the upstream section under free-flow conditions, $n_{i-1}(k)$, and what can be *received* by the downstream section during congestion, $(w/v)[N_i - n_i(k)]$. Daganzo showed in [23] that the CTM is *consistent* with the LWR model in the sense that their solutions converge as $\Delta t$ and $L_i$ become small. The model also has the important property that it dissipates acceleration shock waves. That is, it finds the entropy increasing solution.
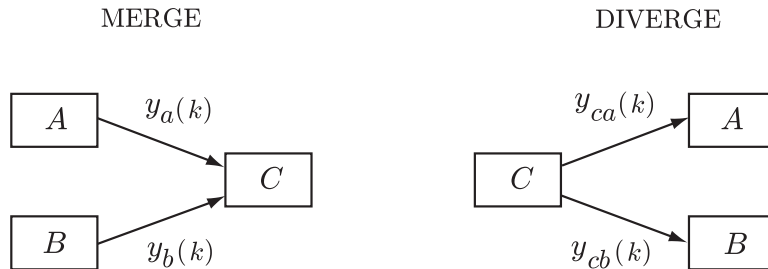


Figure 2.4: Merging and diverging flows in the CTM.

The CTM was extended in [22] to cover more general *network* topologies, including traffic merges and diverges (see Figure 2.4). In the case of merges, the flows from each of the upstream sections are computed with:

$$y_a(k) \;=\; \mathrm{mid}\{\,S_a\;;\;R_c - S_b\;;\;p\,R_c\,\} \tag{2.14}$$
$$y_b(k) \;=\; \mathrm{mid}\{\,S_b\;;\;R_c - S_a\;;\;(1-p)R_c\,\} \tag{2.15}$$

15

where

$$S_a = \min\{ Q_a \; ; \; n_a(k) \} \tag{2.16}$$

$$S_b = \min\{ Q_b \; ; \; n_b(k) \} \tag{2.17}$$

$$R_c = \min\{ Q_c \; ; \; (w/v)[N_c - n_c(k)] \} \tag{2.18}$$

The mid{} function in Eqs. (2.14) and (2.15) returns the middle value, or the median, of its three arguments. $S_i$ and $R_i$ are respectively the number of vehicles that a section $i$ can *send* or *receive*. $p$ is a given number between 0 and 1 that determines the priority assigned to flows coming from sections $A$ or $B$ at the junction. These formulas ensure that the density in section $C$ will never exceed $N_c$.

Bifurcating flows in the network CTM are computed using a given *split ratio* $\beta_{ca}(k) \in [0, 1]$, which determines the portion of the total flow leaving section $C$ that continues into section $A$, during time interval $k$ [3]. The portion that diverts to section $B$ is $1 - \beta_{ca}(k)$. Using the known split ratios, the total flow out of section $C$, $y_c(k) = y_{ca}(k) + y_{cb}(k)$, is given by:

$$y_c(k) = \min\{ S_c \; , \; R_a/\beta_{ca}(k) \; , \; R_b/(1 - \beta_{ca}(k)) \} \tag{2.19}$$

Then the individual flows can be found with:

$$y_{ca}(k) = \beta_{ca}(k) \, y_c(k) \quad \text{and} \quad y_{cb}(k) = (1 - \beta_{ca}(k)) \, y_c(k) \tag{2.20}$$

The CTM was implemented in the NETCELL simulation software [18, 54], which was used to validate the model using data from I-880 in Oakland, California in [53]. Tan [94] calibrated the CTM using data from I-210 in Pasadena, California.

Another important contribution to the LWR model is Newell's "simplified theory", covered in [69, 70, 71]. This solution method is based on the concept of the *cumulative vehicle count*, $N(x, t)$, defined as "the number of vehicles to pass some location $x$ by time $t$ starting from the passage of some reference vehicle". The usage of $N(x, t)$ instead of $f(x, t)$ and $\rho(x, t)$ as the computed variable has practical justifications. First, several quantities of interest to transportation analysts are readily extracted from $N(x, t)$ curves. For example, the trajectory of a vehicle $m$ is obtained by intersecting the surface $N(x, t)$ with the horizontal plane $N = m$. The plots of Figure 2.5 were constructed by projecting vertical slices of $N(x, t)$ at $x = x_1$ and $x = x_2$ onto the $N/t$ plane. Both the trip time for a vehicle $m$ between points $x_1$ and $x_2$ and the total number of vehicles between $x_1$ and $x_2$ at time $t_1$ can be directly measured from these curves. Also, the total time spent by all vehicles between $x_1$ and $x_2$ during a time interval $[t_1, t_2]$ appears as the shaded area.
Density and flow also have direct interpretations as the partial derivatives of $N(x, t)$:

$$\rho(x, t) = -\frac{\partial N(x, t)}{\partial x} \quad \text{and} \quad f(x, t) = \frac{\partial N(x, t)}{\partial t} \tag{2.21}$$

Another important advantage of stating the LWR equations in terms of cumulative counts, is that the conservation of vehicles turns into a condition of *continuity* of the $N(x, t)$ surface. Substituting Eq. (2.21) into (2.2):

$$\frac{\partial}{\partial t}\left(-\frac{\partial N(x, t)}{\partial x}\right) + \frac{\partial}{\partial x}\left(\frac{\partial N(x, t)}{\partial t}\right) = 0 \tag{2.22}$$

---

[3]A second version of the network CTM assuming known OD information was also given in [22].
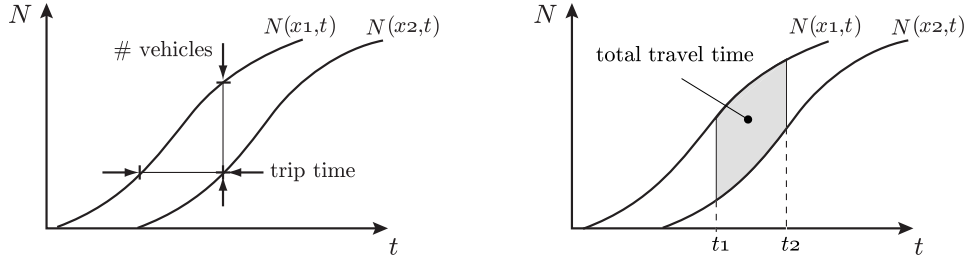
Figure 2.5: Uses of the cumulative vehicle count.

which is trivially satisfied whenever the derivatives exist. Shocks are manifested in the simplified theory as creases in $N(x, t)$, or locations where the second derivatives do not exist.

Newell's graphical method applies to uniform freeways, where the fundamental diagram does not change with $x$. In this case, the cumulative count behaves linearly along characteristics:

$$dN = \underbrace{\left[ -\rho + f/F'(\rho) \right]}_{\text{constant}} dx = \underbrace{\left[ -\rho \, F'(\rho) + f \right]}_{\text{constant}} dt \qquad (2.23)$$

since $\rho$ and $f$ remain constant along the characteristic. Thus, the value of $N(x, t)$ at a particular point can be easily obtained from the value at any other point on the same characteristic (e.g. an initial or boundary point), without evaluating intermediate points. Newell describes several situations in which this may be advantageous. A computer implementation of the method is described in [93].

Despite its simplicity, the LWR model is capable of reproducing many important phenomena of freeway traffic. Notably, it captures the main difference between free-flow and congested traffic, which is that they propagate small disturbances in opposite directions and at different speeds. It also explains the formation and dissipation of queues upstream of a bottleneck, the dynamics of deceleration shock waves, and the absence of naturally forming acceleration shock waves. As opposed to many higher-order models, the LWR theory never predicts negative speeds or flows. However, several authors have noted that some of its predictions are unrealistic, and that it fails to capture other important aspects of traffic behavior. Below follows a list of cited shortcomings.

1. The central hypothesis of the theory – that flow and density are related by a static function $F(\rho)$ – fails in the congested regime [26, 68, 78, 90].

2. Time sequences of field measurements in the density/flow plane seem to follow different $F(\rho)$ curves, depending on whether the traffic stream is decelerating or accelerating. Treiterer and Myers [95] suggest that this is indicative of *hysteresis* in freeway traffic behavior, caused by the fact that vehicle deceleration and acceleration are not symmetric processes. Hysteresis cannot be replicated by LWR [107, 108].

3. The model predicts instantaneous changes in vehicular speeds as they traverse a shock [24, 26, 55, 107].

17

4. The model does not consider the distribution of driver behaviors and desired speeds. Daganzo suggests in [25] that this may lead to errors in the free-flow regime, when significant passing occurs, but that this problem is not likely to be resolved by any macroscopic model.

5. Because of the requirement of continuity made on the fundamental diagram, the model cannot incorporate any abrupt losses in capacity due to congestion [74].

Related to this last deficiency, Papageorgiou argues in [74] that it represents a serious limitation to the usefulness of LWR in the design of on-ramp control strategies. This is because the capacity gained by preserving free-flow conditions is a motivating factor for delaying the onset of congestion by metering the on-ramps (Section 2.2.1). Metering strategies based on the LWR model, such as the optimal policy described in Chapter 5, will therefore be most effective when they are applied to freeways where the *capacity drop* is less prominent, and congestion has other detrimental effects, such as the obstruction of upstream offramps.

Lighthill and Whitham were aware of the limitations stemming from their central hypothesis. They suggested as a possible improvement the addition of higher order terms to capture the effects of *diffusion*, caused by drivers' reaction to perceived changes in density, and *inertia*, caused by the fact that vehicles do not reach their desired speeds instantaneously. Several higher order models, beginning with Payne's in 1971, were later proposed to deal with some of the limitations of the LWR theory.

## 2.1.2 Higher order models

The second and third order PDE models described in this section all retain the principle of vehicle conservation from the LWR model (Eq. (2.2)), but replace the static fundamental diagram with a dynamic *momentum* equation of the form:

$$\underbrace{\frac{\partial v}{\partial t} + \zeta(v)\frac{\partial v}{\partial x}}_{\text{vehicle acceleration}} = \underbrace{\frac{1}{T(\rho)}\left[V(\rho) - v\right]}_{\text{relaxation}} + \underbrace{h(\rho)\frac{\partial \rho}{\partial x}}_{\text{anticipation}} + \underbrace{\nu\frac{\partial^2 v}{\partial x^2}}_{\text{viscosity}} \qquad (2.24)$$

The left hand side of this equation is the total time derivative of speed (vehicle acceleration) whenever $\zeta = v$. The right hand side is composed of three terms representing the reasons why drivers adjust their speed. The first term, speed *relaxation*, accounts for the fact that drivers have a desired speed $V(\rho)$, which they converge to smoothly. The coefficient $T(\rho)$ is often referred to as the *driver reaction time*, due to Payne's original derivation from a microscopic car-following model. It can also be interpreted as a macroscopic *relaxation time*. Del Castillo et al. note in [26] that this double interpretation of $T(\rho)$ has lead to a wide range of values in different applications. The second term represents the tendency of drivers to *anticipate* downstream conditions by accelerating when they are less dense, and decelerating when they are more dense. The third is a *viscosity* term introduced by Kühne in [50] to smooth discontinuous shocks.

Payne [80] developed the first and still the most widely used second order model by substituting quantities from a microscopic model with analogous macroscopic variables. His

momentum equation is:

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \frac{1}{T}[\,V(\rho) - v\,] + \left(\frac{V'(\rho)}{2T\rho}\right)\frac{\partial \rho}{\partial x} \tag{2.25}$$

Payne showed that this PDE is stable to small perturbations about a uniform equilibrium $\rho = \rho_o$ if and only if:

$$-2\,T\,\rho_o^2\,V'(\rho_o) < 1 \tag{2.26}$$

Equation (2.26) defines a critical density which Kühne [50] estimates to be approximately 35 veh/km/lane. Payne also proposed a discrete version of the model which was later implemented as the FREFLO software [81]. The METANET model [46] extends Payne's model to network configurations. It also introduces a couple changes to the discrete version, which improve its numerical properties. For example, the $1/\rho$ in the anticipation term of Eq. (2.25) is replaced in METANET with $1/(\rho + \kappa)$, with $\kappa > 0$, to allow it to handle situations of small or zero density [75]. Several calibration studies using FREFLO, METANET, or other similar implementations have found Payne's model to correspond well to measured freeway data [19, 20, 27, 34, 46, 75, 81, 91]. Although Eq. (2.25) includes a smoothing relaxation term, it does not eliminate all discontinuities from the solution. Del Castillo et al. showed in [26] that deceleration shocks may still form if a *steady compression wave* condition (Eq. (2.30)) is not met. This condition was derived by assuming a moving wave solution of the form:

$$\rho(x,t) = \Phi(x - v_s t) = \Phi(y) \tag{2.27}$$
$$v(x,t) = \Psi(x - v_s t) = \Psi(y) \tag{2.28}$$

where $\Phi$ and $\Psi$ are arbitrary functions. $v_s$ is the speed of the profiles given by $v_s = (\rho_2 v_2 - \rho_1 v_1)/(\rho_2 - \rho_1)$, where subscripts 1 and 2 denote the limits as $y \to -\infty$ and $y \to \infty$. Substituting these profiles into (2.2) and (2.25) leads to the following ODE:

$$\left(\frac{V'}{2T} + \frac{Q_s^2}{\Phi^2}T\right)\Phi' + \Phi(V - v_s) - Q_s = 0 \tag{2.29}$$

where $Q_s = \rho_1(v_1 - v_s) = \rho_2(v_2 - v_s)$. Smooth compression wave solutions (with $\rho_2 > \rho_1$) to Eq. (2.29) can exist only if:

$$\frac{V'(\rho)}{2T} + \frac{Q_s^2}{\rho^2}T > 0 \tag{2.30}$$

Thus, the profile $\Phi(y)$ must include a discontinuity if Eq. (2.30) is not satisfied for either $\Phi = \rho_1$ or $\Phi = \rho_2$.

Kühne's analysis of the Payne model in [50] showed that it supports "start-stop" wave solutions that were not predicted by the LWR theory, but that these special waves must contain discontinuities. To soften the solution he proposed the addition of a *dissipation* or *viscosity* term:

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \frac{1}{T}[V(\rho) - v] - \frac{C_o^2}{\rho}\frac{\partial \rho}{\partial x} + \nu\frac{\partial^2 v}{\partial x^2} \tag{2.31}$$

where $\nu$ is a positive constant and $C_o$ is akin to the speed of sound in fluids.

19

Lyrintzis et al. [60] suggested that, contrary to Payne's model, the reaction time of drivers depends on density, in such a way that traffic reacts quicker at high densities than at low densities. Their model is identical to Payne's with the following reaction time specification:

$$T(\rho) = T_o \left[ 1 + \left( \frac{\rho_{crit}}{\rho} \right)^{\theta} \right] \tag{2.32}$$

where $T_o$ and $\theta$ are positive constants, and $\rho_{crit}$ is the critical density of Figure 2.1.

The two most severe criticism of Payne's model are that it can lead to negative speeds and flows, or *wrong-way travel*, and that it allows vehicles to be influenced by conditions behind them [24]. The latter is because perturbations in the model may travel forward through the traffic stream at a higher speed than the traffic itself; that is, the speed of the fastest characteristic exceeds the free-flow speed. These two obviously unrealistic predictions do not plague the LWR theory. The properties of the fundamental diagram imply that, under LWR, vehicles never travel backwards, and the speed of information is at most equal to the speed of traffic.

Del Castillo et al. [26] proposed a fix to the problem of fast characteristics. They showed that it is removed by making the reaction time dependent on density as follows:

$$T(\rho) = -\frac{1}{2 \, \rho^2 \, V'(\rho)} \tag{2.33}$$

This specification for reaction time is a decreasing function of $\rho$ whenever the fundamental diagram is strictly concave. However, Daganzo showed in [24] that a whole class of higher order models, including Payne's with constant or density dependent reaction time, predicts wrong way travel. This was demonstrated with an example in which a stationary queue of finite length was initially stopped behind a road block. The density within the queue was $\rho_j$ and zero elsewhere. The expected behavior for this scenario is that nothing should happen unless the blockage is removed. Daganzo showed however that a class of models with diffusion terms tends to smooth the tail-end of the queue by moving vehicles backwards. Thus, the problem of wrong way travel cannot be eliminated by making the reaction time density dependent. A more fundamental departure is required.

Zhang proposed the following momentum equation in [107]:

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \frac{1}{T} [V(\rho) - v] - \rho \, (V'(\rho))^2 \, \frac{\partial \rho}{\partial x} \tag{2.34}$$

An interesting property of this model is that it reduces to LWR as $T \to 0$, as can be seen by comparison with Eq. (2.10). Zhang showed that the model does not exhibit wrong-way travel when applied to Daganzo's example, in the limit as $T$ tends to zero. It was not shown however, that negative speeds cannot result when $T$ is given a larger more realistic value. Zhang showed in [108] that this model predicts hysteresis, or stop-start waves, in congested traffic. It is referred to in Table 2.1 as "Zhang I".

In a later paper [109], Zhang gave a precise definition a class of "Payne-like" models that all contain fast characteristics. This definition included Zhang I. He went on to propose a new model ("Zhang II" in Table 2.1) which along with Liu's described below, are the

only two presented here that eradicate both wrong-way travel and fast characteristics. The momentum equation for Zhang II is:

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = -\rho V'(\rho)\frac{\partial v}{\partial x} \qquad (2.35)$$

Liu et al. [55] devised another model also devoid of wrong-way travel and fast characteristics:

$$\frac{\partial v}{\partial t} = \frac{1}{T(\rho)}[\,V(\rho) - v\,] \qquad (2.36)$$

$$T(\rho) = T_o\left[1 + \frac{E}{1 + (\rho/\rho_{crit})^\theta}\right] \qquad (2.37)$$

where $T_o$, $E$, and $\theta$ are positive constants. Notice that these last two models are more simple than the rest because they remove the anticipation and viscous terms. Wrong-way travel is eliminated with the removal of the anticipation term.

Table 2.1 summarizes the models covered in this section. Others include [83] and [90], and the efforts based on alternative techniques, such as catastrophe theory [1], neural networks [112], and heuristics [42, 43, 44].

| Model | $\zeta$ | $T(\rho)$ | $h(\rho)$ | $\nu$ |
|---|---|---|---|---|
| Payne | $v$ | $T$ | $V'(\rho)/(2T\rho)$ | 0 |
| Kühne | $v$ | $T$ | $-C_o^2/\rho$ | $\nu$ |
| Lyrintzis et al. | $v$ | $T_o\left[1 + (\rho_{crit}/\rho)^\theta\right]$ | $-\alpha/\rho$ | 0 |
| Del Castillo et al. | $v$ | $-1/(2\rho^2 V'(\rho))$ | $V'(\rho)/(2T\rho)$ | 0 |
| Zhang I | $v$ | $T$ | $\rho\,(V'(\rho))^2$ | 0 |
| Zhang II | $v + \rho\,V'(\rho)$ | $\infty$ | 0 | 0 |
| Liu et al. | 0 | $T_o\left[1 + \frac{E}{1+(\rho/\rho_{crit})^\theta}\right]$ | 0 | 0 |

Table 2.1: Higher order models.

### 2.1.3 Lane changing models

The models described so far have focused on the longitudinal behavior of vehicles. Several attempts have also been made to model the way in which drivers change lanes, and the effect that this has on density and flow. Gazis [32] proposed a proportional model in 1961, where the rate of lane changing between adjacent lanes was considered proportional to the difference between the density errors, delayed by a constant reaction time $T$:

$$Q_{1\to 2}(t) \; = \; \lambda \left[ \, \rho_2(t-T) - \rho_1(t-T) \, - \, (\rho_{2o} - \rho_{1o}) \, \right] \tag{2.38}$$

$Q_{1\to 2}(t)$ is the rate of lane changing from lane 1 to lane 2, $\lambda$ is a proportionality constant, $\rho_{io}$ is the equilibrium density in lane $i$. Gazis also assumed no external sources or sinks $(Q_{2\to 1} = -Q_{1\to 2})$ and a uniform distribution of densities in the $x$ direction. Then:

$$\dot{\rho}_1(t) \; = \; -\dot{\rho}_2(t) \; = \; Q_{1\to 2}(t) \tag{2.39}$$

From a closed-form solution to the $n$-lane case, Gazis derived stability conditions for the lane change process, depending on the values of $\lambda$, $T$, and $n$. Munjal et al. [66, 67] extended this result by coupling Gazis' model with the LWR theory. By assuming $T = 0$ in Eq. (2.38), Munjal et al. found that stable closed-form solutions were possible in certain situations, and compared the predicted densities to measured data. A few modifications to the Gazis-Munjal model were suggested by Michalopoulos et al. in [63, 64]. They reintroduced the reaction time, and made the proportionality coefficient $\lambda$ dependent on the difference between lane densities, in such a way that it vanished for $|\rho_1 - \rho_2| < k_a$ (arguing that lane changing is rare under near-uniform conditions). These added complexities put closed-form solutions out of reach, and a numerical scheme was proposed instead. Unfortunately, due to the difficulties of collecting lane change data, none of these models have been comprehensively validated.

## 2.2 Freeway on-ramp control

### 2.2.1 The rationale for on-ramp metering

The goal of any on-ramp control system must be, in general terms, to improve the efficiency of the system by regulating the number of vehicles allowed to enter the freeway. How efficiency can be increased by metering the on-ramps and how this improvement is measured are the topics of this section.

Several measures of performance have been used to assess the effectiveness of traffic controllers, however most are variations on two fundamental measures: the *Total Travel Distance* (TTD) and the *Total Travel Time* (TTT). TTD is defined as the sum of the distances travelled by all users of the system, over a given period of time. Equivalently, it is the product of the average trip length and the total number of users. It can be computed as the integral of *flow* over time and space:

$$\texttt{TTD} \; \triangleq \; \int_{x_1}^{x_2} \int_{t_1}^{t_2} f(x,t) \, dt \, dx \tag{2.40}$$

`TTT` is the sum of all trip times incurred by users during a given time period, or the number of users multiplied by the average trip time. It is computed as the integral of *density*:

$$\texttt{TTT} \triangleq \int_{x_1}^{x_2} \int_{t_1}^{t_2} \rho(x,t)\, dt\, dx \tag{2.41}$$

Using Eq. (2.2) and integration by parts, `TTT` can be separated into a control independent term, involving only the initial conditions and input flows, and a weighted sum of exit flows [76]:

$$\texttt{TTT} = \underbrace{(t_2 - t_1) \int_{x_1}^{x_2} \rho(x,t_1)dx + \int_{t_1}^{t_2} (t_2 - t)\,\mathrm{In}(t)\, dt}_{\text{control independent}} - \underbrace{\int_{t_1}^{t_2} (t_2 - t)\,\mathrm{Out}(t)\, dt}_{\text{exit flows}} \tag{2.42}$$

Here, $\mathrm{In}(t)$ is the sum of all flows entering the system. The scope of the system is assumed to be large enough so that the incoming flows are independent of the control. The first two terms on the right hand side are therefore insensitive to on-ramp control. The third term is a weighted integral of $\mathrm{Out}(t)$, the flows leaving the system. The weights $-(t_2 - t)$ are linearly increasing and negative on $[t_1, t_2]$. Equation (2.42) demonstrates the intuitive notion that to minimize `TTT`, the control strategy must maximize the outgoing flows. That is, it must remove vehicles from the system as quickly as possible.

Ideally, the spatial boundaries of the system should be chosen such that the entering flows are not affected by the state of traffic inside it, and the flows exiting are not affected by traffic outside. A necessary condition for this to happen is that both the upstream and downstream boundaries must remain in free-flow throughout the test period. For a freeway this means that the scope of the test section should encompass any queues generated by internal bottlenecks, as well as the bottlenecks themselves and some stretch of freeway downstream. The downstream boundaries should not be reached by queues from external bottlenecks.

Although it is needed, a state of free-flow at the upstream boundaries is not sufficient to guarantee that incoming flows will not be affected by the control. Drivers often make on-the-spot routing decisions based on their assessment of travel time and congestion. They can make the decision to enter or leave the freeway at different points, or they may avoid the freeway altogether and use only the surface streets. Thus, incoming flows can be affected by traffic *diversion*. The scope of the system for which control is being evaluated should therefore also encompass any alternative routes to which drivers may divert.

Yet another condition relates to the selection of the *temporal* boundaries for the experiment. In order to capture the full effect of the control strategy, the test section should be left in an uncongested state at the end of the test period. This is to ensure that the performance measure represents a fair account of all of the system users.

Finally, to compare the performance of competing control strategies, the demands on the system should be the same in each of the experiments. This of course can be accomplished only approximately in actual field tests, but it is important nevertheless, since the day-to-day variations can be on the same order as the expected savings due to ramp metering. To minimize the differences in demand, one might restrict comparisons to experiments carried out on the same day of the week, with similar weather conditions, and without incidents.

Assuming that the experiments meet the four above-mentioned specifications, it can be argued that TTT, and not TTD, is the meaningful measure of performance. The total number of users is the same for all experiments, since the demands are the same, and the time period is long enough to satisfy all of the demand. TTD and TTT are therefore proportional to the average trip length and the average trip time, respectively. In the case of a system without routing alternatives, the trip length is fixed for all users, so the TTD is the same for all experiments, regardless of the control measure. TTT is clearly the sole significant performance measure in this case.

In the case that drivers have more than one routing option, it can be expected that they will choose the route that minimizes their individual travel times. This was postulated by J.G. Wardrop in [97], and is known as "Wardrop's first principle". It implies that individual users consider travel time savings, and not travel distance, as the determining factor – it is the individual reward which they seek to maximize. It also implies that the route choice of some drivers may have a negative impact on others. For example, a relatively small number of drivers with short trips may decide to use the freeway, even though their decision to do so produces congestion. This congestion may, in turn, cause a delay to a large number of drivers with longer trips, and this induced delay may exceed the travel time savings of the short trip population. This uncontrolled situation results in larger average travel times, and more variation in the individual travel times. The function of the centralized controller should then be to provide a global perspective for achieving a more efficient use of the infrastructure, in order to increase the average reward for all users. Because the individual reward is travel time savings, the conclusion is that the controller should seek to minimize the average travel time – and therefore the total travel time.

But, can on-ramp metering alone reduce total travel time? Equation (2.42) shows that this goal is achieved by keeping the exit flows high. There are three mechanisms by which metering the on-ramps can increase exit flows. We will refer to these as *capacity drop*, *exit blockage*, and *diversion*. [4]

The first, capacity drop, is based on the observation that the maximum sustainable flow during congestion, known as the *queue discharge rate*, is sometimes less than the true capacity, which can be realized only in free-flow conditions. The onset of congestion is characterized by a sharp decrease in flow from capacity flow to the queue discharge rate, measured downstream of the bottleneck. This phenomenon has been reported in several field studies, where measurements appear to be best represented by a discontinuous fundamental diagram, as in Figure 2.2. The implication of this observation for on-ramp metering is that *congestion on the freeway should always be avoided*, no matter how long the required on-ramp queueing times, since it leads directly to a decrease in mainline, and therefore offramp flows.

Congestion has several causes related to on-ramp flow. The most obvious is when the demand exceeds the nominal capacity of the freeway. Congestion caused by *excessive demand* can be avoided simply by keeping the surplus on the ramps. Cassidy et al. observe in [12] that congestion is sometimes triggered by short *surges* in on-ramp flow. This type of congestion can be avoided with on-ramp metering by attenuating the spikes in demand. Congestion can also be caused by inefficient merging of the on-ramp and mainline streams

---

[4] Banks [7] identifies the reduction of accidents as a fourth mechanism. This and others, such as the inducement of temporal shifts in demand, are not considered here.

[14]. A class of metering strategies known as "gap-acceptance" strategies aims at improving merging efficiency by synchronizing the release of vehicles from the on-ramp with vehicles on the freeway to produce a "zipper effect".

However there is no consensus among traffic analysts on the capacity drop. Hall [37] noted that whether or not an abrupt drop exists could not be concluded from some of the existing studies, because the data was collected upstream of the bottleneck. The observed gap in the flow-density data could therefore be explained by the passage of a shock wave. Yet other studies have found a modest but non-negligible drop using downstream data: Hall and Agyemang-Duah [36], Agyemang-Duah and Hall [2], and Cassidy and Bertini [13] all found capacity drops ranging from 4% to 10%. These three studies used data from a section of the Queen Elizabeth Way in Ontario, Canada. On the other hand, Banks [5] found no significant change in average flow at several sites. In conclusion, the capacity drop does not appear to be a universal feature of freeway traffic.

A second mechanism by which on-ramp flow can influence travel time is by avoiding the blockage of the freeway offramps [10, 11]. Left unchecked, mainline queues may grow past the offramps located upstream of the bottleneck, and thus prevent vehicles from leaving the freeway. The exiting flow can be increased by restricting on-ramp flow so that the mainline queue does not overrun the upstream offramps. For this to be effective, it is not necessary, as with the capacity drop, to eliminate congestion completely, but only to reduce it sufficiently. In fact, further restricting the on-ramps will induce unnecessary delays. The offramp blockage mechanism therefore warrants for a measured amount of congestion on the freeway.

The two mechanisms mentioned so far, capacity drop and offramp blockage, lead to different strategies for minimizing total travel time. The optimal strategy for capacity drop alone, for example in a single-destination network, is to reduce on-ramp flow sufficiently to eliminate all mainline congestion, but not more. When offramp blockage is the only concern, the optimal strategy is to reduce on-ramp flow enough to unclog the offramps, but not more. The difference between these two competing strategies was noted by Banks in [7]. He found that the optimal plan for delay reduction (i.e. minimum travel time) is different from the one for minimizing congestion, and concluded that this is because delay is not minimized by *eliminating* congestion, but by unclogging the offramps. Zhang et al. studied a model without route choice in [113], and found that the freeway is best left uncontrolled (i.e. no improving controller exists) if it is either "uniformly congested" or "uniformly uncongested", meaning that the amount of congestion cannot be affected by on-ramp control alone. This seems reasonable, since the two mentioned mechanisms depend on the ability of the on-ramp controller to reduce congestion.

The third mechanism, *diversion*, becomes important when certain users have a choice of route within the system, but tend to select one that causes significant delay to others. For on-ramp metering to improve travel time in this situation, it must be able to create sufficient additional delay on the ramps for the troublesome users to switch to the system-preferred route, while decreasing the total travel time. That is, the added on-ramp queueing time should be sufficient to induce a particular diversion, but not so much as to increase total travel time. This balance may not be attainable, for example, if the difference in travel time between the user-preferred and system-preferred routes is large. It is also not possible unless

drivers have some propensity to divert when the on-ramp queue is large [111].

## 2.2.2   Limitations to on-ramp control

The previous section outlined three ways in which on-ramp metering can reduce travel time: by avoiding the capacity drop and offramp blockage, and by encouraging certain diversions. The ability of the on-ramp controller to realize these reductions, however, is limited by physical and practical considerations [6]. The two most important of these are the bounds placed on the *metering rate* and the limited *storage capacity* of the on-ramps.

The metering rate is typically restricted by a lower bound of 180 to 240 vphpl[5], and an upper bound of 900 vphpl [86, 89]. This upper bound is dictated by the one-vehicle-per-green policy, whereby every vehicle must stop at the meter and wait for a 2-second green light before proceeding. The minimum time required for a vehicle to approach the stop line is about 2 seconds, so the minimum cycle time is 4 seconds, or 900 vph. This rate can be increased if more than one vehicle is allowed entry per green light, or by increasing the number of metered lanes. The lower bound of between 180 and 240 vphpl is attributed to the fact that drivers will not wait longer than 15 to 20 seconds at the stop line.

Both the capacity drop and offramp blockage mechanisms call for a reduction of congestion. As was explained in the previous section, this can be accomplished by throttling the excess demand. Ignoring the effects of diversion, this requires that the surplus be stored upstream of the on-ramp meter, either in the on-ramp queue or further upstream on the surface streets. However, storing vehicles on the streets without disrupting traffic is only possible with a more sophisticated coordination between the on-ramp meter and the street signals. In the case of independent freeway and surface street control systems, the only storage space available is on the on-ramps themselves. These can typically hold about 30 vehicles each, which is a small number compared to the number of vehicles that can be stored on the mainline. When the on-ramp queue grows too long and threatens to spill onto the street, many systems will override the controller and set the meter to its maximum rate. The amount and distribution of available storage capacity is thus seen to be a very important factor which should be considered in the calculation of the metering rates. However, very few controller designs take storage capacity into consideration. Exceptions include Di Febbraro et al. [28], Alessandri et al. [3], and the strategy proposed in Chapter 5 of this dissertation. These three impose *hard* constraints on the on-ramp queue lengths. Kotsialos and Papageorgiou [48, 49] used *soft* constraints – i.e. penalties in the objective function – to prevent the on-ramp queues from overflowing.

The use of mathematical models in the design of on-ramp metering strategies began with the linear programming approach proposed by Wattleworth [98] in 1965. Since then, a large number of strategies have been proposed, using an array of modeling and control techniques. These designs can be categorized in several ways, but perhaps the most meaningful is to separate them into *predictive* and *reactive* strategies. The reactive type, also known as *traffic-responsive* or *feedback* controllers, use real-time measurements from the freeway to compute the metering rates. Predictive or *time-of-day* strategies, use only historical data. The latter are susceptible to deviations of the traffic demands from their expected values, and should

---

[5]vphpl: vehicles per hour per lane.

therefore only be implemented within a larger control structure that includes feedback (e.g. model predictive control, hierarchical control). The next section gives a historical account of predictive on-ramp metering designs.

## 2.2.3 Predictive on-ramp control

Predictive on-ramp metering strategies use a model of the freeway and an estimate of future demands to *predict* the response of the system to on-ramp control. The metering rates are found as the solution to an optimization problem. The important components of the optimization problem are:

1. the objective function,

2. the traffic model,

3. the definition of the state, inputs, disturbances, and tunable parameters,

4. the bounds on the control and state variables.

For multiple-origin/destination networks, in addition to the conservation and flow equations, the model must also include a method for deciding how to divide the traffic stream at bifurcations. There are two ways of representing this information: with an *OD matrix* and with *split ratios*.

The OD matrix is a time varying matrix $A(t)$ whose components $a_{ij}(t)$ are the numbers of vehicles that enter the system at source $i$ and time $t$ that wish to exit through sink $j$. In networks such as linear freeways, with only one route for every origin/destination pair (i.e. no alternative routes), the OD matrix provides sufficient information for resolving the flows at the bifurcations (i.e. at offramps). Additional information is needed when there are multiple routes.

The OD matrix can be difficult to obtain in practice because it cannot be determined from aggregate flow measurements alone. This is because any particular set of flow measurements can be generated by a number of different OD matrices, even though a specific OD matrix generates only one set of flows. Many useful techniques have been developed for the difficult problem of OD matrix estimation [16, 56, 57, 92, 102, 103], however many remain unverified.

In the absence of a reliable OD matrix, the most common method for computing the flows at bifurcations is with *split ratios*. These are the time varying proportions of the total flow that take each branch of a bifurcation. In contrast to the OD matrix, the split ratios can be easily deduced from aggregate measurements before and after the bifurcation. When the application is on-ramp metering design, the main drawback of using split ratios is that it introduces the false assumption that they remain unchanged, despite the control of the on-ramps. In reality the composition of origin/destination types at any point in the network, and therefore the split ratios, will almost surely be affected by on-ramp control. This is because metering can induce diversion, and also because different on-ramps with different OD mixes may be metered at different rates. The decision of whether to use an OD matrix or split ratios to model bifurcations is therefore not always easy. One must consider the

drawbacks of each: the uncertainty implied in the estimation of the OD matrix, and the false assumption of insensitive split ratios.

One way of classifying predictive metering strategies is according to how their underlying model deals with bifurcations. They can also be classified according to the type of model (1st order, higher order), the scope of the system (linear freeway, freeway with parallel route, freeway plus network of streets), or the control measure (on-ramp metering only, on-ramp metering and variable speed limits, on-ramp metering and variable message signs). Yet another criterion, which is most useful for this presentation, is whether or not the assumption of a *steady-state* is made. Depending on this assumption, predictive strategies are referred to as either *static* or *dynamic*.

**Static predictive control**

Static or steady-state models assume that the demands (entry flows, OD matrix, split ratios), and the traffic state (density, flow, speed) remain constant for the duration of the control period. This assumption, although limiting, allows important simplifications for the control problem. Equation (2.2) implies that the flow is then uniform along portions of the freeway without on-ramps or offramps. The flow at any section of the system can be found by adding and subtracting upstream entry and exit flows. Therefore, under this assumption, the density state and the fundamental diagram are not needed to determine flow.

Wattleworth [98] was the first to pose the predictive on-ramp control problem. The goal of this early formulation was to select the on-ramp flows such that travel time was minimized. Travel time is minimized by maximizing offramp flows, or equivalently under the steady-state assumption, by maximizing on-ramp flows ($r_i$). Wattleworth assumed knowledge of the OD matrix, which was used to find proportions $w_{ij}$ of flow through the $j$th bottleneck that originated from on-ramp $i$. The total flow through bottleneck $j$, $\sum_i w_{ij} r_i$, was constrained by its capacity $b_j$. On-ramp flows $r_i$ were limited by the available demand $d_i$. The resulting linear program is stated below.

$$\max \sum_i r_i \tag{2.43}$$
$$\text{s.t.} \quad \sum_i w_{ij}\, r_i \ \leq \ b_j \qquad \text{for all bottlenecks } j$$
$$0 \ \leq \ r_i \ \leq \ d_i \qquad \text{for all on-ramps } i$$

Wattleworth also described how additional features might be included by adding constraints to Eq. (2.43). Two possible restrictions on the accumulation of vehicles in the on-ramp queues were suggested:

$$s_i \ \leq \ q_i \qquad \text{and} \qquad s_i \ = \ s_{i+1} \tag{2.44}$$

where $s_i$ is the slack variable associated with constraint $r_i \leq d_i$. The first, $s_i \leq q_i$, establishes an upper bound on the number of vehicles not granted access to the freeway, or equivalently, an upper bound on the rate of growth of the $i$th on-ramp queue. It does not imply a strict upper bound on the length of the queue, however. The second less restrictive option, $s_i = s_{i+1}$, distributes the queues evenly amongst all on-ramps.

Later contributions built upon Wattleworth's formulation by suggesting alternative objective functions and adding constraints. Yuan and Kreer [106] minimized the sum of the squared difference between the metering rate and the demand, claiming that this would lead to a better distribution of the queuing times. Chen et al. [17] used the total travel distance as the objective function to maximize, instead of total travel time [6]. Eldor [29] considered the capacity $b_j$, not as a given number, but as a normally distributed random variable. The capacity constraint was re-stated in this context as a restriction on the *probability* of overwhelming the bottleneck. Chance-constrained linear programming was used to cast this stochastic constraint in a deterministic form:

$$\text{Prob}\left\{\sum_i w_{ij}\, r_i \leq b_j\right\} \geq \alpha_j \quad \Leftarrow \quad \sum_i w_{ij}\, r_i \ \leq \ \text{E}\left\{b_j\right\}(1 - m_j) \tag{2.45}$$

where $\text{E}\left\{b_j\right\}$ is the expected value of $b_j$ and $m_j \in [0, 1]$ is a function of the probability $\alpha_j$ and the first order statistics of $b_j$.

Other efforts in the static predictive category have accounted for the effects of driver diversion. Wang and May [96] enhanced Wattleworth's formulation by considering the possibility of drivers deciding not to use the freeway because of the added waiting times. They used a very simple diversion model, which stated only that drivers with shorter trips would be more likely to divert than drivers with longer trips. Aside from this restriction, the diversion rates were allowed to take their optimal values. The linear program proposed by Wang and May is the following:

$$\max \sum_i r_i \tag{2.46}$$

$$\text{s.t.} \quad r_i = \sum_j p_{ij}\, a_{ij} \qquad \qquad \text{for all on-ramps } i$$

$$p_{ij} \leq 1 \qquad \qquad \text{for all on-ramps } i \text{ and offramps } j$$

$$p_{ij} \leq p_{ij+1} \qquad \qquad \text{for all on-ramps } i \text{ and offramps } j$$

$$\sum_i \sum_j \delta_{ijk}\, p_{ij}\, a_{ij} \ \leq \ b_k \qquad \qquad \text{for all bottlenecks } k$$

$p_{ij}$ is the portion of the demand $a_{ij}$ from on-ramp $i$ to offramp $j$ that does not divert. OD flows that do not divert at all have $p_{ij} = 1$. The third constraint captures the propensity of shorter trip users to divert (offramp $j$+1 is downstream of offramp $j$). The fourth constraint is an equivalent form of the capacity constraint of Eq. (2.43), with $\delta_{ijk} = 1$ if section $k$ is used by OD pair $ij$, and 0 otherwise.

Payne and Thompson [82] expanded the scope of the problem by considering the freeway and a system of parallel streets. More importantly, they introduced the use of Wardrop's first principle as determining the diversion behavior of drivers. This entailed a considerable complication in the statement and solution of the problem, as linearity was lost. Routing and metering sub-problems, referred to as *traffic assignment* and *optimal allocation*, had to

---

[6]The use of travel distance instead of travel time was first suggested by A.D. May in the discussion following [98].

be posed separately. A sub-optimal numerical method combining a heuristic algorithm and dynamic programming was proposed to solve the overall problem.

Iida et al. [39] recognized the bi-level assignment/metering problem as an instance of a Stackelberg planning game, where the control system plays the role of the *leader*, and acts by setting the metering rate, and the users act as *followers*, adjusting their routes in response to the action of the leader. The goal of the leader is to minimize total travel time, while the followers strive to minimize their individual travel times (Wardrop's principle). Although numerical methods for the general Stackelberg problem existed, Iida et al. proposed a heuristic fixed-point procedure, where solutions to the individual problems were re-inserted into each other, until convergence was achieved. Yang et al. [105] designed an improved numerical method for the bi-level problem using successive linear programming. Yang and Yagar [104] applied the method to the particular case of a freeway with a parallel street system (similar to [82]'s).

### Dynamic predictive control

All static predictive control designs assume that free-flow conditions prevail on the freeway and surrounding surface streets. Maintaining free-flow conditions requires that sufficient demand be either stored on the on-ramps or diverted. For freeways where the effect of capacity drop is pronounced, such a congestion-avoiding strategy might actually be optimal. However, when offramp blockage is dominant, a total elimination of congestion may be counterproductive (Section 2.2.1). Moreover, in either case it may not even be possible to store or divert the excess demand, due to limitations in on-ramp storage. Static models become poor predictors of freeway traffic when congestion cannot be avoided.

Dynamic models provide an important potential for improvement over static models for several reasons. First, they are not restricted to free-flow traffic. Second, limitations on allowable storage can be included by placing constraints on the on-ramp queue lengths[7]. Third, the limiting assumption of constant demands over the design period can be relaxed. This permits the design period to be extended over longer intervals, perhaps even the entire peak-period. Unfortunately, the consideration of congestion also introduces the nonlinearity of the fundamental diagram, which severely complicates the problem. To simplify, all of the on-ramp metering designs discussed below have reverted to the assumption that vehicles do not divert. The effects of voluntary diversion have apparently only been studied in the context of static models [39, 82, 104, 105].

Kaya [41] posed a fairly complete statement of the dynamic predictive problem in 1970. The objective function included terms penalizing the time spent in on-ramp queues, the queue length, and the mainline density in excess of the critical density. Constraints included a discretized form of LWR, constraints on the metering rate, and bounds on the queue lengths and waiting times. The complete problem contained nonlinearities in both the objective function and the constraints. These were approximated with piece-wise linear functions. The approximate optimization problem was solved using a specialized numerical package that could deal with the piece-wise linear constraints.

---

[7]Wardrop's principle can be used to obtain equilibrium queue lengths for steady-state models, as in [82], but this requires the solution of the traffic assignment problem.

Although Kaya's approach was practically complete in its formulation, its main drawback lied in the numerical method, which limited its application to small networks and short time horizons. Several more recent designs have employed more powerful numerical methods, and used higher order models of freeway traffic. Kotsialos et. al [45, 47] used METANET, the network extension of Payne's second order model, and considered the use of on-ramp metering in conjunction with routing suggestions through variable messaging signs (VMS) as the control. They solved the resulting nonlinear problem with a gradient method. Kotsialos and Papageorgiou [48, 49] applied the same technique to a network without VMS (ramp metering only) and simulated the method using data from the Amsterdam Orbital Motorway. Zhang [110] posed a similar problem using a first order model, and also solved it with a gradient method.

Several designs have been developed for the case in which the *speed limit*, in addition to the on-ramp flows, can be controlled. Di Febbraro et al. [28] posed this problem using Payne's model, modified to incorporate the effect of the variable speed limits. They assigned a neural network to the control law, and solved its coefficients with a gradient method. Hegyi et al. [38] also posed the problem using Payne's model, and used sequential quadratic programming to find a solution.

Other authors have assumed even stricter forms of control over the system. The design of Chang et al. [15] generates timing plans for the on-ramps and for signals on a parallel street route. It also assumes that vehicles can be removed from the freeway and made to complete their trips on the parallel route. Ziliaskopoulos [114] posed the problem assuming *complete* control over the freeway: on-ramp metering rates as well as mainline, bifurcation, and merging flows were controlled. This highly unrealistic scenario was studied in order to provide insights into the properties of the optimal plan for single destination networks. The paper reached a useful conclusion: *the optimal traffic assignment for a single destination network is such that the cost from any point to the single destination – along a used route – equals the marginal cost of an extra unit of demand at that point in space and time, and the cost along an unused route is larger.* Laval et al. [51] made use of this principle to derive an explicit optimal assignment for a network consisting of a linear freeway with an alternative route, and with a single downstream destination.

A different approach to the problem was taken by Lovell [58] and Lovell et al. [59]. Unlike the designs described above, they assumed knowledge of the OD matrix, instead of the offramp split ratios. They also assumed that the freeway remained uncongested, but preserved the order of vehicles in the on-ramp queues. That is, the first-in-first-out discipline was observed. This introduced the additional unknown of the *waiting time* for each on-ramp, which was assumed to be time variant but OD-independent. Also in contrast to previous works, the optimization problem of [58, 59] was posed in continuous time. It was shown that despite its nonlinear form, global solutions could be found in four special cases: single origin, single bottleneck, single destination, and constant OD proportions. Erera et al. [30] proved that the general form of Lovell's problem is NP-complete.

# Chapter 3

# The Asymmetric Cell Transmission Model

The formulation of the optimal ramp metering control problem includes a model of freeway traffic among its components. The mathematical form and precision of the model will influence both the effectiveness of the optimal plan, and the difficulty of finding a good solution. Ideally, the model should capture all of the travel time reducing mechanisms defined in Section 2.2.1. However, an overly complicated model will also be difficult to solve, and the solution that can be found within a limited time frame may be less than optimal. A higher quality solution to a simpler – i.e. less complete – model may be preferred.

In this chapter we describe the model that is used to formulate the predictive metering design of Chapter 5. This model has been dubbed the Asymmetric Cell Transmission Model (ACTM), due to its similarities with the original CTM [21, 22] described in Section 2.1. The main difference resides in the treatment of merging flows. The CTM merge model is *symmetric* in the sense that it is not altered if labels 'a' and 'b' in Eqs. (2.14) through (2.18) are swapped, and $p$ is replaced with $1 - p$. On the other hand, a clear distinction is made in the ACTM between *through* and *joining* flows at the merge.

The advantage of using the ACTM for optimization is that its merge model involves a concave min{} function, as opposed to the non-convex/non-concave mid{} function of the CTM (Eqs. (2.14) and (2.15)). This is shown in Chapter 5 to be tremendously beneficial to the numerical solution of the optimization problem.

The chapter is organized as follows. The model equations are presented in Section 3.1. Section 3.2 provides proof of an important property of the model: despite its simple concave structure, the flow and density variables are guaranteed to remain within certain implicit bounds. An extension to the model that allows a slower update frequency for the ramp metering rates is presented in Section 3.3. The validity of the model is tested using a realistic freeway configuration in Section 3.4.

## 3.1   Model Equations

The ACTM is intended to model one-way roads with on-ramps and offramps, as illustrated in Figure 3.1. To use the ACTM, it should be possible at any merging junction to distinguish

between the *through* or *mainline* branch and the *joining* branch. The mainline branch carries traffic that can move through the merge without changing lanes. The joining branch is usually a secondary access ramp. This restriction is not too limiting: typical configurations for on-ramp/freeway and even freeway/freeway junctions permit an unambiguous definition of the *through* and the *joining* traffic streams.
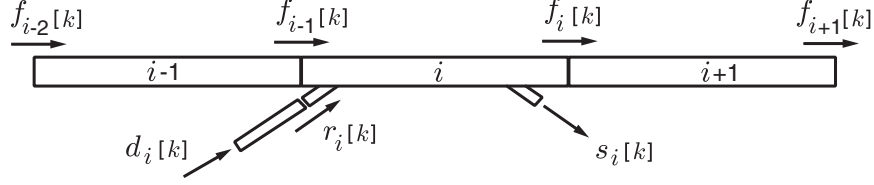


Figure 3.1: Interpretation of model variables.

The freeway mainline is divided into I sections. Section boundaries are chosen such that every section contains at most one on-ramp and/or one offramp. Furthermore, in sections containing both an on-ramp and an offramp, the on-ramp should be upstream of the offramp, as shown in Figure 3.1. Sections are numbered $0 \dots I-1$, starting from the upstream-most section, and proceeding sequentially downstream. Time is divided into K uniform intervals of length $\Delta t$. The following are sets of section and time indices:

| | | | |
|---|---|---|---|
| $\mathcal{I}$ | : | set of all freeway sections | $\mathcal{I} = \{0 \dots I-1\}$ |
| $\mathcal{K}$ | : | set of time intervals | $\mathcal{K} = \{0 \dots K-1\}$ |
| $\mathcal{E}n$ | : | set of sections with on-ramps | $\mathcal{E}n \subseteq \mathcal{I}$ |
| $\mathcal{E}n^+$ | : | set of sections with metered on-ramps | $\mathcal{E}n^+ \subseteq \mathcal{E}n$ |

All traffic variables and boundary conditions are normalized to vehicle units. Flow variables – $f_i[k]$, $r_i[k]$, $r_i^c[k]$, $d_i[k]$, and $s_i[k]$ – are interpreted as a number of vehicles per time interval. Density variables – $\rho_i[k]$ and $l_i[k]$ – represent respectively the number of vehicles on the mainline and on-ramp portions of section $i$ at time $k\Delta t$. Definitions for each of these quantities are given below.

| | | |
|---|---|---|
| $\rho_i[k]$ | : | number of vehicles in section $i$ at time $k\Delta t$ |
| $l_i[k]$ | : | number of vehicles queueing in the on-ramp of section $i \in \mathcal{E}n$ at time $k\Delta t$ |
| $f_i[k]$ | : | number of vehicles moving from section $i$ to $i+1$ during interval $k$ |
| $r_i[k]$ | : | number of vehicles entering section $i \in \mathcal{E}n$ from its on-ramp during interval $k$ |
| $r_i^c[k]$ | : | metering rate for on-ramp $i \in \mathcal{E}n^+$ |
| $d_i[k]$ | : | demand for on-ramp $i \in \mathcal{E}n$ |
| $s_i[k]$ | : | number of vehicles using offramp $i$ during interval $k$ |
| $\beta_i[k]$ | : | dimensionless split ratio for offramp $i$ |

An on-ramp indicator $\delta_i$ is used to simplify the notation:

$$\delta_i \triangleq \begin{cases} 1 & \text{if} \quad i \in \mathcal{E}n \\ 0 & \text{else} \end{cases} \tag{3.1}$$

The parameters of the model are listed below. Their relation to a truncated triangular fundamental diagram is illustrated in Figure 3.2.

| | | | |
|---|---|---|---|
| $v_i$ | : | normalized free-flow speed | $\in [0, 1]$ |
| $w_i$ | : | normalized congestion wave speed | $\in [0, 1]$ |
| $\bar{\rho}_i$ | : | jam density | [veh] |
| $\bar{f}_i$ | : | mainline capacity | [veh] |
| $\bar{s}_i$ | : | offramp capacity | [veh] |
| $\alpha_i, \gamma_i , \xi_i$ | : | influence parameters | $\in [0, 1]$ |
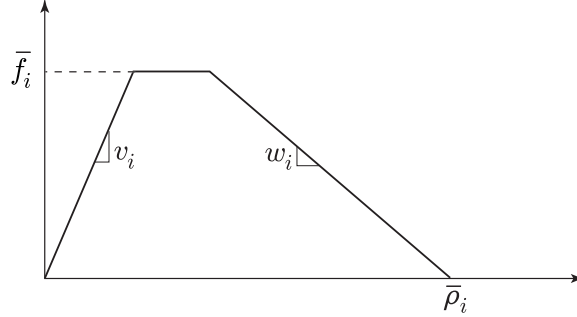


Figure 3.2: Interpretation of model parameters.

The model has five basic components: the mainline and on-ramp conservation equations, mainline and on-ramp flows, and offramp flows. Offramp flows are assumed to be related to the mainline flows by a known, possibly time-varying, split ratio $\beta_{i[k]} \in [0, 1]$. Similarly to the CTM, $\beta_{i[k]}$ is defined as the ratio of one of the partial flows to the total flow:

<u>Offramp flows</u>    $\forall\, i \in \mathcal{I}\,,\ k \in \mathcal{K}$ :

$$s_{i[k]} \;=\; \beta_{i[k]}\big(s_{i[k]} + f_{i[k]}\big) \tag{3.2}$$

$$\therefore \quad s_{i[k]} \;=\; \frac{\beta_{i[k]}}{1 - \beta_{i[k]}} f_{i[k]} \;=\; \frac{\beta_{i[k]}}{\bar{\beta}_{i[k]}} f_{i[k]} \tag{3.3}$$

where $\bar{\beta}_{i[k]} \triangleq 1 - \beta_{i[k]}$ has been defined to simplify the equations. Also for convenience, the split ratio is defined for every section, and set to 0 if the section does not contain an offramp. One important difference between the ACTM and the CTM relates to the asymmetry of this diverge model. The ACTM only supports diverges where one branch is an *exiting* branch, and the other is a *continuing* branch. The exiting branch removes traffic from the simulation, whereas the continuing branch transfers vehicles to the next downstream section. Moreover, the exiting branch is assumed to feed an imaginary section with infinite storage capacity, so no congestion can propagate into the mainline from the exiting branch (although an exit capacity $\bar{s}_i$ is included). This is not the case with the CTM, where both branches are identical.

Mainline flows are calculated, similarly to the CTM, as the minimum of what can be sent by the upstream section assuming maximum speed and what can be received by the downstream section without causing overflow. The mainline flow is calculated as the largest

value of $f_{i[k]}$ that that complies with the following constraints:

$$
\begin{aligned}
f_{i[k]} + s_{i[k]} &\leq v_i\big(\rho_{i[k]} + \delta_i\gamma_i r_{i[k]}\big) & \ldots \text{free-flow term} & \quad (3.4)\\
f_{i[k]} &\leq w_{i+1}(\bar{\rho}_{i+1} - \rho_{i+1[k]}) - \delta_{i+1}\alpha_{i+1} r_{i+1[k]} & \ldots \text{congestion term} & \quad (3.5)\\
f_{i[k]} &\leq \bar{f}_i & \ldots \text{mainline capacity} & \quad (3.6)\\
s_{i[k]} &\leq \bar{s}_i & \ldots \text{offramp capacity} & \quad (3.7)
\end{aligned}
$$

Equation (3.4) limits the total flow that can leave section $i$ during time interval $k$, assuming that traffic moves at the free-flow speed $v_i$. This total includes a portion $\gamma_i$ of the traffic that enters section $i$ from its on-ramp that can reach the downstream boundary within the same time interval. $\gamma_i$ should be chosen based on the distance from the on-ramp to the downstream boundary.

Equation (3.5) represents the restriction placed on $f_{i[k]}$ by the amount of space available in section $i+1$. The right hand side of this equation is the portion $w_{i+1}$ of the total unoccupied space $(\bar{\rho}_{i+1} - \rho_{i+1[k]})$ that can be filled by vehicles coming from section $i$, discounted by an amount of space $\alpha_{i+1}r_{i+1[k]}$ that is occupied by traffic from the downstream on-ramp. The parameter $\alpha_i$ determines the influence of the on-ramp flow on the mainline flow that enters section $i$. It should be chosen to reflect the position of the on-ramp within the section, with larger values of $\alpha_i$ corresponding to on-ramps that are closer to the upstream edge.

Equations (3.6) and (3.7) are, respectively, the mainline and offramp capacity limits. Considering Eq. (3.3), this leads to the following expression for $f_{i[k]}$:

Mainline flows $\quad \forall\, i \in \mathcal{I}\,,\ k \in \mathcal{K}:$ $\hspace{5cm}$ (3.8)

$$
f_{i[k]} = \min\left\{\, \bar{\beta}_{i[k]}\, v_i\big(\rho_{i[k]} + \delta_i\gamma_i r_{i[k]}\big)\ ;\ w_{i+1}(\bar{\rho}_{i+1} - \rho_{i+1[k]}) - \delta_{i+1}\alpha_{i+1} r_{i+1[k]}\ ;\ \bar{f}_i\ ;\ \frac{\bar{\beta}_{i[k]}}{\beta_{i[k]}}\bar{s}_i \,\right\}
$$

Similarly, on-ramp flows are computed such that none of the following limits are exceeded:

$$
\begin{aligned}
r_{i[k]} &\leq l_{i[k]} + d_{i[k]} & \ldots \text{demand} & \quad (3.9)\\
r_{i[k]} &\leq \xi_i(\bar{\rho}_i - \rho_{i[k]}) & \ldots \text{mainline space} & \quad (3.10)\\
r_{i[k]} &\leq r^c_{i[k]} & \ldots \text{ramp metering (for } i \in \mathcal{E}n^+) & \quad (3.11)
\end{aligned}
$$

Equations (3.9) and (3.11) represent the fact that the flow on the on-ramp cannot exceed either the available demand $(l_{i[k]} + d_{i[k]})$, or the rate imposed by the on-ramp meter ($r^c_{i[k]}$) when $i \in \mathcal{E}n^+$). Equation (3.10) is the restriction to $r_{i[k]}$ due to limited space on the mainline. The parameter $\xi_i$ determines the allotment of available space to vehicles entering from the on-ramp. Its value should reflect the geometrical layout of the section. For example, if the on-ramp is located at the midpoint, incoming vehicles will only have access to the downstream half of the section. On multi-lane freeways, vehicles entering from a right-side on-ramp may not be able to reach open space in the leftmost lanes. Equations (3.9) through (3.11) lead to the following expression for $r_{i[k]}$:

On-ramp flows $\quad \forall\, i \in \mathcal{E}n\,,\ k \in \mathcal{K}:$

$$
r_{i[k]} = \begin{cases} \min\big\{\, l_{i[k]} + d_{i[k]}\ ;\ \xi_i(\bar{\rho}_i - \rho_{i[k]})\,\big\} & \text{if } i \in \mathcal{E}n \setminus \mathcal{E}n^+ \\ \min\big\{\, l_{i[k]} + d_{i[k]}\ ;\ \xi_i(\bar{\rho}_i - \rho_{i[k]})\ ;\ r^c_{i[k]}\,\big\} & \text{if } i \in \mathcal{E}n^+ \end{cases} \qquad (3.12)
$$

The number vehicles in the on-ramps ($l_{i[k]}$) and on the mainline ($\rho_{i[k]}$), evolve according to conservation equations (3.13) and (3.14).

<u>On-ramp conservation</u>   $\forall\, i \in \mathcal{E}n\,,\ k \in \mathcal{K}:$

$$l_{i[k+1]} \;=\; l_{i[k]} \;+\; d_{i[k]} \;-\; r_{i[k]} \tag{3.13}$$

with initial condition $l_{i[0]}$ and boundary condition $d_{i[k]}$.

<u>Mainline conservation</u>   $\forall\, i \in \mathcal{I}\,,\ k \in \mathcal{K}:$

$$\begin{aligned}
\rho_{i[k+1]} &\;=\; \rho_{i[k]} \;+\; f_{i-1[k]} \;+\; \delta_i\, r_{i[k]} \;-\; f_{i[k]} \;-\; s_{i[k]} \\
&\;=\; \rho_{i[k]} \;+\; f_{i-1[k]} \;+\; \delta_i\, r_{i[k]} \;-\; f_{i[k]}/\bar{\beta}_{i[k]}
\end{aligned} \tag{3.14}$$

with initial condition $\rho_{i[0]}$. The boundary condition for this equation is the flow entering the mainline at its upstream boundary, $up_{[k]}$. It can be represented either as the mainline flow into section 0, i.e. $f_{-1[k]} = up_{[k]}$, or as the demand into a fictitious on-ramp, i.e. $d_{0[k]} = up_{[k]}$ and $f_{-1[k]} = 0$. The latter approach helps to avoid numerical problems related to mainline congestion obstructing the upstream source, but increases the dimensionality of the model.

Equations (3.8), (3.12), (3.13), and (3.14) constitute the basic ACTM model. A second important distinction of the ACTM with respect to the CTM, in addition to the bifurcation model, is in the merge model implied by Eqs. (3.8) and (3.12). Unlike the CTM, the merge junction in the ACTM is composed of a *through* branch and a *joining* branch, and distinct equations apply to each one. Three *influence parameters*, $\alpha_i$, $\gamma_i$, and $\xi_i$, are involved in the merge model. Their influence on the mainline and on-ramp flows is illustrated in Figure 3.3. In short, $\gamma_i$ determines the direct influence of the on-ramp flow on the downstream mainline flow, $\alpha_i$ the influence of the on-ramp on the upstream mainline flow, and $\xi_i$ the influence of the mainline density on the on-ramp flow.
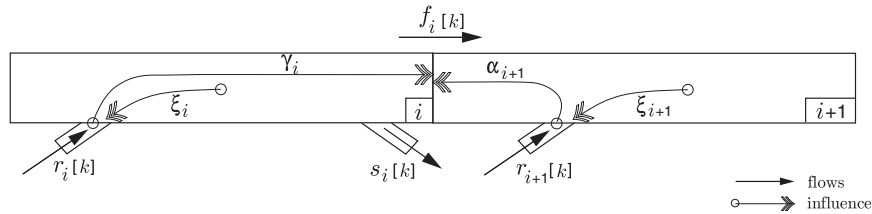


Figure 3.3: Directions of influence.

The main advantage of the ACTM over the CTM is in the replacement of the non-concave / non-convex mid{} function of the CTM with concave min{} functions. This structural simplification is the basis for the predictive controller presented in Chapter 5. Section 3.4 demonstrates that the ACTM is capable of replicating freeway traffic behavior. But first, it is shown that the model preserves some important properties such as positivity (i.e no wrong-way travel) from the original CTM.

## 3.2   Implicit bounds

An important property of the CTM is that it never produces negative flows or densities, nor do the densities ever exceed the jam density. That is, in the notation of the CTM defined

on page 15, the following constraints automatically apply:

$$n_i(k) \in [\, 0, N_i \,] \qquad , \qquad y_i(k) \geq 0 \tag{3.15}$$

These *implicit bounds* are a minimum requirement for any model to be considered as a reasonable approximation of freeway traffic. In the case of the CTM, they are a consequence of the consistency of the model with the LWR theory, and of the particular rules used for merges and diverges. Some models, such as the one proposed by Ross in [90], impose bound constraints *explicitly*. Papageorgiou suggests in [55] that explicit constraints can be used to solve the problem of backward moving traffic associated with higher order models. Such an artificial "fix" has the drawback that it requires the nominal model to be ignored (i.e. violated) whenever one of the additional constraints is reached. This annuls the interpretation of the model equations in these situations, and compromises its usefulness as a tool for understanding traffic behavior.

The question that arises is whether the ACTM has the property expressed by Eq. (3.15). The answer is not obvious. For example, the right hand side of Eq. (3.5) could be negative. The following theorem provides conditions on the model parameters and boundary conditions that ensure a physically reasonable evolution of the model, thus avoiding the need for explicit constraints.

**Theorem 1** *Given initial and boundary conditions, ramp metering rates, and model parameters satisfying,*

| | | |
|---|---|---|
| *Initial conditions :* | $\rho_i[0] \in [0, \bar{\rho}_i]$ | $\forall\, i \in \mathcal{I}$ |
| | $l_i[0] \geq 0$ | $\forall\, i \in \mathcal{E}n$ |
| *Boundary conditions :* | $d_i[k] \geq 0$ | $\forall\, i \in \mathcal{E}n \,,\, k \in \mathcal{K}$ |
| | $f_{-1}[k] = 0$ | $\forall\, k \in \mathcal{K}$ |
| *On-ramp metering rates :* | $r_i^c[k] \geq 0$ | $\forall\, i \in \mathcal{E}n^+ \,,\, k \in \mathcal{K}$ |
| *Model parameters :* | $v_i \,,\, w_i \in [0, 1]$ | $\forall\, i \in \mathcal{I}$ |
| | $\alpha_i \,,\, \gamma_i \in [0, 1]$ | $\forall\, i \in \mathcal{E}n$ |
| | $\xi_i \in \left[\, 0 \,,\, \min\left(\frac{w_i}{\alpha_i}, \frac{1-w_i}{1-\alpha_i}\right) \,\right]$ | $\forall\, i \in \mathcal{E}n$ |
| | $\bar{f}_i \,,\, \bar{s}_i \geq 0$ | $\forall\, i \in \mathcal{I}$ |
| | $\beta_i[k] \geq 0$ | $\forall\, i \in \mathcal{I} \,,\, k \in \mathcal{K}$ |

*the evolution of the ACTM in time is guaranteed to remain bounded as follows:*

$$\rho_i[k] \in [0, \bar{\rho}_i] \quad , \quad f_i[k] \geq 0 \qquad \forall\, i \in \mathcal{I} \,,\, k \in \mathcal{K}$$
$$l_i[k] \geq 0 \qquad , \quad r_i[k] \geq 0 \qquad \forall\, i \in \mathcal{E}n \,,\, k \in \mathcal{K}$$

**Proof**
The proof is by induction. Assuming that $\rho_i[k] \in [0, \bar{\rho}_i]$ and $l_i[k] \geq 0$ holds for some $k$, we show that $f_i[k] \geq 0$ and $r_i[k] \geq 0$. We then show that this implies $\rho_i[k+1] \in [0, \bar{\rho}_i]$ and $l_i[k+1] \geq 0$. Because $\rho_i[k] \in [0, \bar{\rho}_i]$ and $l_i[k] \geq 0$ holds for $k = 0$, the result follows.

First, from Eq. (3.12), with $l_i[k] \geq 0$, $d_i[k] \geq 0$, $\xi_i \geq 0$, $\rho_i[k] \leq \bar{\rho}_i$, $r_i^c[k] \geq 0$, it follows that $r_i[k] \geq 0$. To show $f_i[k] \geq 0$, we need to check that each of the four terms in Eq. (3.8) is

positive. The only non-obvious one is the second. However,

$$
\begin{aligned}
\xi_{i+1} \;\leq\; \frac{w_{i+1}}{\alpha_{i+1}} \quad \Rightarrow \quad & \xi_{i+1}(\bar{\rho}_{i+1} - \rho_{i+1}[k]) \;\leq\; \frac{w_{i+1}}{\alpha_{i+1}}(\bar{\rho}_{i+1} - \rho_{i+1}[k]) \\
\Rightarrow \quad & \delta_{i+1} r_{i+1}[k] \;\leq\; \frac{w_{i+1}}{\alpha_{i+1}}(\bar{\rho}_{i+1} - \rho_{i+1}[k]) \\
\Rightarrow \quad & 0 \;\leq\; w_{i+1}(\bar{\rho}_{i+1} - \rho_{i+1}[k]) - \delta_{i+1}\alpha_{i+1} r_{i+1}[k]
\end{aligned}
$$

Therefore, $f_{i[k]} \geq 0$. Using the above, we can deduce $l_{i[k+1]} \geq 0$ and $\rho_{i[k+1]} \in [0, \bar{\rho}_i]$:

$$
\begin{aligned}
l_{i[k+1]} \;=\;\; & l_{i[k]} + d_{i[k]} - r_{i[k]} \\
\geq\;\; & l_{i[k]} + d_{i[k]} - (l_{i[k]} + d_{i[k]}) \\
=\;\; & 0 \\
\rho_{i[k+1]} \;=\;\; & \rho_{i[k]} + f_{i-1[k]} - f_{i[k]} / \bar{\beta}_{i[k]} + \delta_i r_{i[k]} \\
\geq\;\; & \rho_{i[k]} - f_{i[k]} / \bar{\beta}_{i[k]} + \delta_i r_{i[k]} \\
\geq\;\; & \rho_{i[k]} - \bar{\beta}_{i[k]} v_i(\rho_{i[k]} + \delta_i \gamma_i r_{i[k]}) / \bar{\beta}_{i[k]} + \delta_i r_{i[k]} \\
=\;\; & (1 - v_i)\rho_{i[k]} + \delta_i(1 - v_i \gamma_i) r_{i[k]} \\
\geq\;\; & 0 \\
\rho_{i[k+1]} \;=\;\; & \rho_{i[k]} + f_{i-1[k]} - f_{i[k]} / \bar{\beta}_{i[k]} + \delta_i r_{i[k]} \\
\leq\;\; & \rho_{i[k]} + f_{i-1[k]} + \delta_i r_{i[k]} \\
\leq\;\; & \rho_{i[k]} + w_i(\bar{\rho}_i - \rho_{i[k]}) - \delta_i \alpha_i r_{i[k]} + \delta_i r_{i[k]} \\
=\;\; & (1 - w_i)\rho_{i[k]} + \delta_i r_{i[k]}(1 - \alpha_i) + w_i \bar{\rho}_i \\
& \vdots \\[6pt]
\vdots \qquad & \\
\leq\;\; & (1 - w_i)\rho_{i[k]} + \delta_i \xi_i(\bar{\rho}_i - \rho_{i[k]})(1 - \alpha_i) + w_i \bar{\rho}_i \\
=\;\; & \begin{cases} (1 - w_i)\rho_{i[k]} + w_i \bar{\rho}_i & \text{if } i \notin \mathcal{E}n \\ (1 - \bar{w}_i)\rho_{i[k]} + \bar{w}_i \bar{\rho}_i & \text{if } i \in \mathcal{E}n \end{cases} \\
\leq\;\; & \bar{\rho}_i
\end{aligned}
$$

where $\bar{w}_i \triangleq w_i + \xi_i(1 - \alpha_i)$. The last line holds since by assumption both $w_i$ and $\bar{w}_i \in [0, 1]$.
∎

   This theorem ensures that unrealistic behaviors such as backward moving traffic, negative densities, and densities above the jam density, are not predicted by the model. Most of the conditions of the theorem are covered by the physical definitions of the model parameters and variables; e.g. $v_i, w_i \in [0, 1]$, $d_{i[k]} \geq 0$, etc. The only two that are not trivially satisfied are $f_{-1[k]} = 0$ and the upper bound placed on $\xi_i$. The first, $f_{-1[k]} = 0$, is met if the upstream mainstream boundary is represented as an additional fictitious uncontrolled on-ramp into section $i = 0$. The only restrictive condition is then the upper bound for $\xi_i$, which requires it to be less than both $w_i/\alpha_i$ and $(1 - w_i)/(1 - \alpha_i)$. Section 5.5 demonstrates that, at least in the case of the test site considered in that section, this bound is larger than any practical choice of $\xi_i$.

## 3.3 Dual time scales

Consistency of the mainline conservation equation, Eq. (3.14), requires that the time interval $\Delta t$ satisfy an analogous form of Eq. (2.11):

$$\Delta t \ \leq \ \min_i \left( L_i / v_i \right) \tag{3.16}$$

where $L_i$ is the length of the $i$th section. This means that $\Delta t$ should be small enough so that no vehicle travelling at the free-flow speed can traverse an entire section within one time interval. To give an idea of the typical length of the time interval, we consider the I-210 test section described in Chapter 6, whose shortest section is 962 feet long. At 65 mph, this corresponds to a maximum interval duration of about 10 seconds. However, it is not possible for the control system on I-210 to update the metering rates at such a high frequency. Data exchange between the freeway and the District 7 Traffic Management Center occurs only once every 30 seconds, and it takes two transfers, i.e. 1 minute, to update the control plan. Aside from this technical problem, it may not make practical sense to change the metering rate more than once every few minutes, since doing so may confuse drivers.

These two factors – the restriction of the time interval used for the mainline conservation equation to about 10 seconds, and the natural choice of a much slower control update frequency – motivate the use of a coarser time scale for the on-ramp metering rates $r_i^c[k]$. The benefit of increasing the size of the control time interval is that, as explained in Section 5.4.1, it dramatically reduces the dimension of the optimization problem.
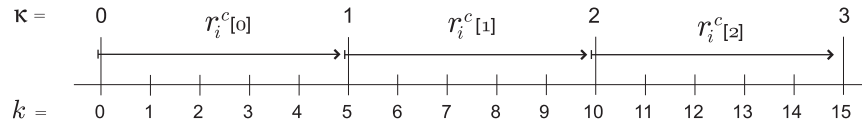


Figure 3.4: Dual time scales ($p = 5$).

In the dual-scale ACTM (Eqs. (3.18) through (3.21)) $\Delta t_c$ denotes the larger time interval used for control, $\Delta t$ still denotes the nominal time interval satisfying Eq. (3.16), and the integer $p$ is their ratio:

$$\Delta t_c \ = \ p \, \Delta t \tag{3.17}$$

As illustrated in Figure 3.4, $\kappa$ is a time index used for the on-ramp metering rates, and $k$ is the index used for all other variables. The mainline model of the dual-scale ACTM, Eqs. (3.18) and (3.19), and the on-ramp conservation equation (3.20) are identical to their original versions. The only difference is in the specification of the on-ramp flow, where the "$r_i^c[k]$" term from Eq. (3.12) is replace in Eq. (3.21) with "$r_i^c[\kappa]/p$".

**Dual-scale ACTM (dACTM-c)**:

$$\rho_{i[k+1]} \;=\; \rho_{i[k]} \;+\; f_{i-1[k]} \;+\; \delta_i r_{i[k]} \;-\; f_{i[k]}/\bar{\beta}_{i[k]} \qquad \forall\, i \in \mathcal{I}\,,\; k \in \mathcal{K}^p \qquad (3.18)$$

$$f_{i[k]} = \min\left\{ \bar{\beta}_{i[k]} v_i(\rho_{i[k]} + \delta_i \gamma_i r_{i[k]}) \;;\; w_{i+1}(\bar{\rho}_{i+1} - \rho_{i+1[k]}) - \delta_{i+1}\alpha_{i+1} r_{i+1[k]} \;;\; \bar{f}_i \;;\; \frac{\bar{\beta}_{i[k]}}{\beta_{i[k]}}\bar{s}_i \right\}$$

$$\forall\, i \in \mathcal{I}\,,\; k \in \mathcal{K}^p \qquad (3.19)$$

$$l_{i[k+1]} \;=\; l_{i[k]} \;+\; d_{i[k]} \;-\; r_{i[k]} \qquad \forall\, i \in \mathcal{E}n\,,\; k \in \mathcal{K}^p \qquad (3.20)$$

$$r_{i[k]} = \begin{cases} \min\left\{ l_{i[k]} + d_{i[k]} \;;\; \xi_i(\bar{\rho}_i - \rho_{i[k]}) \;;\; r^c_{i[\kappa]}/p \right\} & \forall\, i \in \mathcal{E}n^+\,,\; k \in \mathcal{K}^p \\[2mm] \min\left\{ l_{i[k]} + d_{i[k]} \;;\; \xi_i(\bar{\rho}_i - \rho_{i[k]}) \right\} & \forall\, i \in \mathcal{E}n \setminus \mathcal{E}n^+\,,\; k \in \mathcal{K}^p \end{cases} \qquad (3.21)$$

In Eq. (3.21), $\kappa$ is the integer part of $k/p$, as illustrated in Figure 3.4. $\mathcal{K}^p$ is the set of $\Delta t$-sized time intervals: $\mathcal{K}^p = [0, \ldots, pK - 1]$. This version of the ACTM is labelled dACTM-c ('d' for dual-scale and 'c' for control) to distinguish it from a third version described in Section 5.4.1, which is labelled dACTM-o ('o' for on-ramps).

## 3.4 Tests

This section demonstrates the use of the ACTM on a 15-mile stretch of Interstate 210 in Pasadena, California. The details of the test site and data collection procedures are described in Chapter 6. The I-210 site contains 20 metered on-ramps and a single uncontrolled freeway connector from I-605 NB. A measured speed contour plot, constructed with the methods described in Chapter 6, is shown in Figure 3.5. The darker shaded areas indicate average speeds below 40 mph.
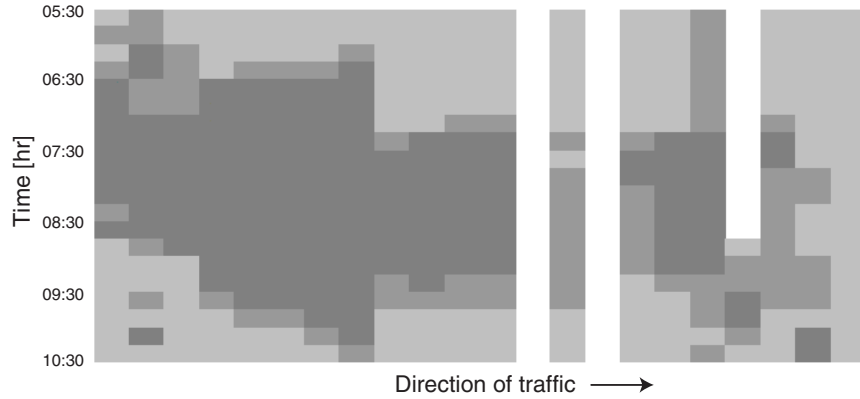


Figure 3.5: Measured speed contour plot [mph].

To apply the ACTM, the site was partitioned into 40 sections of varying lengths, as illustrated in Figure 3.6. The lengths of the individual sections are indicated in the figure in feet. A simple manual parameter calibration was performed, with resulting speed contour plot shown in Figure 3.7. The speed variable used in Figure 3.7 was calculated with:

$$\text{vel}_{i[k]} \triangleq \frac{f_{i[k]}/\bar{\beta}_{i[k]}}{\rho_{i[k]} + \gamma_i\, r_{i[k]}} \left( \frac{L_i}{\Delta t} \right) \qquad (3.22)$$

40

This equation ensures $\text{vel}_{i[k]} = v_i$ when the freeway section is free-flowing. The tuned parameter set is provided in Table 3.1. Notice that the values in the table are fairly uniform: free-flow speed is 65 mph, congestion propagation speed is 8.4 mph, jam density is 260 veh/mile/lane, capacity is 2200 veh/hr/lane, $\alpha_i = \gamma_i = 0.0$, and $\xi_i = 0.0$, almost everywhere. The only exceptions are sections 8, 14, and 15. These correspond to the merge of I-210 with I-605, in section 8, and the Huntington Av. bottleneck (identified as bottleneck "B1" in Section 6.8), at sections 14 and 15. The indicated adjustment to $w_8$, $\bar{\rho}_8$, and $\xi_8$ produced the correct distribution of congestion between the I-605 and I-210 branches of the merge. The changes to $w_i$ and $\bar{\rho}_i$ in sections 14 and 15 were made based on the bottleneck activation time, discharge rate, and duration.
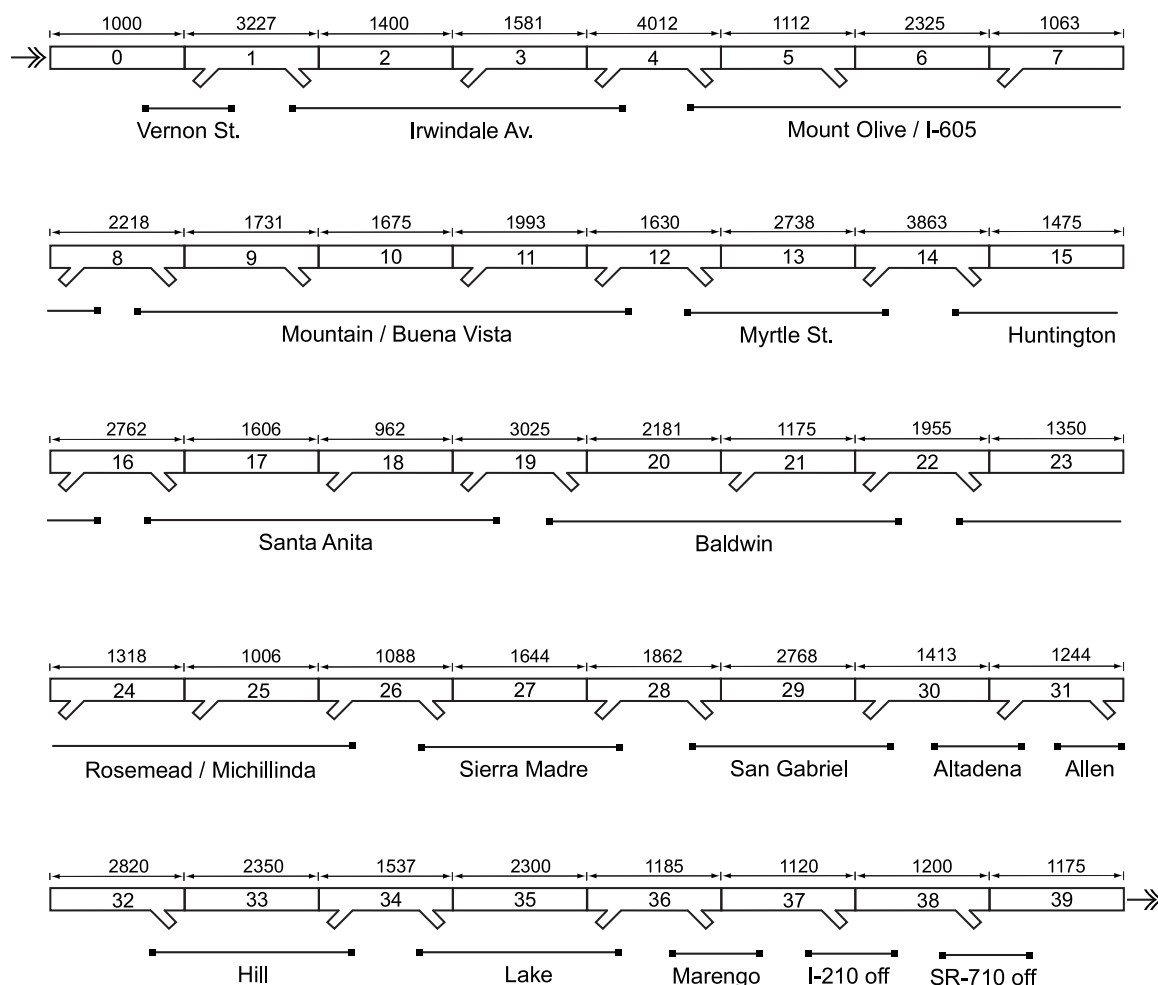


Figure 3.6: 40 ACTM sections in the I-210 test site (section lengths in feet).

This simple calibration effort is intended to establish the feasibility of the ACTM as a model of freeway traffic, in order to justify its use in control design. A more thorough calibration effort, using more sophisticated methods such as the gradient method of [20], the Kalman filtering approach of [34], or the integrated approach of [65], has not been performed.

41
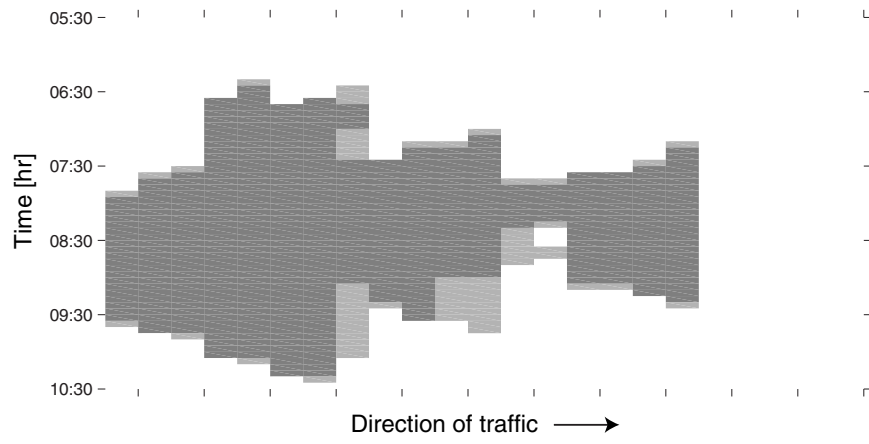
Figure 3.7: Simulated speed contour.

| | Mainline parameters | | | | On-ramp parameters | | | |
|---|---|---|---|---|---|---|---|---|
| $i$ | $v_i \frac{L_i}{\Delta t}$ [mph] | $w_i \frac{L_i}{\Delta t}$ [mph] | $\bar{\rho}_i \frac{1}{L_i \cdot n_i}$ [vpm/lane] | $\bar{f}_i \frac{1}{\Delta t \cdot n_i}$ [vph/lane] | $\delta_i$ | $\alpha_i$ | $\gamma_i$ | $\xi_i$ |
| 0 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 1 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 2 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 3 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 4 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 5 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 6 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 7 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 8 | 65 | 8.0 | 250 | 2,200 | 1 | 0 | 0 | 0.027 |
| 9 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 10 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 11 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 12 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 13 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 14 | 65 | 7.7 | 240 | 2,200 | 1 | 0 | 0 | 0.3 |
| 15 | 65 | 7.7 | 240 | 2,200 | 0 | - | - | - |
| 16 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 17 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 18 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 19 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 20 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 21 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 22 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 23 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 24 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 25 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 26 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 27 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 28 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 29 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 30 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 31 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 32 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 33 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 34 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 35 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 36 | 65 | 8.4 | 260 | 2,200 | 1 | 0 | 0 | 0.3 |
| 37 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 38 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |
| 39 | 65 | 8.4 | 260 | 2,200 | 0 | - | - | - |

Table 3.1: Parameters for I-210 (mph=mile/hr, vpm=veh/mile, vph=veh/hr, $n_i$=# lanes).

# Chapter 4

# A Study of Two Metering Strategies Using a Discontinuous ACTM

In this chapter we apply the ACTM to the analysis of two popular local ramp metering strategies – Alinea and Percent-Occupancy (%-Occ) – in order to compare their merits, and to provide some suggestions for their tuning.
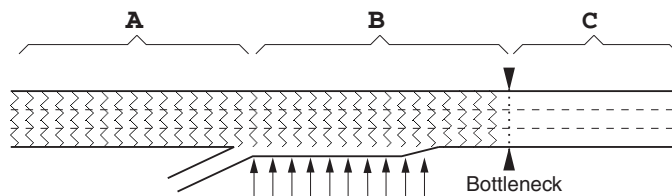


Figure 4.1: Detector placement.

The study focuses on a stretch of freeway in the vicinity of a single metered on-ramp, as illustrated in Figure 4.1. Vehicles entering the freeway from the on-ramp join the mainline stream by changing lanes within the merge section. Congestion appears whenever the combined mainline and on-ramp demands exceed the freeway capacity. The point at which congestion begins is called the *bottleneck*. It is typically somewhere within the merge section or nearby downstream. From this point, congestion propagates upstream, while the portion of freeway downstream of the bottleneck remains uncongested. The geometric features represented in Figure 4.1 suggest a division of the section into three zones: `A`, `B` and `C`, separated respectively by the on-ramp junction and the bottleneck. One of the issues addressed in this study relates to the placement of the feedback sensor on the mainline. Putting it in zone `C` is an obviously bad choice, since the flow out of the bottleneck becomes a constant whenever zone `B` becomes congested. Thus, the density of the freeway upstream of the bottleneck is *unobservable* from zone `C` during congestion. If the sensor is placed in zone `B`, it gains observability in the congested regime. However, as it is moved upstream through zone `B`, a delay appears between the onset of congestion and its detection by the sensor. With the sensor placed in zone `A`, the mentioned delay is large and the density in zones `B` and `C` are unobservable when traffic is in free-flow. Unlike the linear models used in some previous studies (e.g. [79]), the ACTM adequately captures the difference between the congested and

uncongested traffic states, and the dilemma it creates for sensor placement.

The study is based on a *discontinuous* form of the ACTM, which captures the effect of *capacity drop* not present in the original version. This modification is necessary for the analysis, since travel time cannot be reduced in the single-destination setup by avoiding the blockage of freeway offramps.

# 4.1   The discontinuous ACTM

As was described in Section 2.2.1, the term *capacity drop* refers to the decrease in flow measured downstream of a bottleneck after congestion initiates [2, 13, 36]. It is caused by the fact that vehicles cannot accelerate instantaneously to their desired speed once they exit the congested region. Instead, the speed profile downstream of the bottleneck increases gradually, from the reduced speed to the free-flow speed. As vehicles move through this profile, the separation between them gets larger. Whether a capacity drop exists or not depends on the size of the inter-vehicular gaps once they reach the free-flow speed.

The reduced flow rate downstream of the congested bottleneck is known as the *queue discharge rate*. The discontinuous ACTM used in this study mimics the capacity drop phenomenon by introducing a queue discharge rate parameter $(f_d)$ that is less than the maximum flow during free-flow. The model differs from the original ACTM only in its specification of mainline flow:

$$
f_i[k] = \begin{cases}
\bar{\beta}_i[k]\, v_i\big(\rho_i[k] + \delta_i \gamma_i r_i[k]\big) & \text{if } \begin{cases} i \text{ Uncongested} \\ i+1 \text{ Uncongested} \end{cases} \\[2ex]
\min\{\ \bar{\beta}_i[k]\, v_i\big(\rho_i[k] + \delta_i \gamma_i r_i[k]\big)\ ; \\
\quad w_i\big(\bar{\rho}_{i+1} - \rho_{i+1}[k]\big) - \delta_{i+1}\alpha_{i+1}r_{i+1}[k]\ \} & \text{if } \begin{cases} i \text{ Uncongested} \\ i+1 \text{ Congested} \end{cases} \\[2ex]
f_{d,i} & \text{if } \begin{cases} i \text{ Congested} \\ i+1 \text{ Uncongested} \end{cases} \\[2ex]
w_i\big(\bar{\rho}_{i+1} - \rho_{i+1}[k]\big) - \delta_{i+1}\alpha_{i+1}r_{i+1}[k] & \text{if } \begin{cases} i \text{ Congested} \\ i+1 \text{ Congested} \end{cases}
\end{cases}
\tag{4.1}
$$

In Eq. (4.1), a section $i$ is considered to be "Congested" when its density exceeds a critical value $\rho_{c,i}$ defined by:

$$
v_i\, \rho_{c,i} \;=\; w_i(\bar{\rho}_i - \rho_{c,i})
\tag{4.2}
$$

Otherwise, it is "Uncongested". Equation (4.1) is identical to the original Eq. (3.8) (removing the $\bar{s}_i$ term) in all but the $i$ Congested / $i+1$ Uncongested regime, where it equals $f_{d,i}$. This mainline flow function is allowed to be discontinuous whenever $f_{d,i} < v_i\rho_{c,i}$. The min{} rule of the original ACTM, on the other hand, ensures continuity of its flow function. Figure 4.2 compares the shapes of these two functions in the case $\alpha_i = \gamma_i = 0$. The other three components of the model remain unchanged (Eqs. (3.12), (3.13) and (3.14)).

## 4.1.1   Single on-ramp configuration

The study site is a stretch of freeway containing a single metered on-ramp, as shown in Figure 4.3. It is partitioned into 4 sections, numbered 0 through 3, with the on-ramp connected to
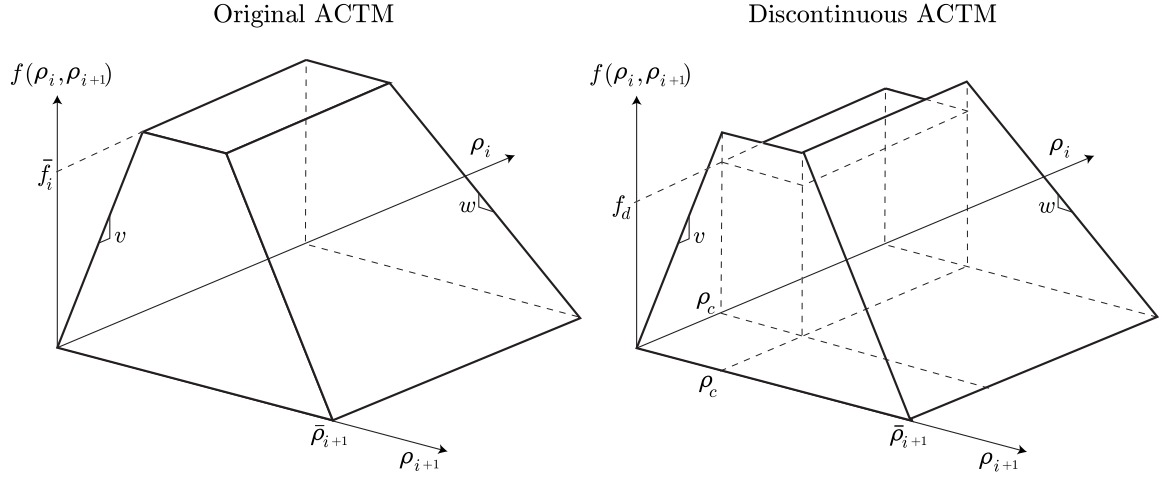
Figure 4.2: Single on-ramp configuration.

section 2.
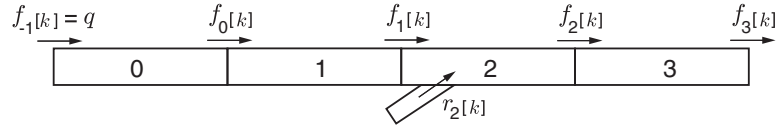


Figure 4.3: Single on-ramp configuration.

The following assumptions are made:

1. The parameters of the model are uniform along the freeway. The section indices will therefore be dropped ($v_i \to v$, $w_i \to w$, etc.).

2. $\gamma_2 = 0$ , $\alpha_2 < 1$

3. Sections 0 and 3 remain *uncongested* at all times (i.e. $\rho_i[k] < \rho_c$ for $i = 0, 3$). This amounts to a restriction on the magnitude of the upstream flow $f_{-1}[k]$ and on the state of the freeway downstream of section 3. It is required in order to establish boundary conditions.

4. $f_{-1}[k]$ is constant: $f_{-1}[k] = q$.

5. $q < f_d < v\rho_c$

Given the uncongested state in sections 0 and 3, the system as a whole can at any time be in one of four *congestion modes*: UU, CU, UC and CC, where the first letter denotes the state (Uncongested or Congested) of section 1, and the second of section 2. Expressions for $f_0[k]$, $f_1[k]$, and $f_2[k]$ in each of the congestion modes are provided in Table 4.1. A total of six different combinations of expressions for the flows can occur. Each combination is referred to as a *dynamic mode*, and denoted I through VI. The dynamic modes are given, along with their associated congestion modes in Table 4.2. Notice that dynamic mode V appears

in congestion modes `UC` and `CC`. Also, that different dynamic modes are possible within the same congestion mode; for example congestion mode `CU` can evolve according to dynamic mode II or III.

| Congestion mode | $f_{0[k]} =$ | $f_{1[k]} =$ | $f_{2[k]} =$ |
|---|---|---|---|
| UU | $v\rho_{0[k]}$ | $v\rho_{1[k]}$ | $v\rho_{2[k]}$ |
| CU | $\min\left\{v\rho_{0[k]}, w(\bar{\rho} - \rho_{1[k]})\right\}$ | $f_d$ | $v\rho_{2[k]}$ |
| UC | $v\rho_{0[k]}$ | $\min\left\{v\rho_{1[k]}, w(\bar{\rho} - \rho_{2[k]}) - \alpha r_{[k]}\right\}$ | $f_d$ |
| CC | $\min\left\{v\rho_{0[k]}, w(\bar{\rho} - \rho_{1[k]})\right\}$ | $w(\bar{\rho} - \rho_{2[k]}) - \alpha r_{[k]}$ | $f_d$ |

Table 4.1: Four congestion modes.

| Dynamic mode | Congestion mode | $f_{0[k]} =$ | $f_{1[k]} =$ | $f_{2[k]} =$ |
|---|---|---|---|---|
| I | UU | $v\rho_{0[k]}$ | $v\rho_{1[k]}$ | $v\rho_{2[k]}$ |
| II | CU | $v\rho_{0[k]}$ | $f_d$ | $v\rho_{2[k]}$ |
| III | CU | $w(\bar{\rho} - \rho_{1[k]})$ | $f_d$ | $v\rho_{2[k]}$ |
| IV | UC | $v\rho_{0[k]}$ | $v\rho_{1[k]}$ | $f_d$ |
| V | UC/CC | $v\rho_{0[k]}$ | $w(\bar{\rho} - \rho_{2[k]}) - \alpha r_{[k]}$ | $f_d$ |
| VI | CC | $w(\bar{\rho} - \rho_{1[k]})$ | $w(\bar{\rho} - \rho_{2[k]}) - \alpha r_{[k]}$ | $f_d$ |

Table 4.2: Six dynamic modes.

The other component of the single on-ramp model is the principle of vehicle conservation for each of the four sections. Because section 3 does not affect the stability analysis, it will not be considered further. The conservation equations are:

$$
\begin{aligned}
\rho_{2[k+1]} &= \rho_{2[k]} + f_{1[k]} - f_{2[k]} + r_{[k]} \\
\rho_{1[k+1]} &= \rho_{1[k]} + f_{0[k]} - f_{1[k]} \\
\rho_{0[k+1]} &= \rho_{0[k]} + q - f_{0[k]}
\end{aligned}
\tag{4.3}
$$

Equation (4.3) and Table 4.2 constitute a complete model for the single on-ramp configuration. The model can be understood as a hybrid system switching among 7 *finite states*: `UU-I`, `CU-II`, `CU-III`, `UC-IV`, `UC-V`, `CC-V`, and `CC-VI`, and evolving within each one according to a discrete-time linear dynamics. Each of the dynamic modes are cast in a state-space form,

$$
\rho_{[k+1]} = A^x \rho_{[k]} + B^x r_{[k]} + W^x
\tag{4.4}
$$

with $\rho_{[k]} = [\rho_{2[k]}, \rho_{1[k]}, \rho_{0[k]}]^T$, $x = $ I . . . VI, and matrices $A^x$, $B^x$, and $W^x$ given in Table 4.3. Table 4.4 provides open-loop transfer functions and eigenvalues for the six dynamic modes.

| $x$ | Congestion mode | $A^x$ | $B^x$ | $W^x$ |
|---|---|---|---|---|
| I | UU | $\begin{bmatrix} 1-v & v & 0 \\ 0 & 1-v & v \\ 0 & 0 & 1-v \end{bmatrix}$ | $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ | $\begin{bmatrix} 0 \\ 0 \\ q \end{bmatrix}$ |
| II | CU | $\begin{bmatrix} 1-v & 0 & 0 \\ 0 & 1 & v \\ 0 & 0 & 1-v \end{bmatrix}$ | $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ | $\begin{bmatrix} f_d \\ -f_d \\ q \end{bmatrix}$ |
| III | CU | $\begin{bmatrix} 1-v & 0 & 0 \\ 0 & 1-w & 0 \\ 0 & w & 1 \end{bmatrix}$ | $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ | $\begin{bmatrix} f_d \\ w\bar{\rho} - f_d \\ q - w\bar{\rho} \end{bmatrix}$ |
| IV | UC | $\begin{bmatrix} 1 & v & 0 \\ 0 & 1-v & v \\ 0 & 0 & 1-v \end{bmatrix}$ | $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ | $\begin{bmatrix} -f_d \\ 0 \\ q \end{bmatrix}$ |
| V | UC/CC | $\begin{bmatrix} 1-w & 0 & 0 \\ w & 1 & v \\ 0 & 0 & 1-v \end{bmatrix}$ | $\begin{bmatrix} 1-\alpha \\ \alpha \\ 0 \end{bmatrix}$ | $\begin{bmatrix} w\bar{\rho} - f_d \\ -w\bar{\rho} \\ q \end{bmatrix}$ |
| VI | CC | $\begin{bmatrix} 1-w & 0 & 0 \\ w & 1-w & 0 \\ 0 & w & 1 \end{bmatrix}$ | $\begin{bmatrix} 1-\alpha \\ \alpha \\ 0 \end{bmatrix}$ | $\begin{bmatrix} w\bar{\rho} - f_d \\ 0 \\ q - w\bar{\rho} \end{bmatrix}$ |

Table 4.3: State-space representations.

## 4.1.2 Controlability and Observability

The observability and controlability properties of the model determine whether the freeway can be effectively controlled by a suitable on-ramp metering strategy. Here we investigate how the position of the feedback sensor affects these properties. It is assumed that the sensor is located either in zone A or in zone B of Figure 4.1, by considering the output of the model to be either $\rho_1$ or $\rho_2$. The observability of each of the 6 dynamic modes is dictated by observability matrices $O_1^x$ and $O_2^x$ ($x=$I ... VI),

$$O_1^x = \begin{bmatrix} C_1 \\ C_1 A^x \\ C_1 (A^x)^2 \end{bmatrix} \qquad O_2^x = \begin{bmatrix} C_2 \\ C_2 A^x \\ C_2 (A^x)^2 \end{bmatrix}$$

with $C_1 = [0, 1, 0]$ and $C_2 = [1, 0, 0]$, corresponding to measurements of $\rho_1$ and $\rho_2$ respectively. Similarly, controlability depends on the rank of controlability matrices $P^x = [B^x, A^x B^x, (A^x)^2 B^x]$. The values of $O_1^x$, $O_2^x$, and $P^x$ can be found in Appendix A.1. Of these, only $P^{VI}$, $O_1^V$, $O_2^I$, and $O_2^{IV}$ have full rank. For the rest, only a subset of the states can be controlled from the on-ramp or their values reconstructed from the measurements. The sets of state variables that can be controlled and/or reconstructed in each dynamic mode are shown in Table 4.5.

Notice in the table that the on-ramp can influence the upstream portion of the freeway

| Dynamic mode | Congestion mode | $r \to \rho_0$ | $r \to \rho_1$ | $r \to \rho_2$ | Open-loop eigenvalues |
|---|---|---|---|---|---|
| I | UU | 0 | 0 | $\frac{1}{z-1+v}$ | $z = \{1-v, 1-v, 1-v\}$ |
| II | CU | 0 | 0 | $\frac{1}{z-1+v}$ | $z = \{1-v, 1-v, 1\}$ |
| III | CU | 0 | 0 | $\frac{1}{z-1+v}$ | $z = \{1-v, 1-w, 1\}$ |
| IV | UC | 0 | 0 | $\frac{1}{z-1}$ | $z = \{1-v, 1-v, 1\}$ |
| V | UC/CC | 0 | $\frac{\alpha(z-1+w/\alpha)}{(z-1)(z-1+w)}$ | $\frac{1-\alpha}{z-1+w}$ | $z = \{1-v, 1-w, 1\}$ |
| VI | CC | $\frac{\alpha w(z-1+w/\alpha)}{(z-1)(z-1+w)^2}$ | $\frac{\alpha(z-1+w/\alpha)}{(z-1+w)^2}$ | $\frac{1-\alpha}{z-1+w}$ | $z = \{1-w, 1-w, 1\}$ |

Table 4.4: Open loop transfer functions.

| Dynamic mode | Congestion mode | Controlable with $r$ | Reconstructable from of $\rho_1$ | Reconstructable from of $\rho_2$ |
|---|---|---|---|---|
| I | UU | $\{\rho_2\}$ | $\{\rho_0, \rho_1\}$ | $\{\rho_0, \rho_1, \rho_2\}$ |
| II | CU | $\{\rho_2\}$ | $\{\rho_0, \rho_1\}$ | $\{\rho_2\}$ |
| III | CU | $\{\rho_2\}$ | $\{\rho_1\}$ | $\{\rho_2\}$ |
| IV | UC | $\{\rho_2\}$ | $\{\rho_0, \rho_1\}$ | $\{\rho_0, \rho_1, \rho_2\}$ |
| V | UC/CC | $\{\rho_1, \rho_2\}$ | $\{\rho_0, \rho_1, \rho_2\} \quad w \neq v$ <br> $\{\rho_1, \rho_0 + \rho_2\} \quad w = v$ | $\{\rho_2\}$ |
| VI | CC | $\{\rho_0, \rho_1, \rho_2\}$ | $\{\rho_1, \rho_2\}$ | $\{\rho_2\}$ |

Table 4.5: Controlable and uniquely recontructable states in each dynamic mode.

($\rho_0$ and $\rho_1$) only in dynamic modes V and VI, when the downstream section is severely congested. Otherwise, in modes I through IV, its influence travels downstream. Also, several observations related to the difference between upstream and downstream detector placements can be made. In the case of an upstream detector ($\rho_1$), the downstream state can only be deduced if section 2 is *congested*, in modes V and VI. Conversely, the state upstream of the on-ramp is only known to a downstream detector if section 1 is *uncongested*, in modes I and IV. In both cases, reconstructability of the other is lost in modes II and III (i.e. the CU congestion mode).

### 4.1.3 Finite state transitions

Part of the control objective described in the next section is to drive the system to the UU-I final state in a fairly direct way. The analysis is simplified considerably by reducing the number of possible transitions with an additional assumption. Figure 4.5 shows a state transition diagram with all 42 possible transitions among the 7 finite states. This number is reduced to a more manageable 16 by assuming that the time step is small enough (i.e. the controller frequency is fast enough) so that certain transitions can be disregarded for being highly improbable. Figure 4.4 illustrates typical positions of the section densities on a fundamental diagram, in each of the 7 finite states. The assumption of a small time step
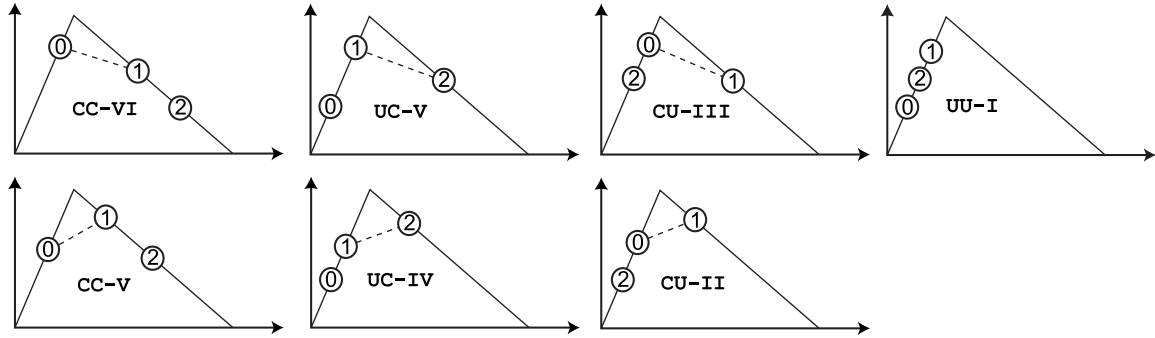
Figure 4.4: Relative section densities in the 7 finite states.
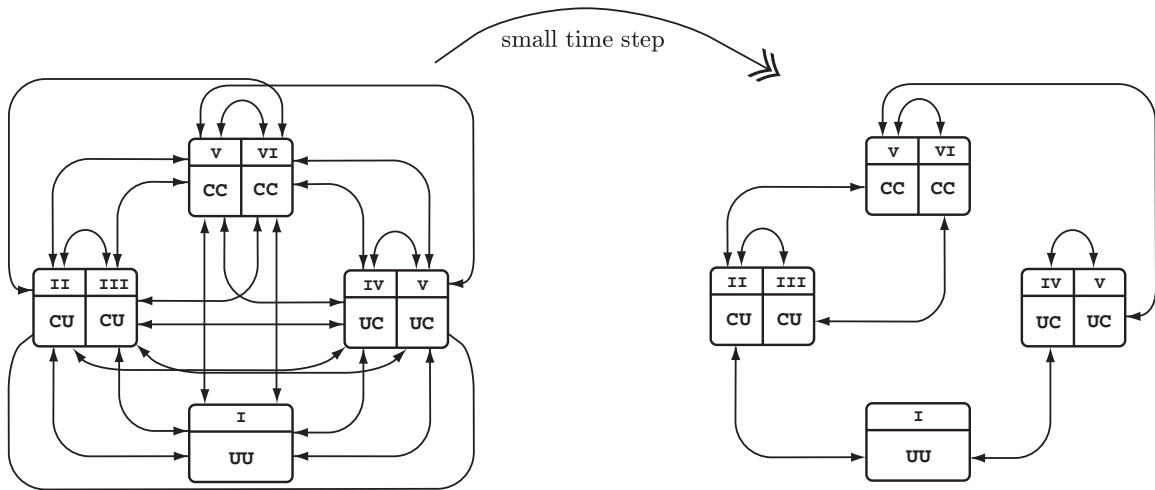


Figure 4.5: Simplification of the finite state transition diagram.

implies that these points should move in a continuous fashion. This leads to the following two types of eliminations:

1. Transitions such as `UU-I`→`CC-VI` and `CU-II`→`UC-IV`, where the congestion state of sections 1 and 2 change simultaneously.

2. Transitions such as `UC-V`→`UU-I`, where the continuous movement of the state would force it to go through a different intermediate state, in this case `UC-IV`.

There are 12 transitions in the first category and 14 in the second. The total number possible transitions is thus reduced from 42 to 16.

## 4.2 Control objective

The main goal of this study is to provide suggestions for the selection of control parameters for Alinea and %-Occ. A typical criterion for parameter selection is closed-loop stability about a desired equilibrium state. However, to prove global asymptotic stability for the hybrid model is difficult. Instead, we will seek parameters that result in a "good strategy",

in the sense that it eliminates congestion and converges to a steady-state flow that is larger than the queue discharge rate. A "good strategy" is defined as one that:

1. Stabilizes the closed-loop transfer function from the on-ramp flow to the measurement, in all dynamic modes where the corresponding open-loop transfer function is non-zero.

2. Never gets stuck in any mode other than `UU-I`, and when in another finite state, it tends to move closer to `UU-I`. This informal statement is made concrete with the conditions of Table 4.6.

3. Has a steady-state flow in `UU-I` exceeding the queue discharge rate. That is, in `UU-I`, $v\,\rho_{2ss} > f_d$. This is to ensure that the flow out of the test section is larger with control than without it. It is assumed that if no control is applied, the freeway section will become congested and $f_2$ will converge to the queue discharge rate $f_d$. Thus, a strategy that does not meet this condition will actually be detrimental to travel time.

| Dynamic mode | Congestion mode | Condition | | |
|---|---|---|---|---|
| I | UU | $\{\rho_{1ss} < \rho_c\}$ | AND | $\{f_d/v < \rho_{2ss} < \rho_c\}$ |
| II | CU | $\{\rho_{1ss} < \rho_c\}$ | AND | $\{\rho_{2ss} < \rho_c\}$ |
| III | CU | $\{\rho_{0ss} < w(\bar{\rho} - \rho_{1ss})\}$ | AND | $\{\rho_{2ss} < \rho_c\}$ |
| IV | UC | $\{\rho_{1ss} < \rho_c\}$ | AND | $\{\rho_{2ss} < \rho_c\}$ |
| V | UC/CC | $\{\rho_{1ss} < \rho_c\}$ | AND | $\{v\rho_{1ss} < w(\bar{\rho} - \rho_{2ss}) - \alpha r_{ss}\}$ |
| VI | CC | $\{v\rho_{0ss} < w(\bar{\rho} - \rho_{1ss})\}$ | OR | $\{\rho_{2ss} < \rho_c\}$ |

Table 4.6: Steady-state requirements for a "good strategy".

Each of the conditions on the steady-state values expressed in Table 4.6 tends to push the system out of its current finite state and into another that is closer to `UU-I` in the following ordering: `CC-VI`, `CC-V`, `UC-V`, `UC-IV`, `CU-III`, `CU-II`, `UU-I`. Note the this definition of a "good strategy" does not imply global asymptotic stability, nor does global asymptotic stability require the strategy to be good. For example, a good strategy may still get caught in an infinite loop between two finite states, due to oscillations. Conversely, a globally stable controller may take a more convoluted trajectory to `UU-I`, and thus not be considered as "good".

## 4.3  Percent-Occupancy control

Percent-occupancy (%-Occ) control is one of the most widespread on-ramp metering schemes in the U.S. due to its simplicity of implementation and observed effectiveness. Along with Alinea, it falls under the category of local traffic-responsive controllers, since the rate at every meter is based on local real-time measurements (i.e. no coordination). It is distinguished from Alinea by the fact that the feedback sensor is placed *upstream* of the on-ramp

junction. %-Occ can be interpreted as a proportional control of the *occupancy* measurement. Occupancy is defined as the portion of time during which a loop detector registers a vehicle presence. It is closely related to the density states of the model used here. The metering rate under %-Occ control is computed with:

$$r_{[k]} = K_1 - K_2 \, \rho_{1[k]} \tag{4.5}$$

where $K_1$ and $K_2$ are tunable controller parameters. The closed-loop dynamics under %-Occ control is given by:

$$\rho_{[k+1]} = A^x_{occ} \rho_{[k]} + B^x_{occ} \tag{4.6}$$

with $A^x_{occ} = A^x - K_2 B^x C_1$ and $B^x_{occ} = W^x + K_1 B^x$ ($x = \text{I} \ldots \text{VI}$). Specific expressions for $A^x_{occ}$ and $B^x_{occ}$ are provided in Appendix A.2.

We derive conditions on $K_1$ and $K_2$ under which %-Occ qualifies as a "good strategy", according to the definition of Section 4.2. To illustrate the methodology, consider dynamic mode $\text{V}$, with closed-loop dynamics:

$$
\begin{aligned}
\rho_{[k+1]} &= \begin{bmatrix} 1-w & 0 & 0 \\ w & 1 & v \\ 0 & 0 & 1-v \end{bmatrix} \rho_{[k]} + \begin{bmatrix} 1-\alpha \\ \alpha \\ 0 \end{bmatrix} (K_1 - K_2 [0\ 1\ 0]\rho_{[k]}) + \begin{bmatrix} w\bar{\rho} - f_d \\ -w\bar{\rho} \\ q \end{bmatrix} \\
&= \begin{bmatrix} 1-w & -(1-\alpha)K_2 & 0 \\ w & 1-\alpha K_2 & v \\ 0 & 0 & 1-v \end{bmatrix} \rho_{[k]} + \begin{bmatrix} (1-\alpha)K_1 + w\bar{\rho} - f_d \\ \alpha K_1 - w\bar{\rho} \\ q \end{bmatrix}
\end{aligned}
$$
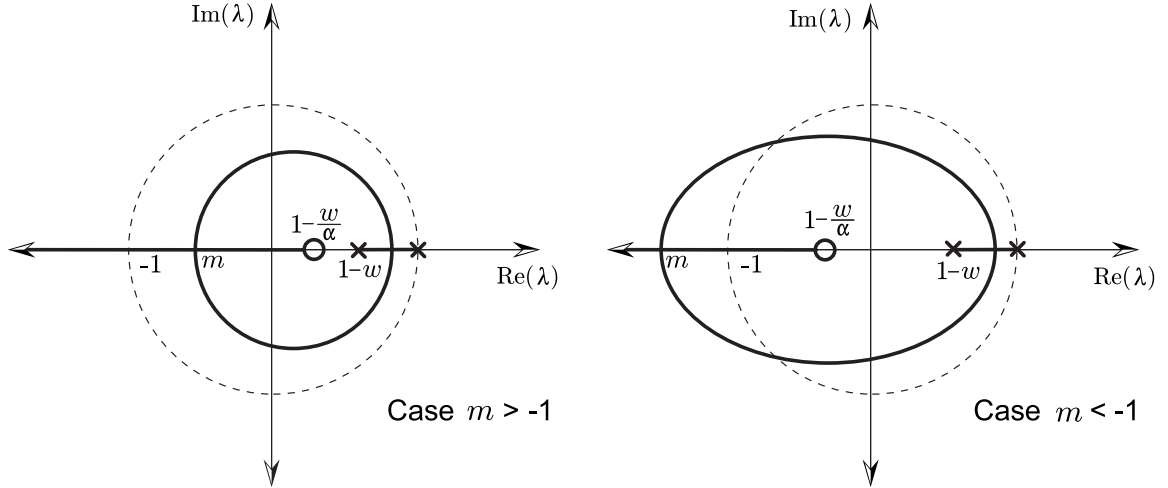


Figure 4.6: Root locus diagram for dynamic mode $\text{V}$ with %-Occ.

A root locus diagram for mode $\text{V}$ is shown in Figure 4.6. For the controller to meet the specifications of Section 4.2, the parameters must be chosen such that the closed loop eigenvalues are within the unit circle, and the stable equilibrium point is as required by Table 4.6. To derive stable gains, we first separate the problem into two cases, depending on whether the

intersection of the root locus with the real axis (point $m$ in Figure 4.6) falls inside or outside of the unit circle. The characteristic polynomial for mode V is:

$$\lambda^2 + (\alpha K_2 + w - 2)\lambda + 1 - w + (w - \alpha)K_2$$

The point $m$ is defined by the occurrence of a repeated real eigenvalue at $\lambda = m$. Then:

$$-2m = \alpha K_2 + w - 2$$
$$m^2 = 1 - w + (w - \alpha)K_2$$

Which leads to:

$$m = 1 - \frac{w}{\alpha} - \frac{w}{\alpha}\sqrt{1 - \alpha}$$

The two cases, $m > -1$ and $m < -1$, are thus translated into conditions on $w$ and $\alpha$:

$$m > -1 \quad \Leftrightarrow \quad 4\alpha - 4w + w^2 > 0$$

In the case $m > -1$, stability is lost when an eigenvalue crosses the -1 point on the real axis. In the $m < -1$ case, both eigenvalues exit the unit circle simultaneously. In both cases, $K_2$ is restricted by an upper bound. The derivation of the upper bounds for $K_2$ in dynamic modes V and VI can be found in Appendix A.2. For dynamic mode V, the upper bound is:

$$\bar{K}_2 = \begin{cases} \frac{2(2-w)}{2\alpha-w} & \text{if } 4\alpha - 4w + w^2 > 0 \\ \frac{w}{w-\alpha} & \text{else} \end{cases} \tag{4.7}$$

With $K_2$ thus chosen, the steady-state equilibrium point is,

$$\rho_{ss} = \begin{bmatrix} \rho_{2ss} \\ \rho_{1ss} \\ \rho_{0ss} \end{bmatrix} = \begin{bmatrix} \bar{\rho} - \frac{1-\alpha}{w}q - \frac{\alpha}{w}f_d \\ \frac{1}{K_2}(K_1 - f_d + q) \\ \frac{1}{v}q \end{bmatrix} \tag{4.8}$$

This steady-state is subject to the conditions from Table 4.6: $\{\rho_{1ss} < \rho_c\}$ and $\{v\rho_{1ss} < w(\bar{\rho} - \rho_{2ss}) - \alpha r_{ss}\}$. Using the values from Eq. (4.8), and after some manipulation, these conditions become:

$$K_1 - \rho_c K_2 < f_d - q \tag{4.9}$$
$$K_1 - (q/v)K_2 < f_d - q \tag{4.10}$$

Given Eq. (4.10), Eq. (4.9) is always satisfied, since $\rho_c > q/v$ by assumption. Therefore, in summary, $K_1$ and $K_2$ must satisfy Eqs. (4.7) and (4.10) for %-Occ to qualify as a "good strategy".

A similar analysis was carried out for the remaining modes. Table 4.7 provides the ranges of stable gains for the two modes with non-zero open loop transfer functions. Table 4.8 shows the long term behavior for the six dynamic modes. In the table, the ↗ symbol indicates unstable states whose values increase without bound, and ↘ indicates states that decrease without bound. Also, labels [IV-a] and [VI-a] in Table 4.8 are constraints on the values of $K_1$ and $K_2$ defined in Table 4.9.

| Dynamic mode | Congestion mode | $\bar{K}_2$ | |
|---|---|---|---|
| V | UC/CC | $\frac{2(2-w)}{2\alpha-w}$ $\frac{w}{w-\alpha}$ | if $4\alpha - 4w + w^2 > 0$ else |
| VI | CC | $\frac{(2-w)^2}{2\alpha-w}$ $\frac{w(2-w)}{w-\alpha}$ | if $2\alpha - 2w + w\alpha > 0$ else |

Table 4.7: Maximum stable gains for %-Occ.

| Dynamic mode | Congestion mode | Long term behavior | | |
|---|---|---|---|---|
| | | $\rho_{0ss}$ | $\rho_{1ss}$ | $\rho_{2ss}$ |
| I | UU | $\frac{1}{v}q$ | $\frac{1}{v}q$ | $\frac{1}{v}[(1 - \frac{1}{v}K_2)q + K_1]$ |
| II | CU | $\frac{1}{v}q$ | $\searrow$ | $\nearrow$ |
| III | CU | $\searrow$ | $\bar{\rho} - \frac{1}{w}f_d$ | $\frac{1}{v}[K_1 - K_2(\bar{\rho} - \frac{1}{w}f_d) + f_d]$ |
| IV | UC | $\frac{1}{v}q$ | $\frac{1}{v}q$ | $\searrow$ if [IV-a] $\nearrow$ else |
| V | UC/CC | $\frac{1}{v}q$ | $\frac{K_1-f_d+q}{K_2}$ | $\bar{\rho} - \frac{1-\alpha}{w}q - \frac{\alpha}{w}f_d$ |
| VI | CC | $\searrow$ if [VI-a] $\nearrow$ else | $\frac{K_1+w\bar{\rho}-f_d}{K_2+w}$ | $\frac{(w\bar{\rho}-f_d)(1+\frac{\alpha}{w}K_2)+(1-\alpha)K_1}{K_2+w}$ |

Table 4.8: Long term behavior with %-Occ.

Table 4.8 offers useful information for parameter selection. Perhaps the most important requirement for a "good strategy" is that the steady state flow in the uncongested regime must exceed the queue discharge rate without exceeding the maximum flow. Given that $\rho_{2ss} = \frac{1}{v}[(1 - \frac{1}{v}K_2)q + K_1]$ in mode I, these two requirements translate into conditions [I-a] and [I-b] in Table 4.9. Table 4.8 also shows that the upstream density in the CU-II mode decreases without bound, while the downstream density increases. Thus, from CU-II, any of the transitions to CU-III, CC-V, or UU-I are possible. The steady state values in modes III and V translate into conditions [III-a], [V-a], and [V-b] in Table 4.9. Notice that conditions [IV-a] and [V-a] are exactly contrary to condition [I-b]. Thus, the controller cannot simultaneously satisfy the steady state requirement for modes IV and V while increasing the flow in UU-I above the queue discharge rate.

Modes UC-IV and CC-VI involve densities that may either increase or decrease without bound, depending on the values of $K_1$ and $K_2$. In mode UC-IV, the preferred behavior for section 2 is for its density to decrease. This is guaranteed by condition [IV-a]. As was mentioned, this condition is in conflict with the important condition [I-b]. Thus, unbounded growth of $\rho_2$ in UC-IV cannot be avoided without sacrificing [I-b]. In CC-VI,

| Dynamic mode | Congestion mode | Requirements for $K_1$ and $K_2$ | Label |
|---|---|---|---|
| I | UU | $K_1 - (q/v)K_2 < v\rho_c - q$<br>$K_1 - (q/v)K_2 > f_d - q$ | [I-a]<br>[I-b] |
| II | CU | - | - |
| III | CU | $K_1 - (\bar{\rho} - f_d/w)K_2 < v\rho_c - f_d$ | [III-a] |
| IV | UC | $K_1 - (q/v)K_2 < f_d - q$ | [IV-a] |
| V | UC/CC | $K_1 - (q/v)K_2 < f_d - q$<br>$K_1 - \rho_c K_2 < f_d - q$ | [V-a]<br>[V-b] |
| VI | CC | $K_1 - (\bar{\rho} - q/w)K_2 < f_d - q$ | [VI-a] |

Table 4.9: Conditions on the %-Occ parameters.

the preferred behavior is for the density in section 0 to decrease, since this section has been assumed to remain uncongested. Condition [VI-a] is enforced. Notice that because $\bar{\rho} - (q/w) > \rho_c$, condition [VI-a] is satisfied whenever [V-b] must therefore be enforced.

In summary, it can be concluded that %-Occ does not qualify as a "good strategy" for several reasons. First, the steady state conditions outlined in Table 4.6 cannot be guaranteed in mode CU-II since both $\rho_2$ and $\rho_1$ diverge. This is not necessarily a major problem: although it is not guaranteed, the decreasing $\rho_1$ may in fact force a transition to UU-I before the increasing $\rho_2$ forces the transition to CC-V. A more troubling observation is the conflict between conditions [I-b], [IV-a], and [V-a]. No single selection of parameters can satisfy all of these conditions. Thus, the controller can either efficiently reject congestion, if [IV-a] and [V-a] are enforced, or increase uncongested flow, if [I-b] is enforced, but not both.

For the single-destination layout studied here, condition [I-b] is required in order to reduce travel time. In the next section we will discard [IV-a] and [V-a] in favor of [I-b], and use computer simulation to test the convergence of %-Occ for an array of initial conditions and parameter values.

### 4.3.1 Computer simulations with %-Occ

Figure 4.7 illustrates restrictions [I-a], [I-b], [III-a], [V-b], and [VI-a] from Table 4.9 in the $K_2/K_1$ plane. These lines and the simulations reported in this section were found with the following parameter values:

$$v = 0.7, \qquad \alpha = 0, \qquad f_d = 0.9\,v\rho_c, \qquad K_r = 0.3,$$
$$w = 0.3, \qquad \bar{\rho} = 10, \qquad q = 0.5\,f_d, \qquad \hat{\rho}_2 = 0.95\,v\rho_c$$

Even though a particular set of parameter values was used, the general shape of the plot applies to any valid set of parameters. This is because the assumptions $\rho_c < \bar{\rho}$ and $q < f_d$
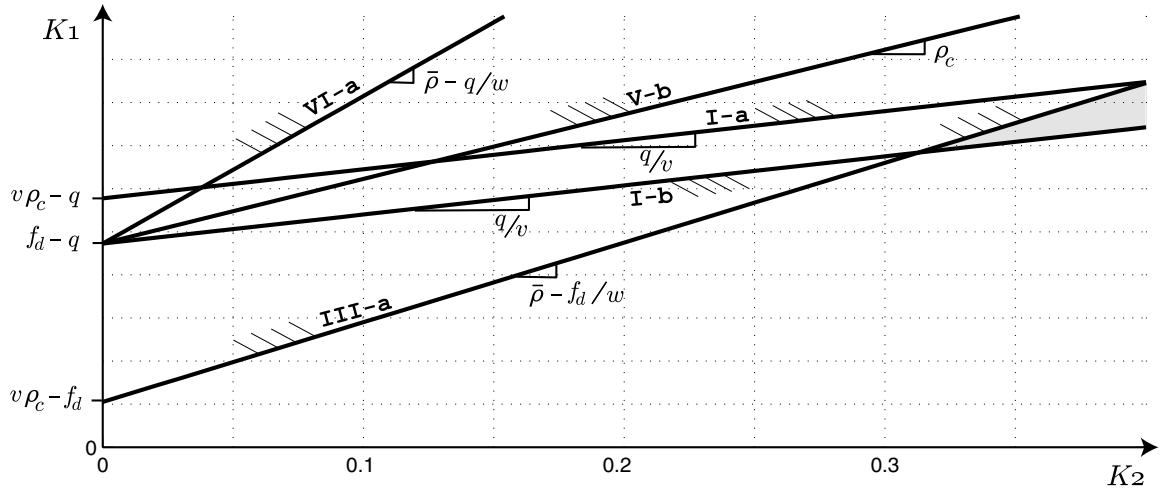
Figure 4.7: Parameter selection and convergence with %-Occ.

lead to the following relations amongst the slopes of the constraints and their intersections
with the $K_1$ axis:

$$(q/v) \ < \ \rho_c \ < \ \bar{\rho} - (f_d/w) \ < \ \bar{\rho} - (q/w)$$
$$v\rho_c - f_d \ < \ v\rho_c - q$$
$$f_d - q \ < \ v\rho_c - q$$

The only ambiguity in the figure is whether the intersection of [III-a] with the $K_1$ axis
occurs above or below $f_d - q$. Figure 4.8 focuses on the region where all of the restrictions
are satisfied. The maximum value of $K_2$, according to Table 4.7 is $\bar{K}_2 = 1.0$. The smallest
allowed value of $K_2$ is 0.31.
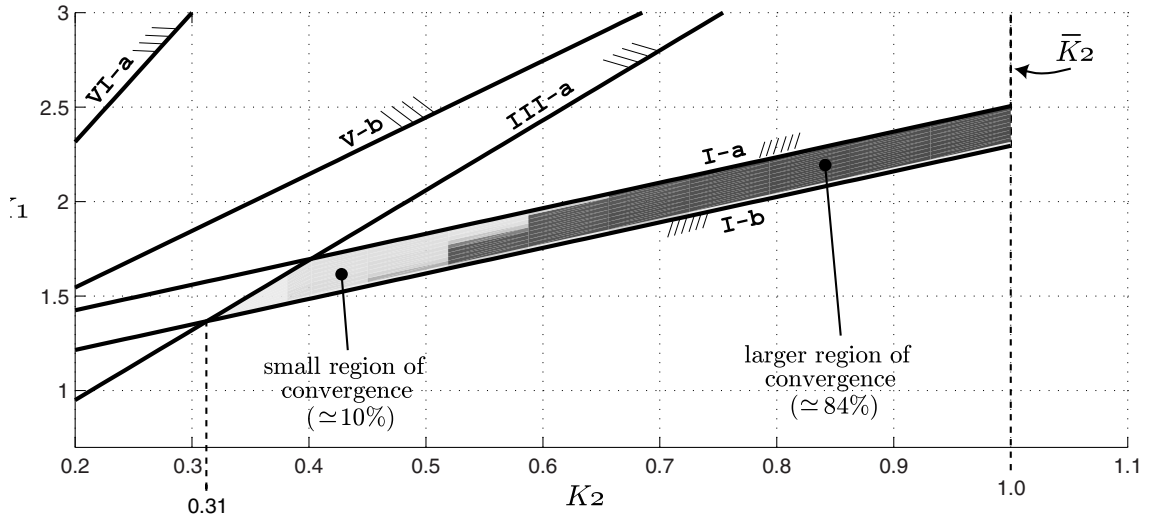


Figure 4.8: Parameter selection and convergence with %-Occ.

The response of the system under %-Occ control was simulated for $K_1/K_2$ pairs on an

irregular 11 X 12 grid covering the shaded region in Figure 4.8. For each of the $K_1/K_2$ pairs, 400 runs were conducted starting from initial conditions on a $20 \times 20$ grid of $\rho_{1[0]}$ and $\rho_{2[0]}$ values between 0 and $\bar{\rho}$. Each run recorded whether or not the desired final condition, with $\rho_1 < \rho_c$ and $f_d/v < \rho_2 < \rho_c$, was reached. The portion of the 400 initial conditions that reached the desired final state was determined for each of the $K_1/K_2$ pairs. This region of convergence is represented in Figure 4.8 in shades of gray. The darker shades represent larger portions of successful runs; i.e. larger regions of convergence. It should be noted that the largest region of convergence, recorded for $K_1 = 2.13$ and $K_2 = 0.86$, covered only 84% percent of the grid of initial conditions. That is, no selection of parameters was found to improve mainline flow and travel time for more than 84% of all possible initial conditions.

## 4.4 Alinea control

The Alinea control law was first introduced by Papageorgiou and others in [79]. It has since been tested in several European cities [35, 77], where it has performed favorably as compared to %-Occ, or a variant of %-Occ since a downstream measurement was used. It is interpreted as integral control of the *downstream* occupancy measurement. Again relating occupancy to density, the control law for Alinea is:

$$r_{[k]} \; = \; r_{[k-1]} \; + \; K_r \left( \hat{\rho}_2 - \rho_{2[k]} \right) \tag{4.11}$$

The two tunable parameters are $\hat{\rho}_2$, the *target density*, and $K_r$, the *feedback gain*. It is assumed that $K_r > 0$ and $\hat{\rho}_2 < \rho_c$. We apply the same analytical procedure to Alinea as to %-Occ. Table 4.10 provides stable ranges of $K_r$ and the resulting steady-states. In almost all cases, Alinea stabilizes the downstream density about $\hat{\rho}_2$, with mode IV being the only exception. The $\odot$ symbol in Table 4.10 in mode IV is meant to indicate that the density error $\tilde{\rho}_2 = \rho_2 - \hat{\rho}_2$ behaves like an undamped oscillator (see Appendix A.3):

$$\tilde{\rho}_{2[k+2]} \; + \; \left( K_r - 2 \right) \tilde{\rho}_{2[k+1]} \; + \; \tilde{\rho}_{2[k]} = 0 \tag{4.12}$$

This, strictly speaking, violates the first requirement of a "good strategy", that the controller should stabilize all dynamic modes with non-zero open-loop transfer functions. It also introduces the possibility of an infinite loop between UC-IV and UC-V (see Figures 4.4 and 4.5). Notice however that the loop is always broken after at most one cycle, since $\rho_2$ oscillates about $\hat{\rho}_2$ in UC-IV, and converges to $\hat{\rho}_2$ in UC-V. Hence, if the system transitions from UC-IV to UC-V, it will eventually come back, and then move through UC-IV to UU-I.

Aside from the marginal stability of mode IV, Alinea meets all of the requirements for a "good strategy". It can be easily verified that the long-term behavior in Table 4.10 is in compliance with Table 4.6, given that $q < f_d < v \rho_c$ and $\hat{\rho}_2 < \rho_c < \bar{\rho}$. Hence, summarizing Table 4.10, and applying the condition $f_{2ss} > f_d$, Alinea is considered a *nearly* "good strategy" whenever:

$$K_r \; \in \; \left[ \, 0 \, , \, \min \left( \, 2\frac{2-w}{1-\alpha} \, , \, 2(2-v) \right) \right] \tag{4.13}$$

$$\hat{\rho}_2 \; \in \; [ \, f_d/v \, , \, \rho_c \, ] \tag{4.14}$$

| Dynamic mode | Congestion mode | $K_r \in$ | Long term behavior | | |
|---|---|---|---|---|---|
| | | | $\rho_{2ss}$ | $\rho_{1ss}$ | $\rho_{0ss}$ |
| I | UU | $\left[\,0\,,\,2(2-v)\,\right]$ | $\hat{\rho}_2$ | $\frac{1}{v}q$ | $\frac{1}{v}q$ |
| II | CU | $\left[\,0\,,\,2(2-v)\,\right]$ | $\hat{\rho}_2$ | $\searrow$ | $\frac{1}{v}q$ |
| III | CU | $\left[\,0\,,\,2(2-v)\,\right]$ | $\hat{\rho}_2$ | $\bar{\rho} - \frac{1}{w}f_d$ | $\searrow$ |
| IV | UC | $\left[\,0\,,\,4\,\right]$ | $\odot$ | $\frac{1}{v}q$ | $\frac{1}{v}q$ |
| V | UC/CC | $\left[\,0\,,\,2\frac{2-w}{1-\alpha}\,\right]$ | $\hat{\rho}_2$ | $\searrow$ | $\frac{1}{v}q$ |
| VI | CC | $\left[\,0\,,\,2\frac{2-w}{1-\alpha}\,\right]$ | $\hat{\rho}_2$ | $\hat{\rho}_2 + \frac{\alpha}{1-\alpha}\left(\frac{1}{w}f_d - (\bar{\rho} - \hat{\rho}_2)\right)$ | $\searrow$ |

Table 4.10: Stable gains and Steady-states with Alinea.

## 4.4.1 Computer simulations with Alinea

Analogous to Section 4.3.1, exhaustive simulation is used here to investigate the performance of Alinea with varying initial conditions and parameter settings. The objective is to refine the recommendation for $K_r$ given in Eq. (4.13). Figure 4.9 shows a sample of the response of the discontinuous ACTM with Alinea, using the model parameters from Section 4.3.1. In this example, the system begins completely congested in the CC-VI finite state. Alinea quickly reduces the ramp metering rate, which eliminates congestion in the downstream section. The system passes through the CU-III and CU-II modes before reaching UU-I. The exiting flow then recovers and settles in the desired range, between the queue discharge rate and the maximum uncongested value. In this example, it took Alinea 18 steps to go from the congested initial condition to the desired final condition of no congestion and $f_2 > f_d$.
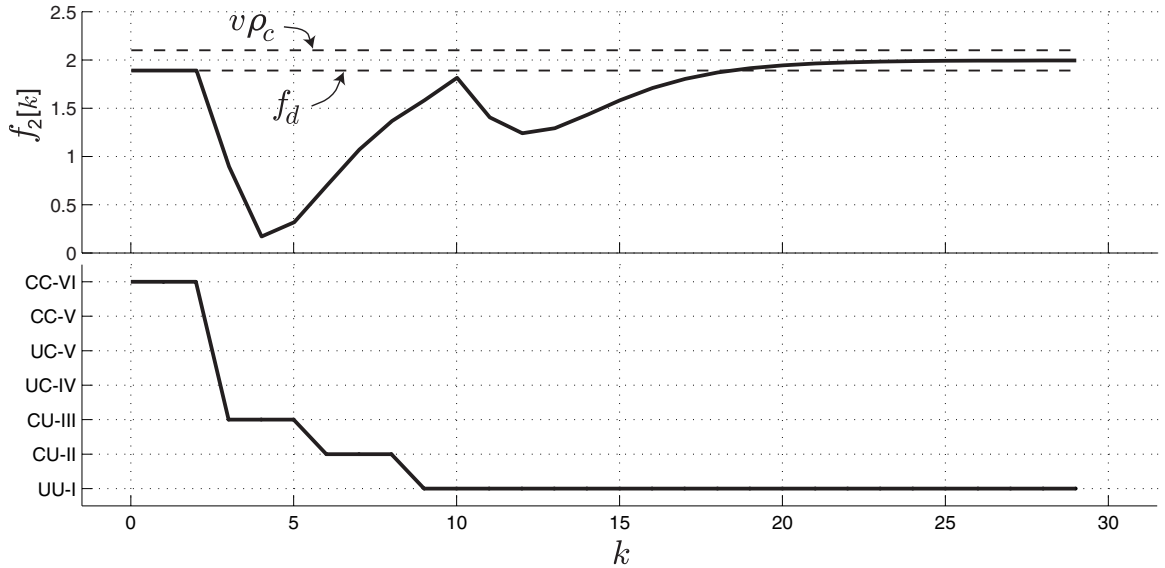


Figure 4.9: Downstream flow and congestion mode with Alinea.

The same experiment was repeated for all initial conditions on a $20 \times 20$ grid covering $\rho_{1[0]}$ and $\rho_{2[0]}$ values between 0 and $\bar{\rho}$. This was done for $K_r$ ranging from 0 to its maximum

value according to Eq. (4.13) of $K_r = 2.6$. For each simulation run, the number of steps it took for the system to reach the desired final state, if it was reached at all, was recorded. The average number of steps, over all initial conditions, was computed for each $K_r$, and is shown in Figure 4.10. Only those values of the gain for which *all* initial conditions reached the desired final state were recorded. Gains for which any initial condition failed to converge appear in the plot as 0 steps[1].

The plot shows that performance is fairly uniform for a range of $K_r$ values, as suggested by [35], but that there is clearly an optimal value of $K_r = 0.65$ that provides the quickest response. Figure 4.10 also illustrates a couple problems that were not considered in the definition of a "good strategy". Decreasing $K_r$ below 0.3 results in a sharp increase in the number of steps to convergence. Also, values above 1.7 produce an infinite loop between `UC-IV` and `UU-I`. This result is in agreement with the observation in [35] that high gains may lead to oscillatory behavior.
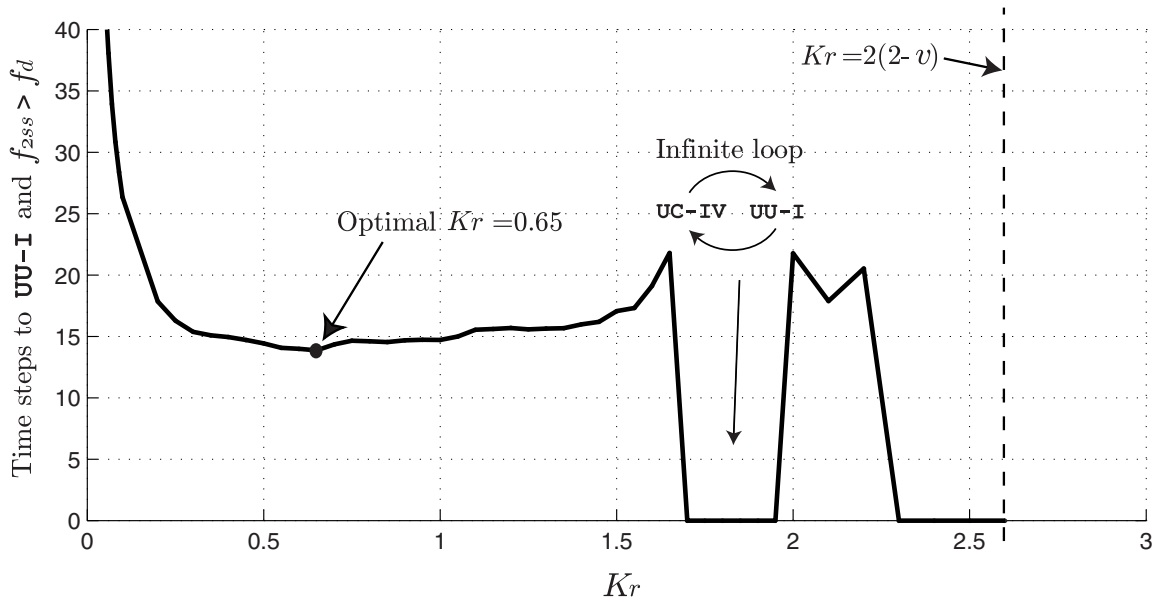


Figure 4.10: Performance of Alinea with varying $K_r$.

## 4.5 Summary

A discontinuous form of the ACTM was described in this chapter, and used to study two local on-ramp metering controllers. The observability and controllability characteristics of this model revealed that a detector placed upstream of the on-ramp may be more effective when the freeway is congested, whereas a downstream detector is preferred during free-flow. The model was also used to analyze two well-known on-ramp metering schemes, and to find ranges of parameter values that qualify them as "good strategies", meaning that

---

[1]The corresponding plot with %-Occ would be identically zero, since it always failed for some initial condition.

they eliminate congestion and increase mainline flow. It was found that neither qualify as a "good strategy", since they do not satisfy all of the criteria for any single choice of parameter values. In addition, exhaustive computer simulations with %-Occ showed that no selection of parameters achieves the stated goals for more than 84% of all initial conditions.

Although it did not strictly qualify, Alinea was found to meet most of the criteria for a "good strategy", whenever its parameters are chosen according to Eqs. (4.13) and (4.14). Only the marginal stability of mode IV prevented it from qualifying. Exhaustive simulation with Alinea showed that, within the range suggested by Eq. (4.13), a best choice of $K_r$ could be found the minimizes the convergence time, and presumably also minimizes total travel time.

# Chapter 5

# Optimal Control of the ACTM

The preceding chapter used a discontinuous form of the ACTM to study the performance of two local ramp metering strategies. The modification to the original continuous model was needed in order to allow the controllers to save travel time by avoiding the *capacity drop*. Alinea, which is based on this principle, was found capable of repelling congestion while increasing downstream flow above the queue discharge rate. This chapter describes the design of a *predictive* and *coordinated* strategy using the original version of the ACTM. Thus, the aim of this design is to avoid the loss of travel time related to *offramp blockage*.

The solution method described here differs from the previous efforts mentioned in Section 2.2.3 in that it does not attempt to directly solve the nonlinear optimization problem, but instead to construct and solve an equivalent *convex* problem. The construction of the convex problem is based on two facts. First, the nonlinearity of the ACTM (and of most freeway models) is in the fundamental diagram, which is a concave function. Hence, the set defined by all values of flow *below* this function is a convex set. The second fact, shown by Eq. (2.42), is that minimizing Total Travel Time is equivalent to maximizing a weighted sum of flows. The relaxed convex problem is obtained by replacing the equality constraint $f = F(\rho)$ with an inequality constraint $f \leq F(\rho)$, as shown in Figure 5.1. Because the travel time objective function favors larger flows, it is not unreasonable to expect the solution of the relaxed problem to "naturally" fall on the upper boundary. In this chapter we derive conditions that guarantee that it does.
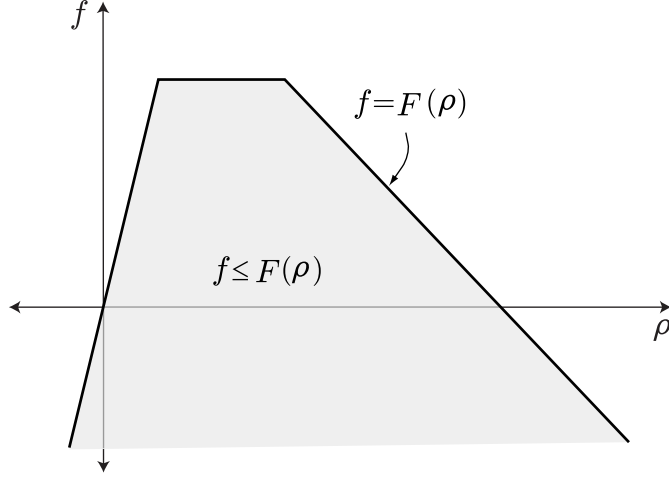
Figure 5.1: Concave fundamental diagram and a convex set.

The model equations of the original ACTM from Chapter 3 are reproduced below for reference here.

$$\rho_{i[k+1]} = \rho_{i[k]} + f_{i-1[k]} + \delta_i r_{i[k]} - f_{i[k]}/\bar{\beta}_{i[k]} \tag{5.1}$$

$$l_{i[k+1]} = l_{i[k]} + d_{i[k]} - r_{i[k]} \tag{5.2}$$

$$r_{i[k]} = \begin{cases} \min\left\{ l_{i[k]} + d_{i[k]} \; ; \; \xi_i(\bar{\rho}_i - \rho_{i[k]}) \right\} & \text{if } i \in \mathcal{E}n \setminus \mathcal{E}n^+ \\ \min\left\{ l_{i[k]} + d_{i[k]} \; ; \; \xi_i(\bar{\rho}_i - \rho_{i[k]}) \; ; \; r^c_{i[k]} \right\} & \text{if } i \in \mathcal{E}n^+ \end{cases} \tag{5.3}$$

$$f_{i[k]} = \min\left\{ \bar{\beta}_{i[k]} \, v_i(\rho_{i[k]} + \delta_i \gamma_i r_{i[k]}) \; ; \; w_{i+1}(\bar{\rho}_{i+1} - \rho_{i+1[k]}) - \delta_{i+1}\alpha_{i+1}r_{i+1[k]} \; ; \; \bar{f}_i \; ; \; \frac{\bar{\beta}_{i[k]}}{\beta_{i[k]}}\bar{s}_i \right\} \tag{5.4}$$

The definitions of both $r_{i[k]}$ and $f_{i[k]}$ in Eqs. (5.3) and (5.4) are concave functions of their arguments, similar to Figure 5.1. The relaxed optimization problem is obtained by restricting $r_{i[k]}$ and $f_{i[k]}$ to less than each of the terms in their defining min{} functions. The objective function is a linear combination of mainline and on-ramp flows termed the *Generalized Total Travel Time* (gTTT). We will see that the solution to the relaxed problem can be guaranteed to fall on the "upper boundary" whenever certain conditions on the *cost weights* are met. The goal of the *Cost Weights Synthesis* (CWS) problem, described in Section 5.2, is to find weights that meet these conditions, given the layout of the freeway. Thus, any global solution to the relaxed linear problem posed with cost weights generated by the CWS problem is also a global solution to the nonlinear problem with the same objective.

Section 5.3 presents a simple example aimed at determining whether the optimal plan performs well in terms of Total Travel Time, in addition to gTTT. This depends on the form of the weights generated by the CWS, and their similarity to the TTT weights (Eqs. (5.9) and (5.10)). It is found that the CWS weights generated with the original approach are qualitatively different from TTT in two ways: first, the mainline flow weights decay too quickly (faster than a straight line), and second, the weights on the on-ramp flows are not identically zero. Two additional assumptions are introduced in Section 5.4 to remedy these problems. Simulation experiments in Section 5.5 demonstrate that,

1. the cost weights generated by the CWS problem under the two additional assumptions are qualitatively similar to TTT,

2. the relaxed linear problem can be posed with or without restrictions on the lengths of the on-ramp queues,

3. the optimal plans derived in both cases (with or without queue length constraints) offer substantial savings in TTT,

4. both assumptions are reasonable.

## 5.1 Problem Formulation

### 5.1.1 Objective function

Equation (2.42) in Section 2.2.1 provided an expression for the Total Travel Time as the sum of a control independent term and a term composed of exiting flows. Below we derive this expression using the notation of the ACTM (Section 3.1). An expression for the total number of vehicles in the system ($N_{[k]}$) is first derived from the conservation equations (5.1) and (5.2) (multiplied by $\delta_i$):

$$\rho_{i[k+1]} = \rho_{i[k]} + f_{i-1[k]} + \delta_i\, r_{i[k]} - f_{i[k]}/\bar{\beta}_{i[k]}$$
$$\delta_i\, l_{i[k+1]} = \delta_i\, l_{i[k]} + \delta_i\, d_{i[k]} - \delta_i\, r_{i[k]}$$

Adding the two:

$$(\rho_{i[k+1]} + \delta_i\, l_{i[k+1]}) = (\rho_{i[k]} + \delta_i\, l_{i[k]}) + f_{i-1[k]} - f_{i[k]}/\bar{\beta}_{i[k]} + \delta_i\, d_{i[k]}$$
$$= (\rho_{i[k]} + \delta_i\, l_{i[k]}) + f_{i-1[k]} - f_{i[k]} - \frac{\beta_{i[k]}}{\bar{\beta}_{i[k]}} f_{i[k]} + \delta_i\, d_{i[k]}$$

Summing both sides over all sections $i$ and cancelling $f_{i[k]}$ terms, an expression for the evolution of the total system occupancy $N_{[k]} \triangleq \sum_i (\rho_{i[k]} + \delta_i\, l_{i[k]})$ is obtained:

$$N_{[k+1]} = N_{[k]} + f_{-1[k]} - f_{I-1[k]} - \sum_{i\in\mathcal{I}} \frac{\beta_{i[k]}}{\bar{\beta}_{i[k]}} f_{i[k]} + \sum_{i\in\mathcal{E}n} d_{i[k]}$$

The total occupancy at time $k$ is then given by:

$$N_{[k]} = N_{[0]} + \sum_{m=0}^{k-1} \left[ f_{-1[m]} - f_{I-1[m]} - \sum_{i\in\mathcal{I}} \frac{\beta_{i[m]}}{\bar{\beta}_{i[m]}} f_{i[m]} + \sum_{i\in\mathcal{E}n} d_{i[m]} \right] \qquad (5.5)$$

63

Equation (5.5) can now be used in the definition of the Total Travel Time (TTT):

$$
\text{TTT} \triangleq \sum_{k \in \mathcal{K}} \left[ \sum_{i \in \mathcal{I}} \rho_{i[k]} \Delta t \; + \; \sum_{i \in \mathcal{E}n} l_{i[k]} \Delta t \right] \tag{5.6}
$$

$$
= \Delta t \sum_{k \in \mathcal{K}} N_{[k]}
$$

$$
= \Delta t \sum_{k \in \mathcal{K}} \left\{ N_{[0]} \; + \; \sum_{m=0}^{k-1} \left[ f_{-1[m]} \; - \; f_{I-1[m]} \; - \; \sum_{i \in \mathcal{I}} \frac{\beta_{i[m]}}{\bar{\beta}_{i[m]}} f_{i[m]} \; + \; \sum_{i \in \mathcal{E}n} d_{i[m]} \right] \right\}
$$

$$
= \Delta t \, \text{K} \, N_{[0]} \; + \; \Delta t \sum_{k=0}^{\text{K}-1} \sum_{m=0}^{k-1} \left( f_{-1[m]} \; + \; \sum_{i \in \mathcal{E}n} d_{i[m]} \right)
$$

$$
- \; \Delta t \sum_{k=0}^{\text{K}-1} \sum_{m=0}^{k-1} \left( f_{I-1[m]} \; + \; \sum_{i \in \mathcal{I}} \frac{\beta_{i[m]}}{\bar{\beta}_{i[m]}} f_{i[m]} \right)
$$

Expanding the double summations and grouping similar terms leads to:

$$
\text{TTT} = \underbrace{\Delta t \, \text{K} \, N_{[0]} + \Delta t \sum_{k \in \mathcal{K}} (\text{K} - k) \left( f_{-1[k]} + \sum_{i \in \mathcal{E}n} d_{i[k]} \right)}_{\text{control independent}}
$$

$$
\underbrace{- \; \Delta t \sum_{k \in \mathcal{K}} (\text{K} - k) \left( f_{I-1[k]} + \sum_{i \in \mathcal{I}} \frac{\beta_{i[k]}}{\bar{\beta}_{i[k]}} f_{i[k]} \right)}_{\text{exit flows}}
$$

The "control independent" portion includes the initial condition and the inlet demands, which are not affected by the on-ramp flows whenever the conditions of Section 2.2.1 are met. The "exit flows" portion gathers offramp flows $s_{i[k]} = (\beta_{i[k]}/\bar{\beta}_{i[k]}) f_{i[k]}$, and the flow leaving the system through the downstream mainline boundary $f_{I-1[k]}$. Thus, minimizing the TTT is equivalent to minimizing:

$$
\overline{\text{TTT}} \triangleq - \sum_{k \in \mathcal{K}} (\text{K} - k) \left( f_{I-1[k]} + \sum_{i \in \mathcal{I}} \frac{\beta_{i[k]}}{\bar{\beta}_{i[k]}} f_{i[k]} \right) \tag{5.7}
$$

This is a particular instance of the *Generalized Total Travel Time*, defined below as a linear combination of mainline and on-ramp flows with positive *cost weights* $a_{i[k]}$ and $b_{i[k]}$:

$$
\text{gTTT} \triangleq - \sum_{k \in \mathcal{K}} \left[ \sum_{i \in \mathcal{I}} a_{i[k]} \, f_{i[k]} \; + \; \sum_{i \in \mathcal{E}n} b_{i[k]} \, r_{i[k]} \right] \tag{5.8}
$$

By using different settings for the cost weights, gTTT can be made to represent different useful cost functions. Equation (5.7) demonstrates that TTT is minimized by setting:

$$
a_{i[k]} = \begin{cases} (\text{K} - k) \, \beta_{i[k]}/\bar{\beta}_{i[k]} & i < I - 1 \\ (\text{K} - k) \left( \beta_{i[k]}/\bar{\beta}_{i[k]} + 1 \right) & i = I - 1 \end{cases} \tag{5.9}
$$

$$
b_{i[k]} = 0 \tag{5.10}
$$

The *Total Travel Distance* is maximized by setting $a_{i[k]}$ and $b_{i[k]}$ equal to their respective mainline and on-ramp section lengths. The cost weights that render `gTTT` equal to `TTT` can be described qualitatively as follows:

1. the $a_{i[k]}$'s are linearly decreasing from an initial positive value to 0 at $k = \mathrm{K} - 1$.

2. the $b_{i[k]}$'s are all zero.

The metering plan derived in this chapter is globally optimal with respect to an objective function that is *similar* to `TTT`, in the sense that they share these qualitative characteristics. The speculation, which is supported by the example of Section 5.5, is that such a plan also performs well in terms of `TTT`.

## 5.1.2   Bounds

Minimum and maximum bounds on the control and state are often used to keep their values within physical limits. Several of these bounds are not required here because they are implicit in the ACTM due to Theorem 1 on page 37. Specifically, $\rho_{i[k]} \in [0, \bar{\rho}_i]$, $l_{i[k]} \geq 0$, $f_{i[k]} \geq 0$, and $r_{i[k]} \geq 0$, are guaranteed. However, bounds are needed to account for the restrictions on the *metering rate* and the *on-ramp storage capacity*. These constraints were described in Section 2.2.2 as the two most important factors limiting the potential savings with on-ramp metering. They are represented in the problem formulation as follows:

| | | | |
|---|---|---|---|
| Metering rate bounds: | $r_i^c[k] \;\geq\; \underline{r}_i^c$ | $\forall\, i \in \mathcal{E}n^+$ | (5.11) |
| | $r_i^c[k] \;\leq\; \bar{r}_i^c$ | $\forall\, i \in \mathcal{E}n^+$ | (5.12) |
| Queue length bounds: | $l_{i[k]} \;\leq\; \bar{l}_i$ | $\forall\, i \in \mathcal{E}n$ | (5.13) |

## 5.1.3   Problem statement

The nonlinear and relaxed linear problems are respectively labelled as $\mathcal{P}_A$ and $\mathcal{P}_B$. In both cases, the on-ramp queue lengths are left unrestricted (i.e. Eq. (5.13) is not enforced). The reason for this omission is explained in Section 5.2.1. Queue length bounds can only be imposed in Problem $\mathcal{P}_C$ (page 75), after the two assumptions of Section 5.4 have been made. In the statement of Problem $\mathcal{P}_B$, the linear constraints (5.16) through (5.18) are a relaxation of (5.4) in $\mathcal{P}_A$, and (5.19) through (5.20) are a relaxation of (5.3).
**Problem $\mathcal{P}_A$**: Given initial and boundary conditions satisfying Theorem 1, find

$$\psi^* = \arg \min_{\psi \in \Omega_A} \mathtt{gTTT}(\psi) \tag{5.14}$$

$$\Omega_A = \Big\{ \; \psi = \{\rho_{i[k]}, l_{i[k]}, f_{i[k]}, r_{i[k]}, r_i^c[k]\} \; :$$

$$\text{Dynamic equations : Eqs. (5.1) and (5.2),}$$
$$\text{Concave fundamental diagram : Eqs. (5.3) and (5.4),}$$
$$\text{Metering rate bounds : Eqs. (5.11) and (5.12)} \Big\}$$

**Problem** $\mathcal{P}_B$: Given initial and boundary conditions satisfying Theorem 1, find

$$\psi^* = \arg \min_{\psi \in \Omega_B} \mathbf{gTTT}(\psi) \tag{5.15}$$

$$\Omega_B = \Big\{ \ \psi = \{\rho_{i[k]}, l_{i[k]}, f_{i[k]}, r_{i[k]}, r_i^c{}_{[k]}\} \ :$$
$$\text{Dynamic equations : Eqs. (5.1) and (5.2)},$$
$$\text{Linear inequality constraints : Eqs. (5.16) through (5.21)},$$
$$\text{Metering rate bounds : Eqs. (5.11) and (5.12)} \Big\}$$

$\forall \, k \in \mathcal{K}, i \in \mathcal{I}$

$$f_{i[k]} \ \leq \ \bar{\beta}_{i[k]} \, v_i \big( \, \rho_{i[k]} \ + \ \delta_i \, \gamma_i \, r_{i[k]} \, \big) \tag{5.16}$$
$$f_{i[k]} \ \leq \ w_{i+1} \big( \, \bar{\rho}_{i+1} - \rho_{i+1[k]} \, \big) \ - \ \delta_{i+1} \, \alpha_{i+1} \, r_{i+1[k]} \tag{5.17}$$
$$f_{i[k]} \ \leq \ \min \left\{ \bar{f}_i \ ; \ \frac{\bar{\beta}_{i[k]}}{\beta_{i[k]}} \bar{s}_i \right\} \tag{5.18}$$

$\forall \, k \in \mathcal{K} \, , \ i \in \mathcal{E}n :$

$$r_{i[k]} \ \leq \ l_{i[k]} \ + \ d_{i[k]} \tag{5.19}$$
$$r_{i[k]} \ \leq \ \xi_i \big( \, \bar{\rho}_i - \rho_{i[k]} \, \big) \tag{5.20}$$

$\forall \, k \in \mathcal{K} \, , \ i \in \mathcal{E}n^+ :$

$$r_{i[k]} \ \leq \ r_i^c{}_{[k]} \tag{5.21}$$

Our goal is to find a globally optimal solution to $\mathcal{P}_A$. This is not a simple task, in principle, since the problem is nonlinear and non-differentiable. Problem $\mathcal{P}_B$, on the other hand, is linear, and can therefore be solved globally and quickly. In the following section we will derive conditions on the cost weights $a_{i[k]}$ and $b_{i[k]}$ that make the two problems *equivalent*. Hence, under these conditions, the global and easily found solution to $\mathcal{P}_B$ also solves $\mathcal{P}_A$.

## 5.2 The Cost Weights Synthesis Problem

The goal of the Cost Weights Synthesis (CWS) problem is to find weights $a_{i[k]}$ and $b_{i[k]}$ that render problems $\mathcal{P}_A$ and $\mathcal{P}_B$ *equivalent* ($\mathcal{P}_A \equiv \mathcal{P}_B$), in the sense that their solution sets are identical:

$$\{\psi^* \text{ solves } \mathcal{P}_A\} \qquad \Leftrightarrow \qquad \{\psi^* \text{ solves } \mathcal{P}_B\} \tag{5.22}$$

Problem $\mathcal{P}_B$ is a relaxation of $\mathcal{P}_A$ since $\Omega_A \subset \Omega_B$ and the two share the same cost function. Therefore, any solution to $\mathcal{P}_B$ that is within $\Omega_A$ must also solve $\mathcal{P}_A$. For the two solution sets to be equivalent, the entire set of solutions to $\mathcal{P}_B$ must be contained in $\Omega_A$. Hence, $\mathcal{P}_A \equiv \mathcal{P}_B$ if:

$$\{\psi^* \text{ solves } \mathcal{P}_B\} \qquad \Rightarrow \qquad \psi^* \in \Omega_A \tag{5.23}$$

Conversely stated, the two problems are equivalent if there are no solutions to $\mathcal{P}_B$ in the set $\Omega_B \setminus \Omega_A$:

$$\psi \in \Omega_B \setminus \Omega_A \qquad \Rightarrow \qquad \{\psi^* \text{ does not solve } \mathcal{P}_B\} \tag{5.24}$$

A feasible point $\psi \in \Omega_B \setminus \Omega_A$ can be shown *not* to solve $\mathcal{P}_B$ if there exists a perturbation $\boldsymbol{\Delta}$ that is both *feasible* and *improving*:

$\boldsymbol{\Delta}$ is *feasible* if $\exists \, \epsilon > 0$ such that: $\qquad\qquad \psi + \epsilon \boldsymbol{\Delta} \in \Omega_B$ $\qquad\qquad\qquad$ (5.25)

$\boldsymbol{\Delta}$ is *improving* if $\exists \, \epsilon > 0$ such that: $\qquad\qquad \text{gTTT}(\psi + \epsilon \boldsymbol{\Delta}) < \text{gTTT}(\psi)$ $\qquad\qquad$ (5.26)

Due to the linearity of gTTT, Eq. (5.26) is equivalent to $\text{gTTT}(\boldsymbol{\Delta}) < 0$. The CWS problem is then restated as follows: Given the layout of the freeway, find weights $a_{i[k]}$ and $b_{i[k]}$ such that for all $\psi \in \Omega_B \setminus \Omega_A$ there exists a feasible perturbation $\boldsymbol{\Delta}$ with $\text{gTTT}(\boldsymbol{\Delta}) < 0$.

The methodology used here for solving the CWS problem is to propose a feasible perturbation for every $\psi \in \Omega_B \setminus \Omega_A$ (Section 5.2.2). Evaluating $\text{gTTT}(\boldsymbol{\Delta}) < 0$ on each of these perturbations leads to conditions on the values of $a_{i[k]}$ and $b_{i[k]}$. We then compute particular values of $a_{i[k]}$ and $b_{i[k]}$ that meet all of these conditions.

## 5.2.1 Feasible perturbations

A feasible perturbation $\boldsymbol{\Delta} = \{\Delta \rho_{i[k]}, \Delta l_{i[k]}, \Delta f_{i[k]}, \Delta r_{i[k]}, \Delta r_{i}^{c}{}_{[k]}\}$ to a point $\psi = \{\rho_{i[k]}, l_{i[k]}, f_{i[k]}, r_{i[k]}, r_{i}^{c}{}_{[k]}\}$ $\Omega_B$ is one for which $\psi + \epsilon \boldsymbol{\Delta}$ satisfies Eqs. (5.1), (5.2), (5.11), (5.12), and (5.16) through (5.21), for some $\epsilon > 0$. This can be expressed in terms of the components of $\boldsymbol{\Delta}$. For example, the mainline conservation equations for $\psi$ and $\psi + \epsilon \boldsymbol{\Delta}$ are:

$$
\begin{aligned}
\rho_{i[k+1]} &= \rho_{i[k]} + f_{i-1[k]} + \delta_i \, r_{i[k]} - f_{i[k]}/\bar{\beta}_{i[k]} \\
(\rho_{i[k+1]} + \epsilon \Delta \rho_{i[k+1]}) &= (\rho_{i[k]} + \epsilon \Delta \rho_{i[k]}) + (f_{i-1[k]} + \epsilon \Delta f_{i-1[k]}) + \delta_i (r_{i[k]} + \epsilon \Delta r_{i[k]}) \\
&\quad - (f_{i[k]} + \epsilon \Delta f_{i[k]})/\bar{\beta}_{i[k]}
\end{aligned}
$$

Subtracting the two and dividing by $\epsilon$,

$$
\Delta \rho_{i[k+1]} = \Delta \rho_{i[k]} + \Delta f_{i-1[k]} + \delta_i \, \Delta r_{i[k]} - \Delta f_{i[k]}/\bar{\beta}_{i[k]}
$$

The same must be done for the on-ramp conservation equation and for all *active* inequality constraints. Thus, a feasible perturbation to $\psi$ is one that satisfies the following:

$$
\begin{aligned}
\Delta \rho_{i[k+1]} &= \Delta \rho_{i[k]} + \Delta f_{i-1[k]} - \Delta f_{i[k]}/\bar{\beta}_{i[k]} + \delta_i \Delta r_{i[k]} \\
\Delta l_{i[k+1]} &= \Delta l_{i[k]} - \Delta r_{i[k]} \\
\Delta f_{i[k]} &\leq \bar{\beta}_{i[k]} \, v_i (\, \Delta \rho_{i[k]} + \delta_i \gamma_i \Delta r_{i[k]} \,) \qquad && \text{if } (5.16)^{ik} \text{ is active in } \psi \\
\Delta f_{i[k]} &\leq -w_{i+1} \Delta \rho_{i+1[k]} - \delta_{i+1} \alpha_{i+1} \Delta r_{i+1[k]} \qquad && \text{if } (5.17)^{ik} \text{ is active in } \psi \\
\Delta f_{i[k]} &\leq 0 && \text{if } (5.18)^{ik} \text{ is active in } \psi \\
\Delta r_{i[k]} &\leq \Delta l_{i[k]} && \text{if } (5.19)^{ik} \text{ is active in } \psi \\
\Delta r_{i[k]} &\leq -\xi_i \, \Delta \rho_{i[k]} && \text{if } (5.20)^{ik} \text{ is active in } \psi \\
\Delta r_{i[k]} &\leq \Delta r_{i}^{c}{}_{[k]} && \text{if } (5.21)^{ik} \text{ is active in } \psi \\
\Delta r_{i}^{c}{}_{[k]} &\geq 0 && \text{if } (5.11)^{ik} \text{ is active in } \psi \\
\Delta r_{i}^{c}{}_{[k]} &\leq 0 && \text{if } (5.12)^{ik} \text{ is active in } \psi
\end{aligned}
$$

The notation $(5.16)^{ik}$ refers to the particular instance of constraint $(5.16)$ applied to the $i$th section, during the $k$th time interval. An inequality constraint is active in a feasible point $\psi$ when it evaluates as an equality.

These conditions for a feasible perturbation are used next to explain the reason why queue length constraints were omitted from the statements of problems $\mathcal{P}_A$ and $\mathcal{P}_B$. The queue length constraint would have appeared in the definition of the feasible perturbation as the addition of the following condition:

$$\Delta l_{i[k]} \leq 0 \qquad\qquad\qquad \text{if } (5.13)^{ik} \text{ is active in } \psi$$

Consider a $\psi \in \Omega_B \setminus \Omega_A$ for which a particular component $f_{i[k]}$ is beneath its "upper boundary". That is, Eqs. $(5.16)^{ik}$, $(5.17)^{ik}$, and $(5.18)^{ik}$ are not active in $\psi$ for some $i$ and $k$. We would like to find a feasible perturbation to $\psi$ that has a positive component $\Delta f_{i[k]} > 0$, in order to move $f_{i[k]}$ closer to its "upper boundary". Considering only causal perturbations, i.e. ones whose components are zero for all times before $k$, then the following derives from the conditions on feasible perturbations:

$$
\begin{aligned}
& \Delta\rho_{i+1[k+1]} \;=\; \Delta f_{i[k]} \;>\; 0 \\
\Rightarrow\quad & \Delta r_{i+1[k+1]} \;\leq\; -\xi_i \Delta\rho_{i+1[k+1]} \;<\; 0 \qquad \text{if } (5.20)^{i+1,k+1} \text{ is active in } \psi \\
\Rightarrow\quad & \Delta l_{i+1[k+2]} \;=\; -\Delta r_{i+1[k+1]} \;>\; 0
\end{aligned}
$$

This last line conflicts with the condition imposed by the queue constraint if $(5.13)^{i+1,k+2}$ is active in $\psi$. Therefore, in the case that both $(5.20)^{i+1,k+1}$ and $(5.13)^{i+1,k+2}$ are active, there is no feasible causal perturbation with $\Delta f_{i[k]} > 0$. One might consider searching amongst non-causal perturbations, however the argument would still hold for $k = 0$. One could also consider perturbations with $f_{i[k]} \leq 0$, however such perturbations would not meet the second criterion of being cost improving (Eq. $(5.26)$). One may further consider "sideways" perturbations, where $f_{i[k]} = 0$ and either $\Delta\rho_{i[k]}$ in Eq. $(5.16)$ or $\Delta\rho_{i+1[k]}$ Eq. $(5.17)$ is perturbed. This however, cannot be done when $k = 0$ since the values of $\rho_{i[0]}$ and $\rho_{i+1[0]}$ are fixed. Thus, it seems that whenever a queue constraint is included, there may be points $\psi \in \Omega_B \setminus \Omega_A$ for which no feasible and improving perturbation exists. In fact, in this case, it may even be that $\mathcal{P}_A$ is not feasible, and therefore $\mathcal{P}_A$ and $\mathcal{P}_B$ are not equivalent for any selection of $a_{i[k]}$ and $b_{i[k]}$.

## 5.2.2 The Maximal Worst-Case Causal perturbation

It is obviously not possible to perturb every one of the innumerable $\psi \in \Omega_B \setminus \Omega_A$. Instead, $\Omega_B \setminus \Omega_A$ is divided into a finite number of subsets, and perturbations are constructed that are feasible for all points in their respective subsets. The subsets are defined as follows:

$$\mathrm{I}_{\iota\kappa} = \big\{\, \psi \in \Omega_B \setminus \Omega_A \;:\; \text{Eq. } (5.4) \text{ is not satisfied with } i = \iota\,,\ k = \kappa \,\big\} \qquad (5.27)$$

$$\mathrm{II}_{\iota\kappa} = \big\{\, \psi \in \Omega_B \setminus \Omega_A \;:\; \text{Eq. } (5.3) \text{ is not satisfied with } i = \iota\,,\ k = \kappa \,\big\} \qquad (5.28)$$

A member of $\mathrm{I}_{\iota\kappa}$ is a point $\psi \in \Omega_B \setminus \Omega_A$ for which Eq. $(5.4)$ is violated with $i = \iota$ and $k = \kappa$, or equivalently one for which constraints $(5.16)^{\iota\kappa}$, $(5.17)^{\iota\kappa}$, and $(5.18)^{\iota\kappa}$ are inactive

($f_{\iota[\kappa]}$ is beneath its upper boundary in Figure 5.1). Subsets of the $\text{II}_{\iota\kappa}$ type have analogous interpretations for the on-ramp flows. There are a total of $\text{I} \times \text{K}$ subsets of the $\text{I}_{\iota\kappa}$ type and $|\mathcal{E}n| \times \text{K}$ subsets of the $\text{II}_{\iota\kappa}$ type.

To each of the $(\text{I} + |\mathcal{E}n|) \times \text{K}$ subsets corresponds a *Maximal Worst-Case Causal* (MWCC) perturbation; $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\text{I}}$ for $\text{I}_{\iota\kappa}$ and $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\text{II}}$ for $\text{II}_{\iota\kappa}$. The definition of the MWCC perturbations is given in Eqs. (5.29) and (5.30). The MWCC is a feasible perturbation for all points in its corresponding subset because it is feasible for the *worst-case* point, where all inequality constraints, aside from the ones that define it, are active. It is *maximal* because it selects the largest (least negative) feasible value for each of the $\Delta f$'s and $\Delta r$'s. This is done in order to maximize its beneficial effect on the cost, and thus minimize the rate of decay of the resulting cost weights. Finally, it is *causal* because its effects only propagates forward in time. That is, all of the components ($\Delta \rho_{i[k]}$, $\Delta f_{i[k]}$, etc.) of $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\text{I}}$ and $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\text{II}}$ equal 0 for $k < \kappa$. This property enables the simple backstepping numerical algorithm described in Section 5.2.3.

$$\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\text{I}} = \{\Delta\rho_{i[k]}, \Delta l_{i[k]}, \Delta f_{i[k]}, \Delta r_{i[k]}, \Delta r_{i[k]}^c\} \text{ is defined for } \kappa \in \mathcal{K}, \; \iota \in \mathcal{I} \text{ as:} \tag{5.29}$$

$$\Delta\rho_{i[k+1]} = \Delta\rho_{i[k]} + \Delta f_{i-1[k]} - \Delta f_{i[k]}/\bar{\beta}_{i[k]} + \delta_i \Delta r_{i[k]}$$

$$\Delta l_{i[k+1]} = \Delta l_{i[k]} - \Delta r_{i[k]}$$

$$\Delta f_{i[k]} = \begin{cases} 1 & i = \iota, \; k = \kappa \\ \min\Big\{ \bar{\beta}_{i[k]} v_i \left(\Delta\rho_{i[k]} + \delta_i \gamma_i \Delta r_{i[k]}\right) \; ; \\ \qquad -w_{i+1}\Delta\rho_{i+1[k]} - \delta_{i+1}\alpha_{i+1}\Delta r_{i+1[k]} \; ; \; 0 \Big\} & i \in \mathcal{I}, \; k > \kappa \\ 0 & i \in \mathcal{I}, \; k < \kappa \end{cases}$$

$$\Delta r_{i[k]} = \min\{\Delta l_{i[k]} \; ; \; -\xi_i \Delta\rho_{i[k]} \; ; \; 0\}$$

$$\Delta r_{i[k]}^c = 0$$

$$\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\text{II}} = \{\Delta\rho_{i[k]}, \Delta l_{i[k]}, \Delta f_{i[k]}, \Delta r_{i[k]}, \Delta r_{i[k]}^c\} \text{ is defined for } \kappa \in \mathcal{K}, \; \iota \in \mathcal{E}n \text{ as:} \tag{5.30}$$

$$\Delta\rho_{i[k+1]} = \Delta\rho_{i[k]} + \Delta f_{i-1[k]} - \Delta f_{i[k]}/\bar{\beta}_{i[k]} + \delta_i \Delta r_{i[k]}$$

$$\Delta l_{i[k+1]} = \Delta l_{i[k]} - \Delta r_{i[k]}$$

$$\Delta f_{i[k]} = \min\Big\{ \bar{\beta}_{i[k]} v_i \left(\Delta\rho_{i[k]} + \delta_i \gamma_i \Delta r_{i[k]}\right) \; ; \; -w_{i+1}\Delta\rho_{i+1[k]} - \delta_{i+1}\alpha_{i+1}\Delta r_{i+1[k]} \; ; \; 0 \Big\}$$

$$\Delta r_{i[k]} = \begin{cases} 1 & i = \iota, \; k = \kappa \\ \min\{\Delta l_{i[k]} \; ; \; -\xi_i \Delta\rho_{i[k]} \; ; \; 0\} & i \in \mathcal{E}n, \; k > \kappa \\ 0 & i \in \mathcal{E}n, \; k < \kappa \end{cases}$$

$$\Delta r_{i[k]}^c = 0$$

Each of the $(\text{I} + |\mathcal{E}n|) \times \text{K}$ MWCC perturbations can be computed offline given the layout of the freeway, the model parameters, and the offramp split ratios. Figure 5.2 shows a sample
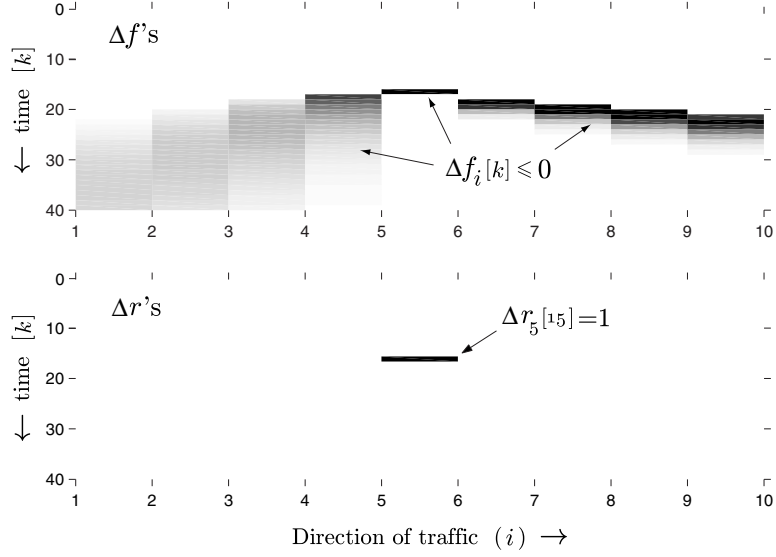
Figure 5.2: $\bar{\boldsymbol{\Delta}}_{5,15}^{\mathrm{II}}$ with parameters of Section 5.3.

MWCC perturbation ($\bar{\boldsymbol{\Delta}}_{5,15}^{\mathrm{II}}$) with the parameters of Section 5.3. Here, a unit increase in $r_5[15]$ produces negative waves that propagate forward in time, and upstream and downstream along the freeway.

With all of the MWCC perturbations computed, the CWS problem can be restated as: Find $a_i[k]$ and $b_i[k]$ such that:

$$\mathrm{gTTT}(\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{I}}) \ < \ 0 \qquad\qquad \forall\,\kappa\in\mathcal{K}\ ,\ \iota\in\mathcal{I} \qquad\qquad (5.31)$$

$$\mathrm{gTTT}(\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{II}}) \ < \ 0 \qquad\qquad \forall\,\kappa\in\mathcal{K}\ ,\ \iota\in\mathcal{E}n \qquad\qquad (5.32)$$

Because each MWCC perturbation is feasible for every point in its member set, conditions (5.31) and (5.32) are sufficient to guarantee non-optimality for all $\psi \in \Omega_B\setminus\Omega_A$, and thus $\mathcal{P}_A\equiv\mathcal{P}_B$.

## 5.2.3  Backstepping numerical method

Using Eq. (5.8), Eqs. (5.31) and (5.32) are expressed as $(\mathrm{I}+|\mathcal{E}n|)\times\mathrm{K}$ linear equations of the form:

$$-\sum_{k\in\mathcal{K}}\left[\ \sum_{i\in\mathcal{I}}a_i[k]\,\Delta f_i[k]\ +\ \sum_{i\in\mathcal{E}n}b_i[k]\,\Delta r_i[k]\ \right] \ = \ -\varepsilon \qquad\qquad (5.33)$$

where $\Delta f$'s and $\Delta r$'s are components of either $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{I}}$ or $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{II}}$, and $\varepsilon$ is some positive number. Because the MWCC perturbation is *causal*, the summations in Eq. (5.33) only contain non-zero elements for $k \geq \kappa$. Equation (5.33) is therefore equivalent to:

$$-\sum_{k=\kappa}^{\mathrm{K}-1}\left[\ \sum_{i\in\mathcal{I}}a_i[k]\,\Delta f_i[k]\ +\ \sum_{i\in\mathcal{E}n}b_i[k]\,\Delta r_i[k]\ \right] \ = \ -\varepsilon \qquad\qquad (5.34)$$

70

Noting that at time $k = \kappa$ all of the components of $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{I}}$ and $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{II}}$ are zero, except for:

$$\begin{aligned}
\Delta f_{\iota[\kappa]} &= 1 && \text{in } \bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{I}} \\
\Delta f_{\iota-1[\kappa]} &= -\alpha_\iota \quad \text{and} \quad \Delta r_{\iota[\kappa]} = 1 && \text{in } \bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{II}}
\end{aligned}$$

the cost weights for section $\iota$ and time $\kappa$ can be written in terms of present and future components of $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{I}}$ and $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{II}}$ as follows:

$$a_{\iota[\kappa]} = -\sum_{k=\kappa+1}^{\mathrm{K}-1} \left( \sum_{i \in \mathcal{I}} a_{i[k]} \, \Delta f_{i[k]} + \sum_{i \in \mathcal{E}n} b_{i[k]} \, \Delta r_{i[k]} \right) + \varepsilon \qquad (5.35)$$

$$b_{\iota[\kappa]} = -\sum_{k=\kappa+1}^{\mathrm{K}-1} \left( \sum_{i \in \mathcal{I}} a_{i[k]} \, \Delta f_{i[k]} + \sum_{i \in \mathcal{E}n} b_{i[k]} \, \Delta r_{i[k]} \right) + \alpha_\iota a_{\iota[\kappa]} + \varepsilon \qquad (5.36)$$

where $\Delta f$'s and $\Delta r$'s in (5.35) are components of $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{I}}$, and $\Delta f$'s and $\Delta r$'s in (5.36) are components of $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{II}}$. These equations can be easily solved by setting $a_{i[\mathrm{K}-1]}$ and $b_{i[\mathrm{K}-1]}$ to some positive value, and computing the rest sequentially backwards.

## 5.3   A simple example

The CWS problem was solved for a simple test freeway consisting of 10 sections and 40 time intervals. A single on-ramp was placed at $i = 5$. Offramps were placed at $i = 4$, 5, and 9 with $\beta_{i[k]} = 0.1$ for all three. $v_i = 0.7$ and $w_i = 0.2$ was used throughout.

Figure 5.3 shows the cost weights resulting from $\alpha_5 = \gamma_5 = 0.2$, and $\xi_5 = 0.06$. A *time decay index* ($D$) was computed for each sequence $a_i[\cdot]$ and $b_i[\cdot]$ as the number of entries in the sequence that exceeded 10% of the first value, divided by the length of the sequence; for example, $D(a_4[\cdot]) = \text{size}\{a_4[\cdot] \geq 0.1 \times a_4[0]\}/40$. The decay index for a constant sequence is 1.0, and the decay index for a linearly decreasing sequence, such as the TTT weights of Eq. (5.9), is 0.9. We inspect the decay indices because they provide some measure of the similarity of gTTT with TTT. Values near 0.9 among the decay indices for $a_i[\cdot]$'s would suggest that the resulting objective function is "close" to total travel time. The decay indices for the sequences of Figure 5.3 are $D(a_4[\cdot]) = 0.51$ and $D(b_5[\cdot]) = 0.54$.

Figure 5.4 shows $\min_i \{D(a_i[\cdot])\}$ (i.e. the decay index of the fastest decaying mainline weight) in the top window, and $\min_i \{D(b_i[\cdot])\}$ in the bottom window, as functions of $\xi_5$, and for several values of $\alpha_5$. The decay index was found to be insensitive to $\gamma_5$. The figure shows that the cost weights degrade less quickly, and are therefore more similar to TTT, for smaller values of $\alpha_i$ and $\xi_i$. However it also shows that the gTTT objective is quite different from TTT in most cases.

## 5.4   Modifications to the original formulation

The preceding example reveals a few deficiencies in the solution to the CWS problem described in Section 5.2. First, the weights on the mainline flows, $a_{i[k]}$'s, decay more quickly than the TTT weights of Eq. (5.9), which decay linearly. Hence, the resulting optimal plan
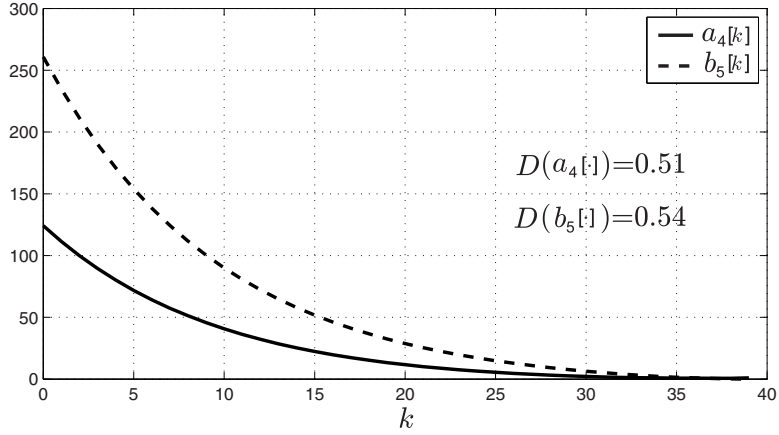
Figure 5.3: Cost weights with $\alpha_5 = \gamma_5 = 0.2$ and $\xi_5 = 0.06$.
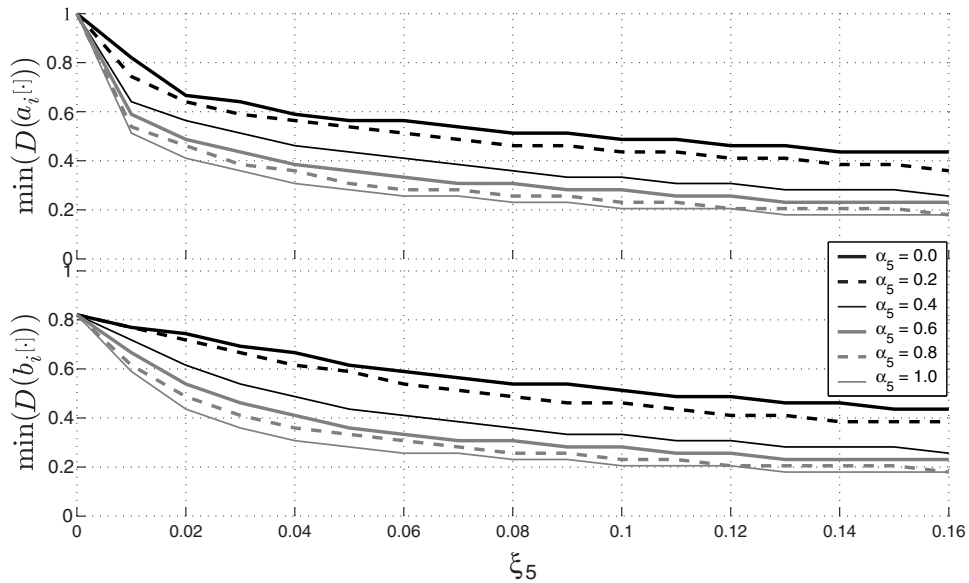


Figure 5.4: Fastest decay index vs. $\xi_5$ and $\alpha_5$. $(\gamma_5 = 0.2)$.

will be more likely to allow congestion to form late in the optimization time window. Also, unlike the TTT $b_i[k]$ weights of Eq. (5.10), the $b_i[k]$'s of the example are not identically zero. A third deficiency, this one in the formulation of the problem, is the omission of upper limits on the on-ramp queue lengths (Eq. (5.13)). Section 5.2.1 explained that including these constraints would have introduced the possibility of a feasible solution in $\Omega_B \setminus \Omega_A$ with no improving perturbation. Finally, the time horizon used in the example spanned only 40 time intervals – about 7 minutes. The linear program resulting from a more useful 1-hour prediction window was found to be unmanageably large.

This section outlines a series of alterations to the original formulation that resolve these four shortcomings. The modified optimization problem, Problem $\mathcal{P}_C$, minimizes an objective function with linearly decaying $a_i[k]$'s and identically zero $b_i[k]$'s, includes on-ramp queue constraints, and is small enough to be solved for up to a 5-hour prediction window.

72

## 5.4.1 An additional restriction: Slowly varying on-ramp flows

Problem $\mathcal{P}_B$ was formulated using the original ACTM of Section 3.1. The number of equations and constraints in the problem can be substantially reduced by using a modified dual-scale ACTM instead. In contrast to the dual-scale model defined in Section 3.3 (dACTM-c), the modified version (dACTM-o) updates *all* of the on-ramp related quantities at the slower rate, not just the control.

The units of the various quantities change as expected: $r_i[k]$, $l_i[k]$, $d_i[k]$, $r_i^c[k]$ are in numbers of vehicles per $\Delta t_c$, and the mainline variables, $\rho_i[k]$ and $f_i[k]$, are in vehicles per $\Delta t$. The mainline model of the dACTM-o is obtained by replacing the $r_i[k]$'s in Eqs. (3.14) and (3.8) with $r_i[\kappa]/p$. The on-ramp conservation equation remains unchanged. The dual-scale version of the on-ramp flow equation, Eq. (5.40), is the only component of the model that changes significantly. Here, instead of a single "$\xi_i$" term as in Eq. (3.12), $p$ separate "$\xi_i$" terms are needed to ensure that the on-ramp flow does not overwhelm the mainline over the entire *upcoming* $\Delta t_c$ interval. This is necessary in order to retain the property of Theorem 1. The modified model equations are given below.

**Dual-scale ACTM (dACTM-o)**:

$$\rho_i[k+1] \;=\; \rho_i[k] \;+\; f_{-1}[k] \;+\; \delta_i r_i[\kappa]/p \;-\; f_i[k]/\bar{\beta}_i[k] \qquad \forall\, i \in \mathcal{I} \;,\; k \in \mathcal{K}^p \tag{5.37}$$

$$f_i[k] = \min\bigg\{ \; \bar{\beta}_i[k] v_i(\rho_i[k] + \delta_i \gamma_i r_i[\kappa]/p) \;;\; \bar{f}_i \;;$$

$$w_{i+1}(\bar{\rho}_{i+1} - \rho_{i+1}[k]) - \delta_{i+1}\alpha_{i+1} r_{i+1}[\kappa]/p \;;\; \frac{\bar{\beta}_i[k]}{\beta_i[k]}\bar{s}_i \; \bigg\} \qquad \forall\, i \in \mathcal{I} \;,\; k \in \mathcal{K}^p \tag{5.38}$$

$$l_i[\kappa+1] \;=\; l_i[\kappa] \;+\; d_i[\kappa] \;-\; r_i[\kappa] \qquad \forall\, i \in \mathcal{E}n \;,\; \kappa \in \mathcal{K} \tag{5.39}$$

$$r_i[\kappa] = \begin{cases} \min\big\{ l_i[\kappa] + d_i[\kappa] \;;\; \min_k\{\xi_i(\bar{\rho}_i - \rho_i[k])\} \;;\; r_i^c[\kappa] \big\} & \forall\, i \in \mathcal{E}n^+ \;,\; \kappa \in \mathcal{K} \\[6pt] \min\big\{ l_i[\kappa] + d_i[\kappa] \;;\; \min_k\{\xi_i(\bar{\rho}_i - \rho_i[k])\} \big\} & \forall\, i \in \mathcal{E}n \setminus \mathcal{E}n^+ \;,\; \kappa \in \mathcal{K} \end{cases} \tag{5.40}$$

In Eqs. (5.37) and (5.38), $\kappa$ is the integer part of $k/p$. $\mathcal{K}^p$ and $\mathcal{K}$ are the sets of mainline and on-ramp time intervals: $\mathcal{K}^p = [0, \ldots, p\mathrm{K} - 1]$ and $\mathcal{K} = [0, \ldots, \mathrm{K} - 1]$. The internal $\min\{\}$ in Eq. (5.40) is taken over $k = \kappa p \ldots (\kappa + 1)p - 1$. That is, $r_i[\kappa]$ is restricted by $p$ terms in the upcoming time interval. Unfortunately, this makes the dACTM-o model *non-causal*, since knowledge of future mainline densities is needed to compute the present on-ramp flow. Unlike the basic ACTM and the dACTM-c, uniqueness of the dACTM-o is therefore not a given. Also, the numerical method for solving the model is not straightforward. Fortunately however, the problem of non-causality is eliminated with the assumption proposed in the next section.

## 5.4.2 Assumption #1: Unobstructed on-ramps

Two of the four deficiencies noted at the beginning of this section – fast decaying $a_i[k]$'s and the lack of on-ramp queue limits – are caused by the interaction between the mainline densities and the on-ramp flows dictated by Eq. (5.20). The non-causality of the dACTM-o model is also due to the "$\xi_i$" terms in Eq. (5.40). All three of these problems disappear with the following assumption.

*Equation (5.20) is not active in the optimal solution.*

In other words, it is assumed that congestion on the mainline does not spill onto any on-ramps, metered or not, when optimal metering is used. Obviously, this does not hold for every freeway. In some cases, on-ramp flows may be restricted by insufficient space on the freeway, even when the freeway is optimally controlled. One scenario in which this may occur is on a heavy uncontrolled freeway-to-freeway connector, such as the I-605/I-210 merge described in Chapter 6. The assumption therefore limits the applicability of the proposed design to freeways where it holds – presumably most well designed freeways. If the assumption does hold, it permits the removal of Eq. (5.20) from the problem statement, and the "$-\xi_i \Delta\rho_{i[k]}$" terms from the definitions of $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{I}}$ and $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{II}}$ (Eqs. (5.29) and (5.30)). Notice that then $\Delta r_{i[k]} = \Delta l_{i[k]} = 0$ in $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{I}}$, and $\Delta l_{i[k]} \leq 0$ in $\bar{\boldsymbol{\Delta}}_{\iota\kappa}^{\mathrm{II}}$. There is no longer a conflict between the MWCC perturbation and the queue length constraint, which requires only $\Delta l_{i[k]} \leq 0$ (Section 5.2.1). Furthermore, removing the "$\xi_i$" terms from Eq. (5.40) leads to the following *causal* definition of on-ramp flows:

$$
r_{i[\kappa]} =
\begin{cases}
\min\left\{ l_{i[\kappa]} + d_{i[\kappa]} \; ; \; r_{i[\kappa]}^c \right\} & \forall\, i \in \mathcal{E}n^+ \,,\; \kappa \in \mathcal{K} \\
l_{i[\kappa]} + d_{i[\kappa]} & \forall\, i \in \mathcal{E}n \setminus \mathcal{E}n^+ \,,\; \kappa \in \mathcal{K}
\end{cases}
\tag{5.41}
$$

As we shall see in Section 5.5, this assumption also eliminates the problem of rapidly decaying $a_{i[k]}$'s. Section 5.5 also shows that the assumption is not overly restrictive, and in fact holds for the I-210 model.

## 5.4.3   Assumption #2: Zero minimum metering rate

With the additional constraint of slowly varying on-ramp flows and the added assumption of unobstructed on-ramps, the variable $r_{i[\kappa]}^c$ can be seen to be practically unnecessary. It only appears in Eq. (5.41) and in the minimum and maximum metering rate bounds, Eqs. (5.11) and (5.12). Its only function is to restrict the on-ramp flows at metered locations. The relaxed linear problem can therefore be equivalently stated without the $r_{i[\kappa]}^c$ variables by setting $r_{i[\kappa]} = l_{i[\kappa]} + d_{i[\kappa]}$ at uncontrolled on-ramps, and $r_{i[\kappa]} \leq \min\{ l_{i[\kappa]} + d_{i[\kappa]} \; ; \; \bar{r}_i^c \}$ at metered on-ramps. This leads to the statement of Problem $\mathcal{P}_C$ in Eq. (5.42). The values of $r_{i[\kappa]}^c$ for metered on-ramps can be constructed after the optimal solution to Problem $\mathcal{P}_C$ is found as the optimal values of $r_{i[\kappa]}$ that are strictly less than $\min\{ l_{i[\kappa]} + d_{i[\kappa]} \; ; \; \bar{r}_i^c \}$. Where the optimal $r_{i[\kappa]}$'s equal $\min\{ l_{i[\kappa]} + d_{i[\kappa]} \; ; \; \bar{r}_i^c \}$, the on-ramp metering rate can be assigned any larger value, up to $\bar{r}_i^c$.

The only pitfall in this simplification is that the optimal on-ramp flows may be less than the minimum metering rate $\underline{r}_i^c$ (typically 180 vph). Theorem 1 guarantees however that they will be positive. Hence, the $r_{i[k]}^c$ variables can be removed without affecting the problem whenever $\underline{r}_i^c = 0$.

$\underline{r}_i^c = 0$.

This is almost never true in practice: the actual minimum metering rate on I-210 is 180 vph, which corresponds to 1 vehicle every 20 seconds. The assumption is adopted nevertheless because it reduces the number of variables and constraints, but more importantly because it leads to $b_{i[k]} = 0$. Recall that this is one of the qualitative features of TTT-like

objective functions. Under Assumption #2, in addition to Assumption #1, and using the dACTM-o model, the following linear program is equivalent to Problem $\mathcal{P}_B$:

**Problem** $\mathcal{P}_C$: Given initial and boundary conditions satisfying Theorem 1, find

$$\psi^* = \arg \min_{\psi \in \Omega_C} \mathbf{gTTT}(\psi) \tag{5.42}$$

$$\Omega_C = \Big\{ \ \psi = \{\rho_i[k], l_i[k], f_i[k], r_i[k]\} \ :$$

Dynamic equations : Eqs. (3.18) and (3.20),

Relaxed fundamental diagram : Eqs. (5.43) through (5.45)

On-ramp flow constraints :  Eqs. (5.46) through (5.48)

On-ramp queue constraint :  Eq. (5.13) $\Big\}$

$\forall \, k \in \mathcal{K}^p \,, \ i \in \mathcal{I} :$
$$f_{i[k]} \ \leq \ \bar{\beta}_{i[k]} \, v_i ( \ \rho_{i[k]} \ + \ \delta_i \, \gamma_i \, r_{i[\kappa]}/p \ ) \tag{5.43}$$
$$f_{i[k]} \ \leq \ w_{i+1} \, ( \ \bar{\rho}_{i+1} - \rho_{i+1[k]} \ ) \ - \ \delta_{i+1} \, \alpha_{i+1} \, r_{i+1[\kappa]}/p \tag{5.44}$$
$$f_{i[k]} \ \leq \ \min \left\{ \bar{f}_i \ ; \ \frac{\bar{\beta}_{i[k]}}{\beta_{i[k]}} \bar{s}_i \ \right\} \tag{5.45}$$

$\forall \, \kappa \in \mathcal{K} \,, \ i \in \mathcal{E}n \backslash \mathcal{E}n^+ :$ $\qquad\qquad r_{i[\kappa]} \ = \ d_{i[\kappa]}$ $\qquad\qquad\qquad$ (5.46)

$\forall \, \kappa \in \mathcal{K} \,, \ i \in \mathcal{E}n^+ :$ $\qquad\qquad r_{i[\kappa]} \ \leq \ l_{i[\kappa]} + d_{i[\kappa]}$ $\qquad\qquad\qquad$ (5.47)

$\forall \, \kappa \in \mathcal{K} \,, \ i \in \mathcal{E}n^+ :$ $\qquad\qquad 0 \ \leq \ r_{i[\kappa]} \ \leq \ \bar{r}_i^c$ $\qquad\qquad\qquad$ (5.48)

Notice that the statement of Problem $\mathcal{P}_C$ does not include $l_{i[\kappa]}$'s for unmetered on-ramps, since these are not allowed to accumulate by Assumption #1. A simple step not shown here is to eliminate constraint (5.46) and the unmetered $r_{i[\kappa]}$ variables by replacing $r_{i[k]}$ with the given demands $d_{i[\kappa]}$.

However, the main advantage of $\mathcal{P}_C$ over $\mathcal{P}_B$ is that neither metered nor unmetered $r_{i[k]}$'s in $\mathcal{P}_C$ are required to fall on their "upper boundary". Perturbations to $r_{i[k]}$'s and the resulting non-zero $b_{i[k]}$'s are therefore not needed. The CWS problem can be solved for Problem $\mathcal{P}_C$ by setting all $b_{i[k]}$'s to zero, and computing $a_{i[k]}$'s with the procedure developed in Section 5.2, with $\bar{\Delta}_{\iota\kappa}^I$'s modified to include the dACTM-o model and Assumption #1.

The optimal plan derived from Problem $\mathcal{P}_C$ may call for some on-ramps to be metered at rates below $\underline{r}_i^c$, or even temporarily closed. This may not be feasible in practice. An *implementable* metering plan, complying with both the maximum and minimum metering rate bounds, can be derived from the solution to Problem $\mathcal{P}_C$ with:

$$r_{i[k]}^c \ = \ \max\{ \ r_{i[k]} \ ; \ \underline{r}_i^c \ \} \tag{5.49}$$

The optimal, perhaps unimplementable metering plan is of course $r_{i[k]}^c = r_{i[k]}$. The following section provides a numerical example using the I-210 layout that demonstrates that the sacrifice to global optimality incurred by applying the *implementable* plan instead of the optimal plan is not large.

## 5.5 Tests using the I-210 layout

The I-210 test site is described in Chapter 6. The details of its representation in the ACTM are explained in Section 3.4. Problem $\mathcal{P}_C$ was solved using this layout for 1, 2 and 5-hour time horizons, with $\Delta t = 10$ seconds and $\Delta t_c = 5$ minutes (i.e. $p = 30$). In all cases, an additional half-hour "cool-down" period was appended to the end of the simulation period. The optimizations were performed over the entire 1.5, 2.5, and 5.5 hour time windows. During the cool-down period, all traffic demands were set to zero, and the freeway was allowed to empty completely.
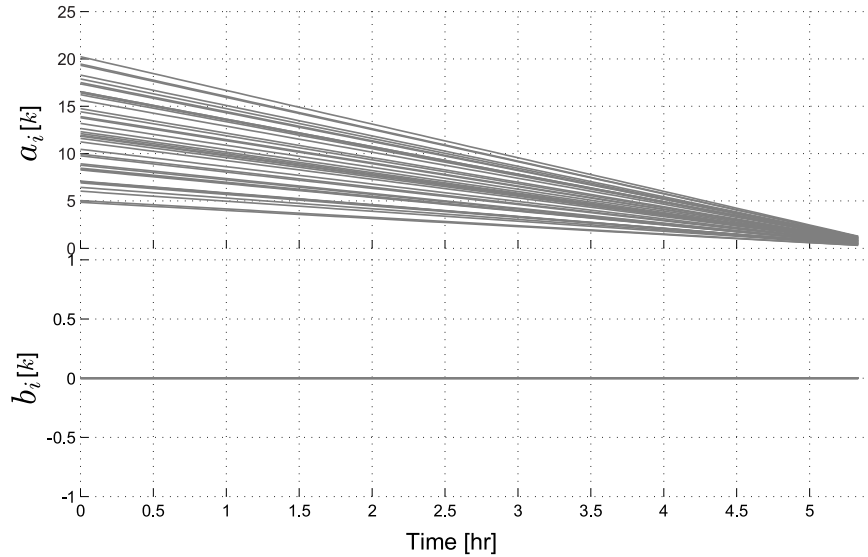


Figure 5.5: Computed cost weights.

The cost weights generated by the CWS problem, under Assumptions #1 and #2, are shown in Figure 5.5. As opposed to the result Section 5.3, these cost weights have the desired TTT-like properties of linear decay and $b_i[k] = 0$. The optimal solution to Problem $\mathcal{P}_C$ posed with these weights is also a global solution to $\mathcal{P}_A$ (with the same objective).

The linear programs were solved with the commercial LP solver MOSEK 3.0. Each of the three time horizons was solved with and without on-ramp queue length restrictions (Eq. (5.13)). The size of the problem ranged from 92,310 constraints and 41,480 variables for the 1-hour problem without queue constraints, to 352,950 constraints and 158,600 variables for the 5-hour problem with queue constraints. Percent improvements in TTT are reported in Table 5.1. The travel time was computed using Eq. (5.6), and including the cool-down period. It was confirmed in every case that the solution to the LP problem satisfied the equations of the model to a high degree of precision – i.e. $\psi^* \in \Omega_A$.

The validity of the two assumptions was also confirmed. For Assumption #1, it was verified that the optimal $r_i[k]$'s never exceeded $\xi_i\, p\, (\bar{\rho}_i - \rho_i[x])$. Assumption #2 was found to have little effect on the solution. This was concluded by running the implementable metering plan suggested in Eq. (5.49) through the model. TTT values for both the optimal and implementable plans are provided in Table 5.1. These results show that increasing the minimum

metering rate from 0 (optimal) to 180 (implementable) induces only a small reduction in travel time savings (0.04% without queue constraints and 1.12% with queue constraints). It should be noted that applying Eq. (5.49) can never cause the queue constraint to be violated, since increasing the metering rate will only make the on-ramp queues shorter.

| Period [hr:min] | TTT no control | TTT optimal | % saved | TTT implem. | % saved |
|---|---|---|---|---|---|
| Without queue constraints | | | | | |
| 1:00 | 1,716 | 1,715 | 0.06% | 1,716 | 0.00% |
| 2:00 | 4,080 | 4,035 | 1.10% | 4,036 | 1.08% |
| 5:00 | 13,075 | 11,535 | 11.78% | 11,540 | 11.74% |
| With queue constraints | | | | | |
| 1:00 | 1,716 | 1,715 | 0.06% | 1,716 | 0.00% |
| 2:00 | 4,080 | 4,053 | 0.66% | 4,060 | 0.49% |
| 5:00 | 13,075 | 11,824 | 9.56% | 11,971 | 8.44% |

Table 5.1: Travel time savings and run times.

It is also interesting to note that the 1-hour and 2-hour time horizons yielded almost no improvement over no control. This is because, as can be seen in Figure 3.7, congestion only begins after the first hour, and starts to dissipate in the fourth hour. These two observations tend to corroborate Zhang's conclusion in [73] that "uniformly uncongested" and "uniformly congested" freeways should not be metered since they provide no opportunity for reducing congestion. The 1-hour experiment demonstrates that nothing can be gained by metering the uncongested freeway, unless there is knowledge of future congestion. In the 2-hour scenario there is no post-peak period. Hence, vehicles retained in the on-ramps cannot be released without increasing congestion. Not much can be gained by metering this situation either. In the 5-hour experiment however, the optimizer is able to shift the surplus demand to the post-peak period by holding vehicles on the on-ramps. Only the 5-hour experiments produced a substantial improvement over no control: 11.78% travel time savings without queue constraints, and 9.56% with queue constraints. This result also emphasizes the importance of using a numerical technique that is efficient enough to produce optimal plans for sufficiently long time horizons, within a relatively short computation time.

Optimized speed contour plots and queue lengths for the 5-hour experiments are shown in Figure 5.6. The left hand plot shows that the optimal strategy when on-ramp queue lengths are left unrestricted is to keep the freeway almost completely uncongested by storing large numbers of vehicles in the on-ramps. In this situation, one of the on-ramp queues accumulates over 500 vehicles. The ride hand side shows that congestion cannot be avoided when the on-ramp queues are limited to at most 50 vehicles. The implementable metering plan nevertheless achieves a reduction of 8.44% in TTT in this case.

77
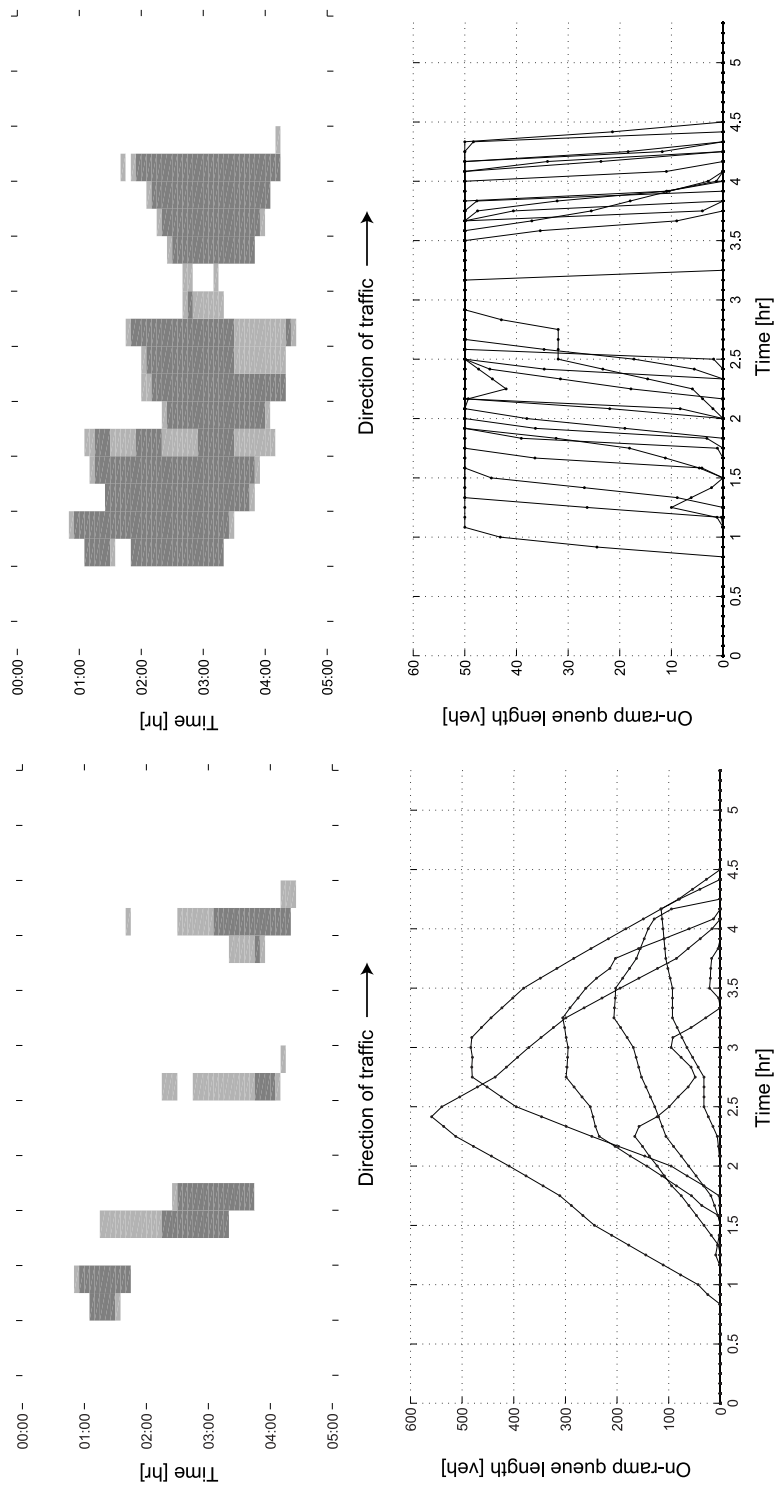
Figure 5.6: Congestion and queue lengths with and without queue constraints.

## 5.6   Summary

This chapter has outlined a complete methodology for solving the feedforward optimal metering problem efficiently and with near-global optimality. The unadulterated result was described in the Sections 5.2 and 5.2.3, and was found to have a few drawbacks. Section 5.4 focussed on solving those problems. The approach was tested in Section 5.5 on realistic freeway setup. The repairs required the use of two additional assumptions. The first, that congestion does not propagate from the mainline onto the on-ramps, can be easily verified in the optimal solution. Future work will consider the question of what to do if this assumption is temporarily violated. The second assumption was that the on-ramp flows could be reduced to zero by the on-ramp meters. This is never true, but it was found to induce only a small sacrifice of global optimality. It was found that, under these two assumptions, the cost weights generated by the CWS problem were qualitatively similar to the values that minimize total travel time, in that 1) the weights on the mainline flows decayed linearly in time, and 2) the on-ramp flow weights were all zero. Also, these assumptions enabled the inclusion of queue length constraints, and allowed the formulation of a simplified but equivalent problem (Problem $\mathcal{P}_C$).

This technique has several advantages over many other predictive on-ramp metering designs. First, it requires only to solve a single linear program, which can be done with extreme efficiency using any modern LP solver. Second, it takes on-ramp storage constraints explicitly into account. Finally, the optimal solution is a global optimum, or near global if slowly varying on-ramp flows are imposed, with respect to a cost function that is qualitatively similar to Total Travel Time.

# Chapter 6

# The I-210 test site

The remainder of this dissertation consists of a microscopic evaluation of several local on-ramp metering schemes. One of the objectives of the study is to determine the influence of certain more detailed features of the freeway and its control system, that could not be included in the macroscopic studies. For example, the specifics of the SATMS controller chip used on I-210 and the controller override rules. Also the differences in behavior between trucks, passenger cars with single occupants, and high-occupancy vehicles (HOVs). The latter are allowed on I-210 to bypass the meters on the on-ramps, and to use a dedicated HOV lane on the mainline. All of these features can be reproduced in the VISSIM microscopic simulation environment.

Interstate 210, also known as the Foothill Freeway, is one of the important east-west links in the Los Angeles County freeway system. It runs along the northern edge of the L.A. metropolitan area, connecting several surrounding cities, including Pasadena, Monrovia, Azusa, and Glendora. The present study focuses on a 15-mile stretch of the westbound direction, between Vernon St. and the 210/134 junction (see Figure 6.1), which sustains heavy congestion daily between about 6:30 am and 9:30 am. This site contains several complicating features that make it a challenge for microscopic simulation. These include:

1. Twenty metered on-ramps, all equipped with upstream mainline loop detectors, a set of on-ramp loop detectors, and 1 or 2 meter heads.

2. A centralized control system that gathers the freeway measurements and issues the control commands remotely from a Traffic Management Center (TMC).

3. A dedicated lane for HOVs, separated from the mixed-traffic lanes by an *intermittent barrier*.

4. Bypass lanes on the on-ramps that allow the HOVs to circumvent the on-ramp meters.

5. A heavy and uncontrolled freeway connector from I-605.

6. Several interacting bottlenecks.

This chapter provides a description of the geometric and traffic related characteristics of the test site. It also describes the procedure that was followed to collect and analyze the traffic

data and to compute the OD tables for the microscopic model. Section 6.8 provides an interpretation of the observed congestion patterns, which was used in the model calibration phase of Chapter 7.

## 6.1 Sources of geometric information

The basic geometric characteristic of the freeway include the location of on-ramps and offramps, the number of lanes and location of lane drops, and the freeway curvature. In the case of I-210, additional features were also considered important, such as the location of the HOV *gates* (gaps in the intermittent barrier where vehicles enter and exit the HOV lane), the positions of the traffic sensors, and the control hardware. Support material used to determine the layout of I-210 included *as-built* maps and a record of aerial photographs provided by Caltrans, and a report from a previous study on I-210 provided by the District 07 Traffic Operations group [72]. This information was encoded in VISSIM with much help from the VISSIM staff.

In VISSIM, the recommended method for entering the geometric data is to construct a scaled map, in bitmap format. This picture can be displayed as a background image in the program, allowing the user to easily trace the links and connectors that constitute the freeway layout (see Figure 6.2). The topographical features that were considered relevant to the description of I-210 are:

1. For the mainline:

    (a) Width and numbers of lanes

    (b) Locations of on-ramps and offramps

    (c) Lane drops

    (d) Auxiliary lanes

    (e) Lane change zones

    (f) Location of the HOV lane and gates

    (g) Position of mainline loop detector stations

2. For on-ramps and offramps

    (a) Number of lanes at the gore of each on-ramp and offramp

    (b) Existence of an HOV bypass lane on on-ramps

    (c) Existence and position of metering lights on on-ramps

    (d) Arrangement of loop detectors on on-ramps and offramps. The position of the on-ramp queue detector with respect to the presence detector is especially important for on-ramp control evaluations, since it determines the maximum storage of the on-ramp.

Three sources of geometric information were used:
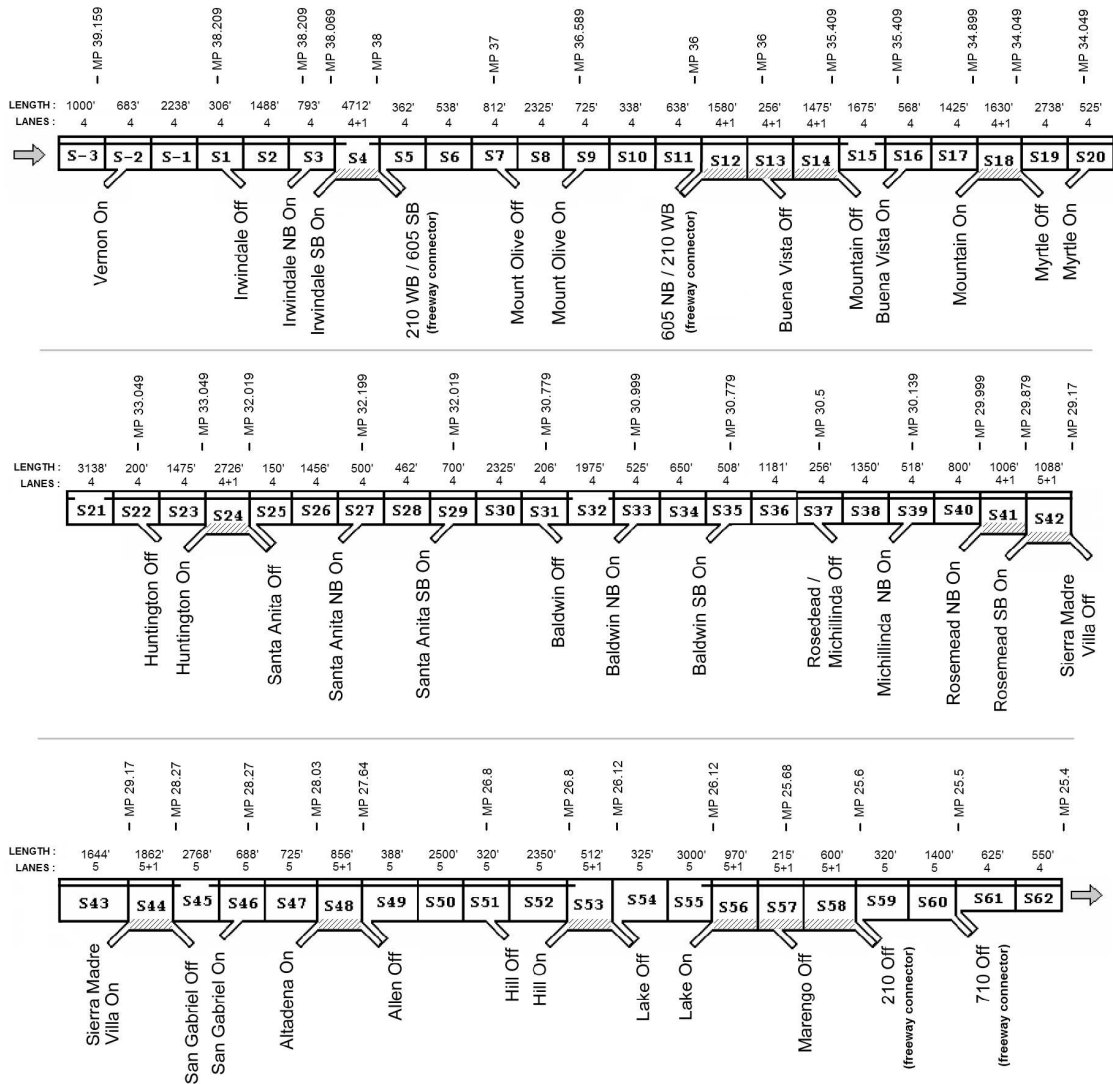
Figure 6.1: 65 sections in the test site. (MP = Mile Post).

1. A set of photocopies of scaled aerial photographs obtained from Caltrans HQ. These photographs are black-and-white and printed on $11'' \times 17''$ paper, with a 1:2400 scale.

2. A set of "as-built" maps indicating the arrangement of loop detectors on on-ramps and the mainline. These were provided by the Caltrans District 07 Ramp Metering Group, headed by Mr. Hanh Pham.

3. Un-scaled aerial photographs in bitmap format downloaded from MapQuest[1].

All of the geometric features were extracted from the aerial photographs (source 1), with the exception of items 1g, 2c, and 2d, which were measured from the as-built maps (source

---

[1]http://www.mapquest.com

2). Each of the important features was assigned a section in Figure 6.1. In total, the site was divided into 65 sections. Boundaries were chosen to isolate each of the important topographical features. For example, section S29 contains a single lane change zone (item 1e) where traffic from the Santa Anita St. on-ramp merges with the mainline stream. Figure 6.1 also provides the lengths (in feet) and the number of mixed-flow lanes in each section. This highway partition was transferred to the large overhead view compiled from source 3 (Figure 6.2).
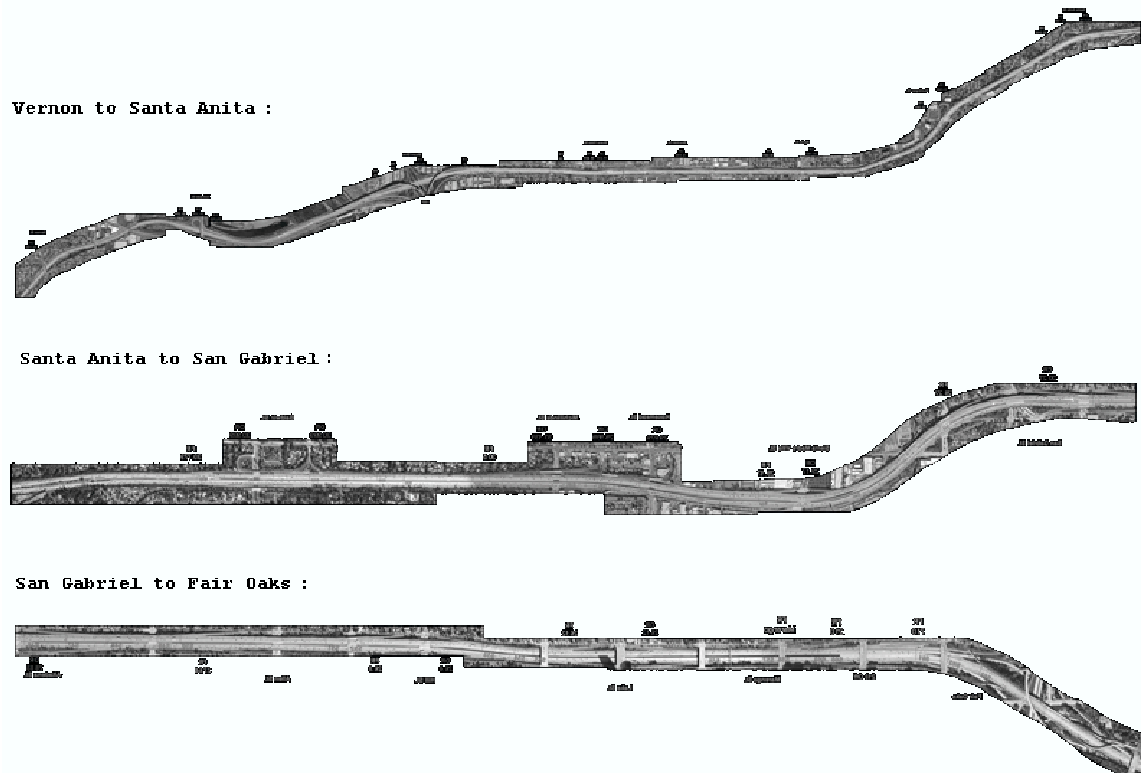


Figure 6.2: Assembled overhead view of I-210.

## 6.2 Traffic data sources

The traffic demand is defined in VISSIM as a set of OD matrices that specify the average numbers of vehicles travelling from every freeway origin to every destination, at 15-minute intervals[2]. This and the next few sections describe the procedure that was followed to gather and process traffic data for generating the OD matrices. The first step was to compile a complete and representative set of boundary flows, covering every on-ramp, offramp, and the two mainline boundaries. FREQ was then used to translate the boundary flows into the required set of OD matrices. Two sources of field data were used:

---

[2]This is one of two available methods. The alternative is to use *turning percentages* (i.e. split ratios).

1. **PeMS**: The PeMS database gathers 30-second and 5-minute data from over 30,000 miles of freeway in California. This database was used to assemble a history of traffic measurements for every loop detector station in the site. A Matlab-based data processing algorithm was created to filter, aggregate, and correct the PeMS data (Section 6.3). Three examples of speed contour maps generated from the processed PeMS mainline data can be found in Appendix B.1. These represent a light, a typical, and a heavy day of congestion on I-210. Speed contour plots such as these were used to identify the three major bottlenecks (Section 6.8), and played a significant role in the calibration effort.

2. **Manual counts**: The District 07 Traffic Operations group provided the results of a biennial survey of freeway ramp volumes conducted between 10/2001 and 1/2002. The collected data consists of 15-minute estimates of volumes on most of the on-ramps and offramps in the test site (all except Marengo St. and the I-210 and SR-710 freeway connectors). The D07 survey did not include any mainline data.
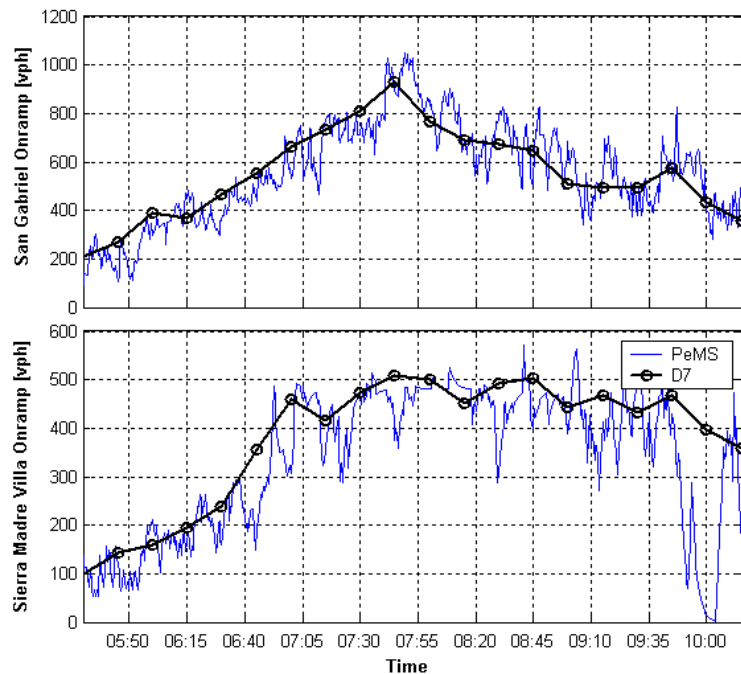


Figure 6.3: Comparison of PeMS 30-second data with the Caltrans D07 survey.

A sample of flow values for the Sierra Madre Villa on-ramp (MP 29.17) from each of the two data sources is shown in Figure 6.3. As in this example, there is close agreement between the two sources in most cases. Instances where significant differences were noted were usually attributable to malfunctioning loop detectors (i.e. errors in PeMS). Manual counts were generally favored over the PeMS loop detector measurements for the ramps. PeMS data was used primarily where mainline measurements were needed. That is, to determine the upstream and downstream mainline flows (needed to estimate the OD matrices in Section 6.7) and to construct the contour plots used for model calibration (Chapter 7).

## 6.3  PeMS data processing with Matlab

PeMS – the Performance Measurement System – has as its primary function to gather, analyze, and disseminate real-time traffic information for California freeways. Its main interface is a web page[3], where users can generate informative graphs and performance indices. Additionally, the raw traffic data is stored in a database. PeMS data was used primarily to identify recurring trends in demands and congestion patterns during the morning commute on I-210. Interpretation of PeMS-derived flow and speed contour maps yielded the critical traffic parameters (e.g. capacity, bottleneck locations) that were used to calibrate the VISSIM model. PeMS data was also used to asses the quality and reliability of real-time data on I-210. Section 6.4 summarizes the conclusions that were reached in this respect. This section gives a brief overview of the filtering and aggregation algorithms that were applied to the raw PeMS data sources prior to their being used in VISSIM.

The PeMS database stores two levels of data resolution: 30 seconds and 5 minutes. The 5-minute data is generated from the 30-second feed, and is aggregated over time before storage. The traffic variables recorded in PeMS include occupancy, flow, speed, and g-factor (estimated effective vehicle length). These can be combined to obtain an estimate of average density. All variables in the PeMS database are per-loop quantities. Samples of data sequences from PeMS are shown in Figure 6.4.
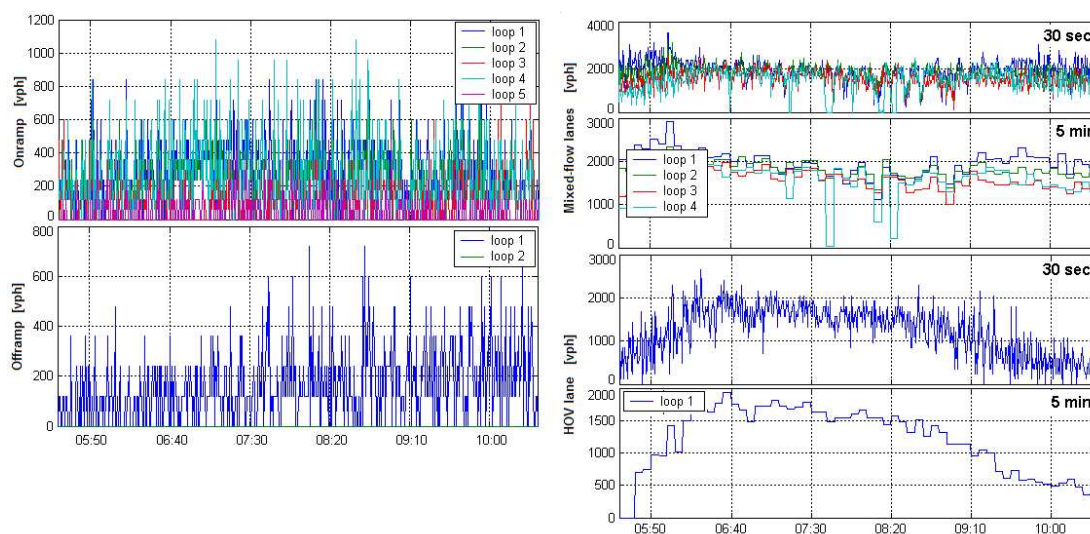


Figure 6.4: Raw 30-second and 5-minute flows from PeMS (`R30` and `R5`).

Figure 6.5 illustrates the stages of data processing that were applied to the raw PeMS feeds. All of these were implemented using Matlab. First, the raw 30-second data (`R30`) was put through a first-order low-pass filter, producing output `S30`. The smoothed and raw per-loop values (`S30`, `R30` and `R5`) were then aggregated over lanes, to obtain values for cross-sections of the freeway at ramps and mainline locations (`AS30`, `AR30` and `AR5`). In each

---

[3]http://pems.eecs.Berkeley.edu
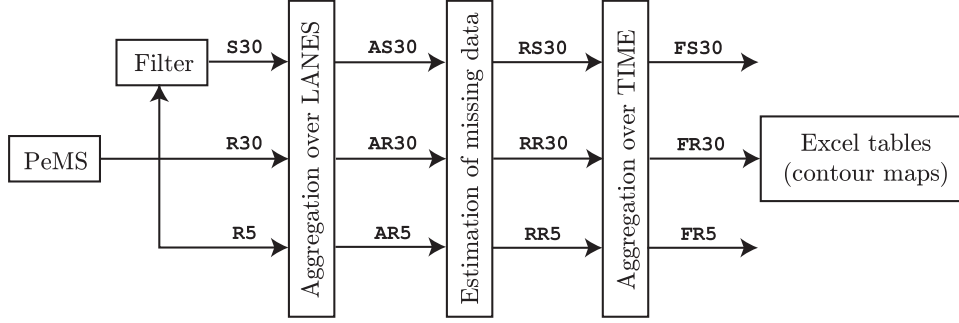
Figure 6.5: PeMS data processing.

case, the aggregation step was performed with:

$$occ^{agg}(i,k) = \sum_j \gamma(i,j) \; occ(i,j,k)$$

$$flow^{agg}(i,k) = \sum_j \gamma(i,j) \; flow(i,j,k) \qquad (6.1)$$

$$speed^{agg}(i,k) = \sum_j \gamma(i,j) \; speed(i,j,k)$$

Here, the values on the left-hand-side are aggregated quantities. They are a linear combination of the per-loop values, with coefficients $\gamma(i,j)$. $i$ denotes the detector station, $j$ is an index for each loop detector within a detector station, and $k$ is the time interval. For on-ramps, the detector station may include *entrance*, *presence*, *passage*, *hov bypass*, and *queue* detectors. In this case, all $\gamma(i,j)$'s were set to zero, except for the one representing the *entrance* loop, which was set to 1.0. For mainline loop detector stations, all $\gamma(i,j)$'s were set to 1.0. The $\gamma(i,j)$ coefficients were also used to perform crude data reconstruction for malfunctioning mainline loops. For instance, the detector on lane 2 of the Myrtle St. mainline station (MP 34.049) did not work on 11/6/2001. Its data was replaced with the average of lanes 1 and 3, by setting the $\gamma$'s on those lanes equal to 1.5.

Next, additional conservation-based data reconstruction methods were applied in cases where more severe data losses could not be compensated with the $\gamma(i,j)$ coefficients. Two examples of this situation that were encountered are the temporary loss of communication with an entire mainline station, and the permanent lack of loop detectors on several offramps. The reconstruction method is based on a static balance of flows on a small section of the freeway. Three reconstructed data sets resulted from this step: RS30, RR30, and RR5. These were fed through a time-aggregation block which generated 15-minute tables. The contour plots of Appendix B.1 were constructed from FR30.

## 6.4 Loop detector reliability

One of the difficulties of using detector data for model input and calibration (as well as for traffic-responsive control) is that in many cases adequate data is not available, due either to

86

an incomplete sensor infrastructure or to failure of the existing system. Large quantities of historical data from the PeMS database were used to asses the dependability of the existing loop detector infrastructure on I-210.

Table 6.1 provides percentages of time during which each on-ramp, offramp, and mainline station registered signal pulses. The percentage values in the table are the averages over all loops in a given station (on-ramp, offramp, or mainline/HOV station) of the ratio of non-zero flow measurements to the total number of measurements. These are optimistic estimates since they do not consider whether the non-zero values were reasonable. The statistics were taken over 11 weeks of PeMS data, using weekdays only, and from 5:30 am to 10:30 am. It can be noted that, in general, mainline and on-ramp detectors are more reliable than offramp detectors. Most remain on-line around 80% of the time. Two exceptions are the Michillinda NB (44%) and Sierra Madre Villa (60%) on-ramp and mainline stations. The only on-ramp lacking a set of loop detectors is the freeway connector from I-605 NB (MP 36). Offramps, on the other hand, are more problematic. Many lack sensors, or at least these are not included in the PeMS database (e.g. Buena Vista - MP 36). Others have sensors that appear to be permanently disconnected from the data collection system (e.g. Lake - MP 26.12).

## 6.5   Ramp flows from the Caltrans D07 survey

The on- and offramp counts collected by the District 07 biennial survey are provided in Appendix B.3. These measurements were gathered manually, by counting the number of vehicles that used every on-ramp and offramp, at 15-minute intervals, throughout the day. Each ramp was surveyed over a period of about 14 consecutive days. The surveyed days are highlighted in the tables of Appendix B.2. This data set constitutes a complete picture of the traffic demand entering and exiting the test site using the ramps, but it does not include any mainline data. Conversely, the PeMS database provides mainline measurements that are practically complete, but doesn't contain information from several key ramps, including the heavy freeway connector from I-605 NB (MP 36), and several offramps where loop detectors had either failed or were missing.

The main difficulty encountered with the D07 boundary data was that there was no single day in which all ramps were surveyed simultaneously. This situation is common in real-world settings, since it is rare to find a complete and reliable sensor structure. As a consequence, it was necessary to assemble a single composite day using ramp counts from several different days considered as *typical*. The set of typical days was created by first discarding all Mondays, Fridays, weekends, and days that did not closely follow the normal (i.e. average) pattern. The remaining days are highlighted with bold grey lines in the time series plots of Figures 9.4 through 9.6. The variances in the counts for the reduced group are plotted in Figure 6.6. These values were computed with:

$$\text{var}(s) \;=\; \frac{1}{K \times D} \sum_{k=1}^{K} \sum_{d=1}^{D} \frac{(f_{s,k,d} - \bar{f}_{s,k})^2}{\bar{f}_{s,k}}$$

$$\text{with} \qquad \bar{f}_{s,k} \;=\; \frac{1}{D} \sum_{d=1}^{D} f_{s,k,d}$$

| Street Name | MP | % non-zero data | | |
| --- | --- | --- | --- | --- |
| | | offramps | on-ramps | mainline/HOV |
| Vernon St On | 39.159 | - | 81.0% | 81.0% |
| Irwindale St. On/Off | 38.209 | 0.0% | 81.6% | 81.6% |
| | 38.069 | - | 81.2% | 81.2% |
| I-605 SB Off | 38 | n.m. | - | - |
| Mt. Olive Off | 37 | n.m. | - | - |
| Mt. Olive On | 36.589 | - | 81.4% | 81.4% |
| I-605 NB On | 36 | - | n.m. | n.m. |
| Buena Vista Off | 36 | n.m. | - | - |
| Mountain Off | 35.409 | 0.0% | - | - |
| Buena Vista On | 35.409 | - | 72.1% | 72.1% |
| Mountain On | 34.899 | - | 65.9% | 65.9% |
| Myrtle On/Off | 34.049 | 79.1% | 79.1% | 79.1% |
| Huntington On/Off | 33.049 | 79.9% | 80.4% | 80.4% |
| Santa Anita Off | 32.019 | 76.5% | - | - |
| Santa Anita NB On | 32.199 | - | 79.4% | 79.4% |
| Santa Anita SB On | 32.019 | - | 80.4% | 80.4% |
| Baldwin Off | 30.779 | 79.1% | - | - |
| Baldwin NB On | 30.999 | - | 80.5% | 80.5% |
| Baldwin SB On | 30.779 | - | 79.1% | 79.1% |
| Rosemead/Michillinda Off | 30.5 | n.m. | - | - |
| Michillinda NB On | 30.139 | - | 44.0% | 44.0% |
| Rosemead NB On | 29.999 | - | 79.9% | 79.9% |
| Rosemead SB On | 29.879 | - | 62.7% | 62.7% |
| Sierra Madre Villa On/Off | 29.17 | 60.2% | 60.2% | 60.2% |
| San Gabriel On/Off | 28.27 | 67.2% | 81.5% | 81.5% |
| Altadena On | 28.03 | - | 81.5% | 81.5% |
| Allen Off | 27.64 | 74.0% | - | - |
| Hill On/Off | 26.8 | 80.0% | 80.0% | 80.0% |
| Lake On/Off | 26.12 | 0.0% | 81.5% | 81.5% |
| Marengo Off | 25.68 | 81.5% | - | 81.5% |
| I-210 connector Off | 25.6 | n.m. | - | - |
| SR-710 connector Off | 25.5 | n.m | - | - |

Table 6.1: Data quality (n.m.=not measured, '-'=does not apply).

$f_{s,k,d}$ is the $k$th 15-minute vehicle count in the 5:30 am to 10:30 am period ($K = 20$), on day $d$, from station $s$. $D$ is the number of days in the reduced set.

From this set, a single day was selected for each on-ramp and offramp. The selected day is marked with a '+' sign in Appendix B.2, and also with '+' markers in Figures 9.4 through 9.6 in Appendix B.3.
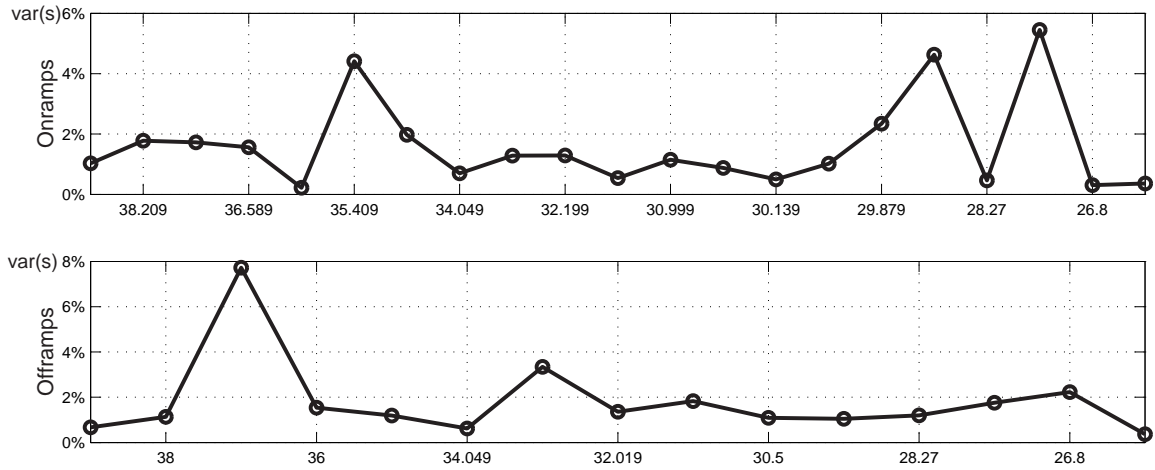
Figure 6.6: Percent variance in selected ramp counts from the Caltrans D07 survey.

## 6.6 Mainline flows from PeMS

Measurements for the two mainline boundaries (Vernon and Fair Oaks) were required to complete the specification of traffic demands for FREQ's OD table estimation. These were obtained from PeMS. Figure 6.7 shows per-lane average flow measurements (AS30) for several days (Tuesdays, Wednesdays, and Thursdays only) during the District 7 survey of freeway ramps. Notice that the flow pattern near Vernon St. does not resemble the expected inverted U shape for the morning period. Flow at this location starts at an extremely high value, near 2200 vph per lane, and slowly decreases throughout the morning. This tendency is odd, but repeats itself from day to day. Again, it was necessary to select a single typical day for the mainline boundary flows from a number of days. This selection was based on the following criteria:

1. completeness of the data set,

2. how well the flow data followed the day-to-day trend,

3. resulting *scale factor*.

Scale factors are defined as the ratio, for each 15-minute period, of the total number of vehicles entering the system to the total number of vehicles that exit. They are computed in FREQ as a first step to finding the OD matrices (Section 6.7). They can also be used to identify possible problems in the data set, since they are expected to fall within 10% of 1.00, for a normal (incident-less) traffic scenario, and their average over a 5-hour period should be very close to 1.00. The scale factors resulting from the final selection of ramp and mainline flows are shown in Figure 6.8. The aggregate scale factor for the 5 hour period is 1.02. Two tables with the final selection of ramp flows can be found in Appendix B.4.
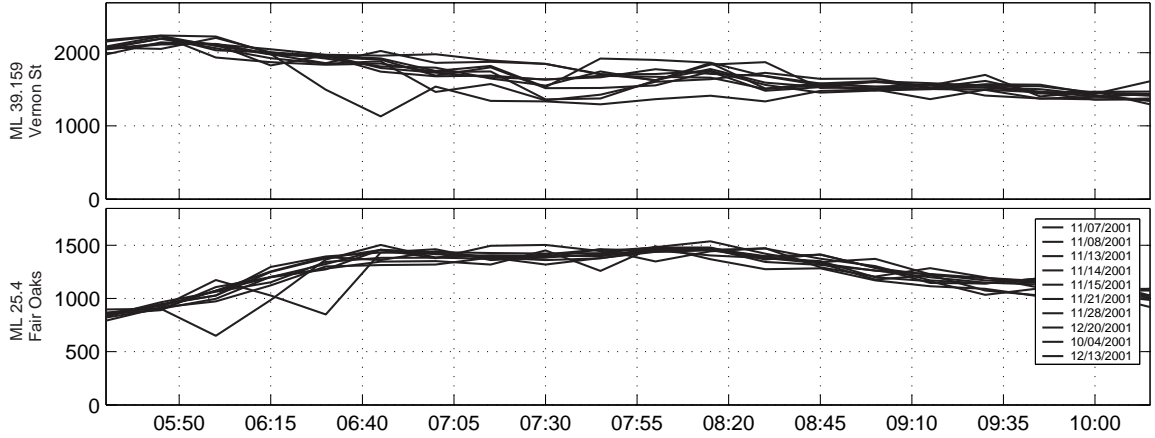
Figure 6.7: Mainline boundary flows for several days.
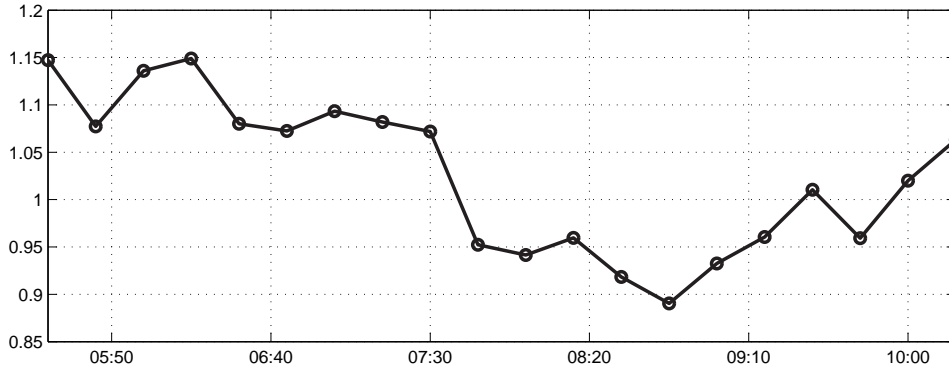


Figure 6.8: Time-dependent scale factors with selected boundary flows (average=1.02).

## 6.7 Estimating OD matrices with the FREQ model

The translation of ramp counts to the set of OD matrices required by VISSIM was achieved with FREQ. FREQ is a macroscopic deterministic freeway corridor model for the development and evaluation of freeway operational strategies, developed by A. D. May and his colleagues at U.C. Berkeley. As was mentioned in the previous section, FREQ was first used as a data verification tool. Specifically, it was used to check scale factors ($\alpha[k]$):

$$\alpha[k] = \frac{\sum_{\text{origins}} f_i^{in}[k]}{\sum_{\text{destination}} f_i^{out}[k]} \qquad\qquad k = \text{ all 15-minute time intervals} \qquad (6.2)$$

The scale factors corresponding to the final selection of flows are shown in Figure 6.8. The OD matrices ($\gamma_{ij}[k]$'s) are computed in FREQ as the solution to the following optimization problem:

For each $k$, given entering flows $f_i^{in}[k]$ and exiting flows $f_j^{out}[k]$, $\forall\ i$ and $j$, find $\gamma_{ij}[k]$'s that minimize a cost function, subject to:

$$\sum_j \gamma_{ij}[k] = f_i^{in}[k] \qquad \forall i$$

$$\sum_i \gamma_{ij}[k] = f_j^{out}[k] \qquad \forall j$$

Additional constraints are imposed to ensure positivity and that vehicles exit through ramps that are downstream of their origin on-ramps. FREQ's OD estimation function was used to convert the on-ramp and offramp demand data into a sequence of 20 OD matrices – one for each 15-minute time interval in the 5 hour period. Each of these matrices has dimensions $(22) \times (19) = (21$ on-ramps $+$ 1 mainline origin$) \times (18$ offramps $+$ 1 mainline destination$)$.

An intermediate step was then performed to incorporate the information of the percentage of HOV vehicles present in each of the source flows. As is explained in Section 7.1.5 of Chapter 7, each OD matrix in VISSIM applies to a specific *traffic composition*. Since the I-210 model includes two traffic compositions (MIX_TC and HOV_TC, defined in Section 7.1.5), each FREQ OD matrix spawned two VISSIM OD matrices, giving a total of 40 matrices. The following assumptions were made based on available data and on suggestions from Caltrans staff. They were sufficient to make the conversion from 20 to 40 OD matrices.

- The number of vehicles using the HOV lane at the upstream mainline boundary (Vernon St.) is a given time-varying fraction of the total (mixed-flow lanes plus HOV lane). This fraction, shown in Figure 6.9, was derived from PeMS data.

- In addition to the HOV vehicles in the HOV lane, 5% of the vehicles in the Vernon St. mixed-flow lanes are also HOV.

- 12% of the vehicles entering the freeway at on-ramps are HOV.

- Of the total number of HOV vehicles that reach the downstream mainline boundary, at Fair Oaks St., 20% are in mixed-flow lanes, and 80% are in the HOV lane.
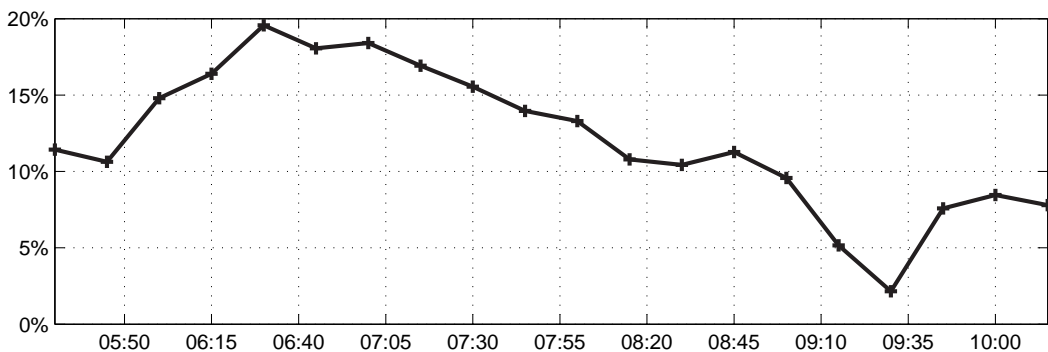


Figure 6.9: Percentage flow in the HOV lane at Vernon St.

# 6.8 Identification of recurring bottlenecks

The first step in the model calibration process is to identify the location and causes of congestion. Appendix B.1 contains three speed contour plots showing the congestion patterns for a heavy, a typical, and a light day of traffic on I-210. From these and other similar contour plots, three distinct problem areas, or bottlenecks, were identified. They are:

**B1**: Near Huntington St. (MP 33.049).
**B2**: Near the Rosemead and Michillinda St. ramps (MP 30.139).
**B3**: Near Hill St. (MP 26.8).

These three bottlenecks are illustrated in Figure 6.10. Mainline loop detector stations are depicted in the figure with a $\times$, $\circ$, or $\otimes$ symbol, depending on whether the station is characterized by heavy congestion (speed is often $< 40$ mph), by free-flow (speed $> 55$ mph), or by decreased speeds not reaching full congestion (speed is between 40 mph and 55 mph). Distances between ramps are marked on the figure, along with the number of mixed-flow lanes in each section. The number accompanying each on-ramp and offramp is a representative (approximately average) level of flow on the ramp when congestion begins.

The following conclusions were reached on the probable causes of congestion at each bottleneck:

**B1**: This bottleneck is not easily explained with a simple comparison of nominal capacities and demands. The Myrtle ramps make no net contribution to the amount of traffic on the freeway (600-600=0). The Huntington ramps supply about 500 vph to the mainline, but this should be easily absorbed by the auxiliary lane between Huntington and Santa Anita. The observed deceleration of the traffic stream must therefore be caused by a reduction in capacity near the Huntington ramps, or somewhere between Huntington and Santa Anita. Localized reductions in capacity have a variety possible causes, including grades, curves, reduced visibility, street signs, and direct oncoming sunlight. In this case, the most probable cause is the series of reverse curves between Myrtle and Huntington (as suggested by Caltrans staff).

**B2**: Bottleneck B2 is less stable than B1, in the sense that its location and congestion pattern changes from day to day. Congestion usually initiates somewhere near the Rosemead and Michillinda ramps (MPs 30.139 to 29.879), however, complete breakdown, with speeds in the 20's and 30's, only occurs upstream near the Baldwin on-ramp (MP 30.779). The Rosemead and Michillinda detectors sometimes register speeds decreasing as low as 40 mph, but seldom less than that. Congestion in this region is probably caused by the two heavy on-ramps from Rosemead and Michillinda, which add approximately 1700 vph to the freeway. These on-ramp flows should be easily accommodated by the two additional auxiliary lanes. However, it appears that this increased capacity is not being fully utilized, probably due to increased weaving in that area.

**B3**: Mainline traffic near Hill St. (MP 26.8) is usually slow, and sometimes fully congested. Traffic near Altadena (MP 28.03) almost always becomes completely congested. As with the previous two, bottleneck B3 is not easily explained by comparing demands

and nominal capacities, since the heavy flow from the Hill on-ramp is supported by an auxiliary lane. The observed congestion must therefore again be explained by a reduction in capacity. In this case, at least two probable causes exist: the S-shaped bend between Hill and Lake, and the heavy weaving that takes place in the 800-foot auxiliary lane feeding the Lake offramp.

## 6.9 Control hardware

Aside from the topographical and traffic data, the characteristics of the centralized control system were also considered important to the representation of I-210 in VISSIM. All of the relevant elements of the current system were gathered through correspondence with the District 07 On-ramp Metering Group. Following is a list of the elements that were incorporated in the VISSIM model:

1. Traffic Management Center (TMC): The TMC is the central location where freeway measurements are processed and control commands for each of the controlled on-ramps are generated. The TMC commands are updated every 30 seconds. The main function of the TMC is to generate target cycle lengths for the upcoming 30-second period. The TMC commands are based on average detector measurements from the previous 30-second period.

2. Communication delay: The commands issued by the TMC take 30 seconds to arrive to the individual control boxes.

3. SATMS chip: All controlled on-ramps on I-210 have either 1 or 2 signal heads. These are locally controlled by a SATMS chip located in the signal control box. The main features of the chip that are reproduced in VISSIM are:

   (a) Each signal can be in one of three modes: *normal*, *queue override*, or *green ball*. When in the normal mode, the cycle length commanded by the TMC is used. The queue override mode is activated whenever the on-ramp queue exceeds a maximum length. The green ball mode is activated if the flow and occupancy on the mainline are low.

   (b) The nominal duration of the green phase is 2 seconds. However, it may last as long as 5 seconds depending on the time it takes for the vehicle to clear the passage detector. This is to allow inattentive drivers enough time to react before the light turns red. Only one vehicle is allowed to pass per green phase. Platoon metering is not used on I-210.

   (c) The signal rests on red in the normal and queue override modes.

   (d) In the case of double signal heads, the release of vehicles is simultaneous. Both stoplights turn green whenever either of the two presence detectors register a waiting vehicle.

**BOTTLENECK B1**

**BOTTLENECK B2**
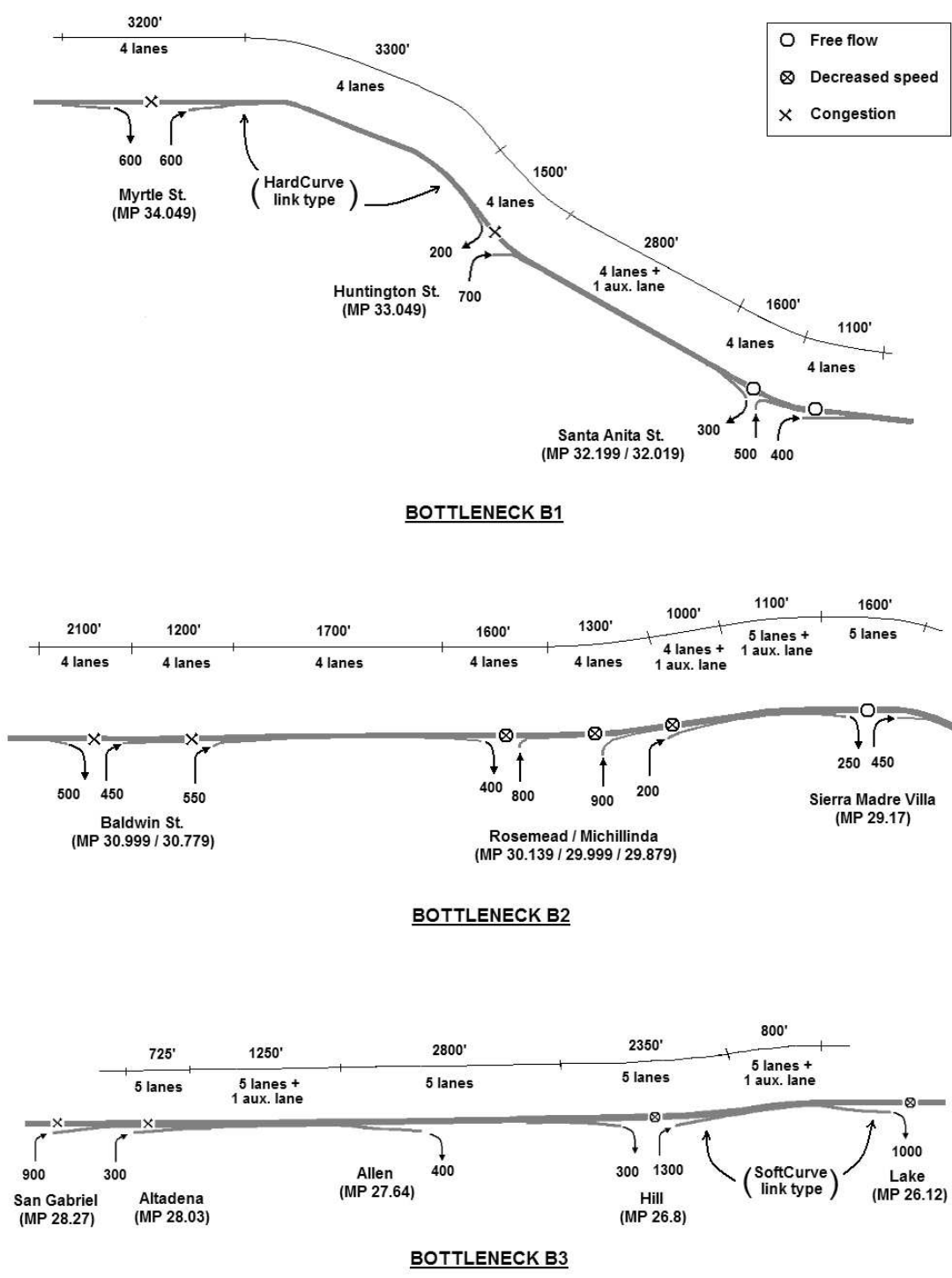
**BOTTLENECK B3**

Figure 6.10: Three major bottlenecks.

(e) The *green ball mode* is activated whenever the average mainline flow is less than 1500 vphpl *and* the average mainline occupancy is less than 15%. Control is then

94

suspended and the meter is set to green. The green ball mode has a minimum duration of 1 minute. After the first minute, the controller may return to the normal mode if the conditions on the mainline surpass any one of the two threshold values.

(f) The *queue override* mode is briefly summarized as follows: If for a particular on-ramp, at least one of its queue detectors remains occupied for more than 3.2 seconds *and* the speed on the mainline is greater than 35 mph, that on-ramp will enter the queue override mode. In this mode, the command rate issued by the TMC is ignored. Instead, the per-lane rate is increased by 120 vphpl (=2 vehicles per minute) at every TMC cycle, until the maximum rate of 900 vphpl (15 veh/min) is reached. If at any point the queue detector is not occupied for 3.2 seconds, the on-ramp is returned to the normal mode in a single step.

## 6.10   On-ramp configurations
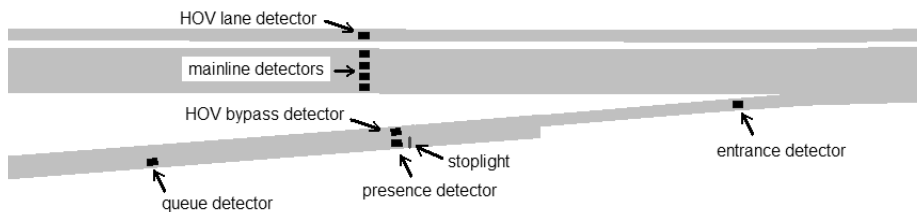


Figure 6.11: Loop detector layout on on-ramps.

Table 6.2 contains basic descriptions of each of the 20 controlled on-ramps in the test site. RH and LH in the second and third columns denote Right Hand and Left Hand on-ramps. Note that most of the on-ramps have one metered lane and an HOV bypass lane, 6 on-ramps have two metered lanes and no HOV bypass, and one on-ramp has a single metered lane and no HOV bypass. The storage capacities are given in total number of vehicles, not in vehicles per metered lane.

Each on-ramp is equipped with a set of loop detectors. A typical loop detector configuration for a single-metered lane on-ramp is shown in Figure 6.11. The *queue* detector is placed at the entrance to the on-ramp, and is used to trigger the queue override. The *presence* and *HOV bypass* detectors are placed side by side near the signal, in the metered and HOV bypass lanes respectively. The *entrance* detector measures the total flow that enters the freeway. The mainline detector bank is located upstream of the on-ramp merge, at approximately the same distance from the junction as the stoplight.

| Street Name | Number of metered lanes | HOV bypass | Number of lanes at gore | Storage capacity |
|---|---|---|---|---|
| Vernon | 1 RH | Yes LH | 1 | 28 |
| Irwindale NB | 1 RH | Yes LH | 1 | 21 |
| Irwindale SB | 1 RH | Yes LH | 1 | 27 |
| Mount Olive | 1 RH | Yes LH | 1 | 33 |
| Buena Vista | 1 RH | Yes LH | 1 | 19 |
| Mountain | 1 RH | Yes LH | 1 | 25 |
| Myrtle | 1 RH | Yes RH | 1 | 12 |
| Huntington | 1 RH | Yes RH | 1 | 26 |
| Santa Anita NB | 2 | No | 1 | 45 |
| Santa Anita SB | 1 RH | Yes RH | 1 | 21 |
| Baldwin NB | 1 RH | Yes LH | 1 | 14 |
| Baldwin-Foothill | 2 | No | 1 | 30 |
| Michillinda | 2 | No | 1 | 19 |
| Rosemead-Foothill | 2 | No | 1 | 17 |
| Rosemead | 1 RH | Yes LH | 1 | 10 |
| Sierra Madre Villa | 1 | No | 1 | 18 |
| San Gabriel | 1 RH | Yes LH | 1 | 23 |
| Altadena | 1 RH | Yes LH | 1 | 26 |
| Hill | 2 | No | 1 | 28 |
| Lake | 2 | No | 1 | 42 |

Table 6.2: On-ramp configurations.

# Chapter 7

# Calibration of VISSIM for I-210

This chapter describes the process through which the geometric and traffic data collected in Chapter 6 were encoded in VISSIM. Section 7.1 describes the VISSIM software, focusing on those features of the model that were used. Section 7.2 provides definitions for the model parameters that were varied in the calibration phase. Sections 7.3 and 7.4 describe the calibration methodology and the findings.

## 7.1 The VISSIM model

VISSIM is the microscopic/stochastic traffic simulator that was used to create the detailed model of I-210 West. In the past, it has been used mostly as a tool for the design of urban public transportation systems, but has been shown to be capable of reproducing freeway traffic behaviors as well. Its traffic model is based on the work of R. Wiedemann [100, 101], which combines a perceptual model of the driver with a vehicle model. The behavioral model for the driver involves a classification of reactions in response to the perceived relative speed and distance with respect to the preceding vehicle. Four driving modes are defined, as shown in Figure 7.1: Free driving, approaching, following, and braking. In each mode the driver behaves differently, reacting either to its following distance, or trying to match a prescribed target speed. These reactions result in a command acceleration given to the vehicle, which is processed according to its capabilities. Drivers can also make the decision to change lanes. This decision can either be forced by a routing requirement, for example when approaching an intersection, or made by the driver in order to access a faster-moving lane.

A useful feature in VISSIM is that it allows stochastic variations of several of its parameters, such as the desired speeds and accelerations. Stochastic sources of boundary flows (rates and compositions) are also supported. Randomness can further be introduced in the ability of the driver population to perceive changes in relative speeds and distances and to determine their mode of driving. More comprehensive descriptions of the VISSIM model and software can be found in [31, 87].

Traffic signals can be simulated, and are controlled in VISSIM by the Signal State Generator (SSG), which is a separate module from the traffic simulation module. One important feature of the SSG is that it is programmable – the user is allowed to specify the signal update logic using a C-language based script. Functions are provided for reading loop detector
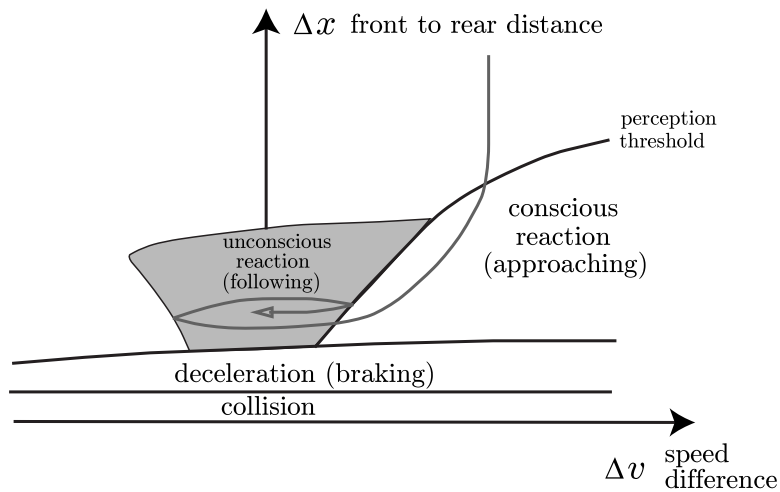
Figure 7.1: VISSIM's driver behavior model.

measurements and for setting the signal states. A trace file can be exported from the simulation to record loop detector and signal related variables. These traffic signaling features can be used, for example, on freeway on-ramps to simulate traffic-responsive metering.

### 7.1.1 Coding of the network geometry

As was described in Chapter 6, the relevant features of the I-210 test site were marked on a composite aerial photograph, which was downloaded from MapQuest (Figure 6.2). Scale was established on this image by matching landmarks with the scaled aerial photographs obtained from Caltrans HQ. Links and link connectors were then traced on this background image in VISSIM. A screenshot of VISSIM is shown in Figure 7.2.



Figure 7.2: Snapshot of VISSIM.

98

### 7.1.2 Control hardware

In addition to the freeway geometry, coding of the supply side of the model also entailed the placement of the control hardware elements: loop detectors and signal heads. In VISSIM, each signal head is associated with a *signal group*. All signal heads in the same group display the same signal status at all times. For I-210, a separate signal group was created for each signal head, in order to allow every on-ramp, and even different signal heads on the same on-ramp, to act independently. Every signal group, in turn, is associated with a *signal controlled junction* (SCJ). An SCJ can contain several signal groups. The control logic corresponding to a particular SCJ determines the signal status of all signal groups and signal heads within that SCJ. A single SCJ was used to control all of the signals in the I-210 model.

### 7.1.3 HOV lanes

Another important aspect of the network coding is the implementation of HOV lanes. VISSIM allows particular lanes of a link to be closed to certain vehicle types (vehicle types are defined in the next section). HOV-only restrictions were enforced by creating a separate vehicle type for the HOV vehicles, and closing the HOV-only lanes to all non-HOV types. This method was used for the HOV lanes on the mainline and HOV bypass lanes on the on-ramps.

### 7.1.4 Freeway connector

Almost all of the on-ramp merges were modeled following the method recommended in [87], where vehicles entering from the on-ramp join the mainline stream by changing lanes within a merge section. It was found however, that this approach only worked well for on-ramps with small or moderate flows. It failed for the heavy freeway connector from I-605 NB (MP36), where it produced a large queue on the ramp. An alternative configuration was designed to shift some of the burden of the merge away from the ramp and onto the mainline, by forcing a percentage of the mainline vehicles to evacuate the right-most lane upstream of the ramp junction, thereby opening space for the flow from I-605 NB. This was accomplished using VISSIM's *partial routing decisions* (see [87] for details).

### 7.1.5 Coding of traffic demands

**Vehicle types and Traffic compositions**

The vehicle population in VISSIM is categorized into *vehicles types*. A single type gathers vehicles that share common physical attributes. These attributes include model, minimum and maximum acceleration, minimum and maximum deceleration, weight, power, and length. All of these, except for model and length, are defined in VISSIM with probabilistic distributions (as opposed to scalars). Four vehicle types were created to model I-210: LOV, HOV, HGV_MED, and HGV_LARGE. The LOV type represents passenger vehicles with a single occupant. HOV vehicles have 2 or more occupants and are allowed to use the HOV and bypass lanes. The vehicle specifications for these two types are identical to those of the

default "CAR" type in VISSIM [87]. The HGV_MED and HGV_LARGE types represent, respectively, medium and large size trucks. Distributions for each of the four vehicle types are provided in Appendix C.3. *Traffic compositions* are the proportions of each vehicle type present in each of the source flows. Two traffic compositions were defined: MIX_TC for mixed-flow lane sources (93% LOV, 3.5% HGV_MED, 3.5% HGV_LARGE) and HOV_TC for HOV lane sources (100% HOV type).

## Dynamic assignment

VISSIM supports two different forms of input for the traffic demands. We chose to use its *dynamic assignment* function, which automatically determines inlet flows and routing information based on a user-supplied set of OD matrices. Each OD matrix is related to a single traffic composition, and to a particular 15-minute period of the simulation. The demand specification for the I-210 model consists of 40 OD matrices - 2 traffic compositions (MIX_TC and HOV_TC) times 20 time intervals. Each OD matrix has entries in the $ij^{th}$ position indicating the average flow of a given traffic composition entering the network at the $i^{th}$ on-ramp, with destination at the $j^{th}$ offramp, during a particular 15-minute period. Routes, or traffic assignments, are generated by the *dynamic assignment module* by assigning a cost to every route available to each OD pair, and then choosing the route with minimum cost. The cost function in VISSIM includes terms penalizing the total distance, total travel time, and a link cost. This last term serves to model factors not covered by the first two, such as tolls. The link cost was used here, as explained below, to encourage the use of the HOV lanes by HOV vehicles.

## HOV lanes and link costs

The idea behind dynamic assignment is that repeated simulations using this method for generating routes, and updating the travel time cost between iterations, should eventually converge to an equilibrium solution, in the sense that traffic assignments and travel times should eventually stop changing between iterations. In the case of I-210, the only routing decision to be made is whether and where the HOV vehicles will access the HOV lane. The calibration runs are based on a single iteration of dynamic assignment. Travel time was therefore not considered in the selection of routes for HOV vehicles (this is because travel time is only known after the first iteration). Instead, the HOV lane was given a favorable cost by using the link cost coefficient. A separate link cost coefficient can be assigned to each vehicle type. The LOV vehicle type's link cost coefficient was set to 0.0, whereas the HOV type was given a value of 1.0. In computing a cost for each route, the program multiplies this coefficient by a link cost associated with each link in a given route, and adds them up. HOV lanes were given a preferred status by attaching a lesser link cost to HOV lanes, as compared to mixed-traffic lanes. Thus, the minimum-cost route available to HOV-type vehicles was always to enter the HOV lane at the gate nearest to its origin, and to exit it at the gate nearest to its destination. Non-HOV vehicles were declined the use of HOV lanes with type-specific lane closures (mentioned in Section 7.1.1).

### 7.1.6  VISSIM output

Two output files were used to generate the contour and time-series plots. First, the C-based script produced a trace file that contains 5-minute averages of flow and occupancy measurements for all of the loop detectors in the model. Second, VISSIM's *link evaluation* was used to export space-aggregated traffic variables, such as link flow, density, and speed, also at 5-minute intervals.

# 7.2  Changeable model parameters - Default values

Section 7.1.5 listed the model parameters related to the physical attributes of the *vehicle*. These were assigned separately for each vehicle type. Fixing the vehicle population, we now look at the parameters of the *driver* model. We have assumed that driver behavior is not correlated with vehicle type, but instead with the position of the driver/vehicle unit on the freeway. For example, drivers might behave differently on curved sections, as compared to straight sections. Thus, the parameters described in this section apply equally to all vehicle types, but can be adjusted for each *link type*. Link types are analogous to vehicle types. They gather links with similar driver behavior parameters. Six link types were created to model I-210. These are described in Section 7.3. The driver behavior parameters that were changed from their default values in each link type are described below. This is a subset of the total number of adjustable driver behavior parameters available in VISSIM. The complete list can be found in [87].

### 7.2.1  Necessary lane change

The dynamic assignment module provides to each driver a sequence of links to follow that will take it from its origin to its destination. The parameters related to necessary lane changes dictate how far in advance each driver will be able to anticipate the next bifurcation or lane drop on its list, and how aggressively that driver will change lanes to reach it. The first two items below – *look-back distance* and *emergency stop distance* – are the only driver behavior parameters that are not grouped into link types, but must be specified for each *link connector* separately[1].

- *Look-back distance*: Distance in anticipation of a bifurcation that the driver will begin maneuvering towards the desired lane. The default value is 200 m.

- *Emergency stop distance*: Distance before the bifurcation where the driver will stop if it has not reached its desired lane. The default value is 5 m.

- *Waiting time before diffusion*: A vehicle that has come to a halt at the emergency stop position will wait at most this amount of time for a gap to appear in the adjacent lane. After the waiting time has elapsed, the vehicle is removed (diffused) from the simulation. The default value is 60 seconds.

---

[1] In VISSIM the *link connector* is the boundary between two links.

- *Acceleration parameters*: The aggressiveness of the lane change maneuver can be adjusted by changing the deceleration applied by the vehicle (Own) and the acceptable deceleration capabilities of the vehicle in the adjacent lane (Trailing). These two numbers vary continuously with distance to the emergency stop position and are determined by six parameters. The relation between the six parameters is illustrated in Figure 7.3. Default values are shown in Table 7.2.
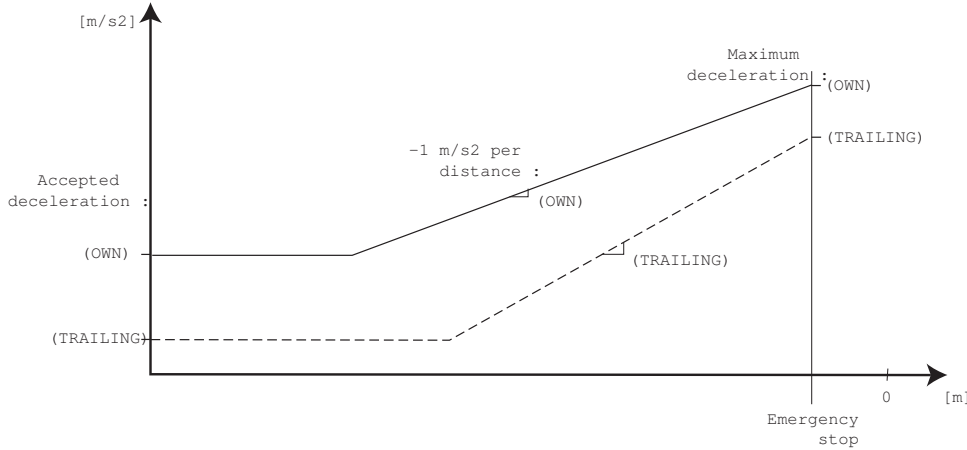


Figure 7.3: Necessary lane change parameters.

## 7.2.2 Vehicle following behavior

VISSIM includes two versions of the Wiedemann model: urban driver and freeway driver. Only the freeway driver type was used. The car-following mode of the freeway driver model involves 10 tunable parameters: CC0 through CC9. Below are described only those CC-parameters that were modified from their default values.

- *CC0 and CC1*: Coefficients used in the calculation of the safe bumper-to-bumper distance (in [m]): dx_safe=CC0+$v$·CC1, where $v$ (in [m/s]) is the speed of the trailing vehicle. According to [87], CC1 is the parameter with the strongest influence on freeway capacity. In fact, it can be related almost directly to capacity by noting that (dx_safe+vehicle length)×capacity = free-flow speed. With reasonable values of capacity, dx_safe, and free-flow speed, and default CC0, this calculation gives CC1=1.5 seconds. Default values are CC0=1.5 m and CC1=0.9 s.

- *CC4 and CC5*: These are dimensionless parameters influencing the coupling between leader and follower accelerations. Smaller magnitudes result in driver behaviors that are more sensitive to changes in the speed of the preceding vehicle. It is recommended in [87] that these two parameters have opposite signs and equal magnitude. Default values are CC4=-0.35 and CC5=0.35. The absolute value of CC4 (or CC5) can be understood as the inverse of a stiffness coefficient between consecutive vehicles.

102

These three CC-parameters (CC0, CC1, and the CC4/CC5 pair) were used to model the curvature-induced capacity reductions that are the supposed culprits of bottlenecks B1 and B3 (Section 6.8). We can infer from their definitions that increasing CC0, CC1, or the absolute value of CC4/CC5 will lead to reductions in freeway capacity.

## 7.3  Parameter variations

With model inputs (network supply and traffic demand) fixed as described in Section 7.1, an initial simulation experiment was conducted using default driver behavior parameters. The resulting speed contour plot is shown in Figure 7.4. The immediate observation here is that there is a severe blockage near the downstream end of the freeway that produces a queue which quickly overruns the entire site. This problem was caused by the large number of vehicles attempting to exit through the last two offramps (the I-210 and SR-710 freeway connectors), but were unable to complete the necessary lane changes before reaching and stopping at the emergency stop position. Several adjustments to the routing-imposed lane change parameters were made to correct this problem.

| time | 39.159 | 38.209 | 38.069 | 36.589 | 35.409 | 34.899 | 34.049 | 33.049 | 32.199 | 32.019 | 30.999 | 30.779 | 30.139 | 29.999 | 29.879 | 29.17 | 28.27 | 28.03 | 26.8 | 26.12 | 25.68 | 25.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5:30 | 63 | 34 | 24 | 62 | 61 | 63 | 62 | 63 | 63 | 63 | 62 | 64 | 62 | 63 | 63 | 63 | 62 | 63 | 64 | 63 | 30 | 61 |
| 5:45 | 12 | 6 | 7 | 62 | 58 | 63 | 62 | 63 | 63 | 63 | 62 | 63 | 63 | 64 | 64 | 63 | 62 | 64 | 64 | 62 | 8 | 61 |
| 6:00 | 6 | 5 | 6 | 62 | 59 | 63 | 62 | 63 | 63 | 63 | 62 | 63 | 60 | 62 | 62 | 62 | 59 | 62 | 51 | 6 | 5 | 61 |
| 6:15 | 5 | 6 | 8 | 62 | 46 | 61 | 62 | 62 | 62 | 63 | 59 | 61 | 59 | 63 | 63 | 62 | 60 | 63 | 5 | 4 | 4 | 62 |
| 6:30 | 7 | 7 | 8 | 63 | 41 | 61 | 60 | 61 | 60 | 62 | 52 | 58 | 51 | 54 | 59 | 47 | 14 | 4 | 3 | 4 | 4 | 63 |
| 6:45 | 7 | 7 | 9 | 61 | 41 | 59 | 60 | 59 | 57 | 59 | 45 | 40 | 24 | 21 | 11 | 3 | 4 | 3 | 3 | 3 | 4 | 63 |
| 7:00 | 6 | 7 | 8 | 61 | 51 | 62 | 62 | 62 | 29 | 23 | 3 | 2 | 2 | 3 | 2 | 2 | 3 | 3 | 3 | 4 | 5 | 62 |
| 7:15 | 5 | 5 | 6 | 62 | 53 | 60 | 61 | 9 | 2 | 2 | 4 | 3 | 3 | 2 | 3 | 4 | 3 | 3 | 4 | 5 | 5 | 63 |
| 7:30 | 5 | 7 | 8 | 61 | 50 | 45 | 5 | 2 | 2 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 4 | 3 | 4 | 5 | 6 | 61 |
| 7:45 | 7 | 6 | 8 | 61 | 14 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 3 | 4 | 5 | 5 | 6 | 6 | 5 | 62 |
| 8:00 | 7 | 7 | 8 | 43 | 2 | 2 | 2 | 2 | 3 | 3 | 4 | 4 | 4 | 5 | 4 | 4 | 4 | 3 | 3 | 5 | 6 | 61 |
| 8:15 | 7 | 7 | 8 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 61 |
| 8:30 | 4 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 4 | 5 | 6 | 6 | 61 |
| 8:45 | 3 | 3 | 3 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 5 | 4 | 5 | 5 | 4 | 4 | 5 | 5 | 63 |
| 9:00 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 4 | 3 | 4 | 5 | 4 | 4 | 5 | 4 | 5 | 5 | 6 | 62 |
| 9:15 | 4 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 3 | 3 | 4 | 4 | 4 | 5 | 4 | 4 | 5 | 4 | 4 | 5 | 5 | 63 |
| 9:30 | 3 | 3 | 4 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 4 | 3 | 4 | 5 | 4 | 4 | 5 | 5 | 5 | 6 | 6 | 61 |
| 9:45 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 5 | 4 | 5 | 5 | 4 | 4 | 5 | 5 | 4 | 5 | 6 | 61 |
| 10:00 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 4 | 5 | 6 | 6 | 56 |
| 10:15 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 4 | 4 | 6 | 5 | 6 | 7 | 7 | 61 |

Direction of traffic →

Figure 7.4: Speed contour plot with default driver behavior parameters (in [mph]).

### 7.3.1  Adjustments to the look-back distance

It was determined that the default look-back distance of 200 m was too small for large numbers of vehicles crossing over several lanes of traffic to reach their exits. On the other hand, increasing this value too much had the unrealistic effect of bunching up all of the exiting vehicles in the right-most lane, far upstream of their intended offramp. These vehicles then obstructed other upstream offramps and on-ramps. It was therefore necessary to tune the look-back distances individually for each offramp, in a way that allowed vehicles sufficient weaving space while ensuring that these lane-change regions did not overlap. The list of

tuned look-back distances is given in Table 7.1. Figure 9.9 in Appendix C.1 shows the contour plot resulting from this adjustment. Note that the offramp blockage problem was corrected almost completely by tuning the look-back distances.

| MP | Street Name | Mainline look-back | Offramp look-back |
|---|---|---|---|
| 38.209 | Irwindale | 800' | 800' |
| 38 | I-605-SB | 1450' | 1450' |
| 37 | Mount Olive | 800' | 800' |
| 36 | Buena Vista | 400' | 400' |
| 35.409 | Mountain | 200' | 400' |
| 34.049 | Myrtle | 800' | 800' |
| 33.049 | Huntington | 800' | 800' |
| 32.019 | Santa Anita | 200' | 800' |
| 30.779 | Baldwin | 800' | 800' |
| 30.5 | Rosemead | 800' | 800' |
| 29.17 | Sierra Madre Villa | 800' | 800' |
| 28.27 | San Gabriel | 800' | 800' |
| 27.64 | Allen | 800' | 800' |
| 26.8 | Hill | 800' | 800' |
| 26.12 | Lake | 800' | 800' |
| 25.68 | Marengo | 600' | 700' |
| 25.6 | I-210 connector | 250' | 700' |
| 25.5 | SR-710 connector | 200' | 700' |

Table 7.1: Adjusted look-back distances for mainline/offramp bifurcations.

## 7.3.2 Adjustments to the lane change parameters

The parameters influencing the aggressiveness of necessary lane changes were modified in order to further reduce the number of vehicles reaching the emergency stop position. The final selection is provided in Table 7.2.

| | Default | | Final | |
|---|---|---|---|---|
| | Own | Trailing | Own | Trailing |
| Maximum Deceleration [m/s$^2$] | -4.00 | -3.00 | -5.00 | -5.00 |
| -1 m/s2 per Distance [m] | 200 | 200 | 10 | 10 |
| Accepted Deceleration [m/s$^2$] | -1.00 | -0.50 | -4.00 | -4.00 |

Table 7.2: Adjusted lane change parameters.

### 7.3.3   Adjustments to the Waiting time before diffusion

Another modification that was found useful for eliminating the offramp blockages was to decrease the *waiting time before diffusion* parameter from its default 60 seconds to 1 second. With this setting, vehicles that reached the emergency stop position were immediately removed from the simulation, thereby minimizing the obstruction to the freeway. Eliminating these vehicles has little impact on the total travel time, since they are few and very close to their exit anyway. However, this adjustment is only recommended after the number of affected vehicles has been minimized by tuning the look-back and lane change parameters. Also, one should be careful not to affect other parts of the network where longer waiting times are expected. In the case of I-210, vehicles attempting to *enter* the freeway also frequently reached the emergency stop position at the end of the on-ramp acceleration lanes. To avoid these vehicles from being evaporated, a set of *merge* link types was created. These mimic their non-merge counterparts in all features except for the waiting time, which was kept at 60 seconds for the merge types (see Table 7.4). Merge link types were used on all on-ramps and on-ramp merge areas.

### 7.3.4   Link types - Adjustments to the CC parameters

The remainder of the calibration effort focused on finding a suitable set of values for the CC-parameters defined in Section 7.2. Three separate sets of CC-parameter values were defined: Freeway, HardCurve, and SoftCurve. Each was paired with a corresponding merge link type (with a 60 second diffusion time), giving a total of 6 link types. The Freeway and Freeway Merge types were used almost everywhere. The HardCurve and SoftCurve link types were applied only to the curved sections that affect bottlenecks B1 and B3 respectively (see Figure 6.10). As is described in the next section, one of the findings of the calibration is that only modest adjustments to the CC-parameters were required to produce the desired simulation response. Also, that capacity reductions due to curvature can be reproduced with changes to the CC1 parameter alone.

## 7.4   Calibration goals and parameter selection

Having assembled the on-ramp and offramp flow inputs using data from several different days, it is not obvious how the simulation results should be evaluated. The usual method of computing an error norm with respect to the measured data, and tuning the model parameters to minimize that norm is not applicable in this case because of the composite nature of the input data. The question arises, should a single typical day be used, or a composite day, as was done with the boundary flows? Added to this difficulty is the fact that none of the data sets considered as typical had a complete set of mainline measurements. Furthermore, there seems to be more variability in the mainline measurements than appears in the on-ramp flows, suggesting the influence of unseen factors, such as weather, day-to-day variations in driver behavior, traffic incidents, etc.

Instead, the goal for the calibration was to match more qualitative aspects of the freeway operation. These are:

1. location of the three identified bottlenecks,

2. initial and final times for each of the three mainline queues,

3. extent of the queues,

4. utilization of the HOV lane,

5. on-ramp performance.

The first three items on this list pertain to the simulated response of the mixed-flow lanes. Target values for these were extracted from contour plots similar to those in Appendix B.1, and are listed in Table 7.3. The goal for the HOV lane was to approximately match the flow values from PeMS. For the on-ramps, the only objective was to avoid large on-ramp queues that might obstruct the vehicle sources.

The parameter selection methodology consisted of iterated runs, visual evaluation of the results using speed contour plots (e.g. Figure 7.6), and manual adjustments of the parameters. These adjustments were limited to the CC-parameters described in Section 7.2, and were aided by the bottleneck analysis of Section 6.8 and by the physical interpretation of the parameters of Section 7.2. The iterative procedure was stopped when all of the qualitative calibration goals were met. This approach was favored over a more exhaustive automated search method because of the potentially huge number of parameter variations, as well as the approximately 3-hour running time[2], and the advantage that it leads to a more sensible result.

The final selection of driver behavior parameters is given in Table 7.4. This parameter set is the most parsimonious among those sets that also met the calibration goals. Notice that the CC4/CC5 parameter was increased (in absolute value), but was kept uniform throughout the freeway. It was found that this parameter, in addition to CC1, also has an important influence on capacity. Its default value of -0.35/0.35 produced almost no congestion. The CC0 parameter was increased globally from 1.5 to 1.7. As expected from its definition, CC0 was more influential at low speeds (i.e. within the mainline queues), and was used to regulate the queue lengths. The CC1 parameter on the other hand, was changed only locally, at two locations. The HardCurve link type was used on the reverse curve near Huntington St. and the SoftCurve type was used on the curved section between Hill and Lake St. (see Figure 6.10). CC1 was adjusted in both cases to achieve the correct activation times for bottlenecks B1 and B3. Interestingly, bottleneck B2 did not require a separate CC1 value. This result supports the interpretations provided in Section 6.8 of Chapter 6 for the causes of the three bottlenecks; that B1 and B3 are probably influenced by curvature, whereas B2 is probably due mostly to weaving.

### 7.4.1 On-ramp response

One of the qualitative goals for the calibration was to avoid unrealistic queues on the on-ramps that might obstruct the vehicle sources. The only on-ramp queuing problem that arose

---

[2]Using a PC/Windows XP, 2.6 GHz, 512 Mb machine.

|          | Bottleneck | Location | Start time | End time | Queue length |
|----------|-----------|----------|-----------|----------|--------------|
|          | B1 | MP 33.049 | 6:00 - 6:30 | 10:00 - 10:30 | To MP 39.159 |
| Measured | B2 | MP 30.779 / 30.139 | 6:45 - 7:15 | 9:00 - 9:45 | Into B1 |
|          | B3 | MP 28.03 / 26.8 | 7:00 - 7:30 | 9:15 - 9:45 | To MP 29.17 |
|          | B1 | MP 33.049 | 6:00 | 10:15 | To MP 39.159 |
| Simulated | B2 | MP 30.779 | 7:00 | 9:45 | Into B1 |
|          | B3 | MP 26.8 | 7:15 | 9:30 | To MP 29.17 |

Table 7.3: Measured and model predicted congestion pattern.

| Link type | CC0 | CC1 | CC4 / CC5 | Waiting time |
|-----------|-----|-----|-----------|--------------|
| Default | 1.5 | 0.9 | -0.35 / 0.35 | 60 |
| Freeway | 1.7 | 0.9 | -2.0 / 2.0 | 1 |
| SoftCurve | 1.7 | 1.1 | -2.0 / 2.0 | 1 |
| HardCurve | 1.7 | 1.4 | -2.0 / 2.0 | 1 |
| Freeway Merge | 1.7 | 0.9 | -2.0 / 2.0 | 60 |
| SoftCurve Merge | 1.7 | 1.1 | -2.0 / 2.0 | 60 |
| HardCurve Merge | 1.7 | 1.4 | -2.0 / 2.0 | 60 |

Table 7.4: Default and calibrated CC values.

was on the freeway connector from I-605 NB (MP 36). As was mentioned in Section 7.3, this was corrected at an earlier stage with partial routing decisions and was not a factor in tuning CC-parameters. All other on-ramps were checked by comparing the supplied on-ramp flows with the simulated on-ramp flows. These were a close match in all cases (see Appendix C.4), indicating that none of the vehicles sources were obstructed by overflowing on-ramp queues.

### 7.4.2 HOV lane response

The goal of matching the utilization of the HOV lane was verified by checking the simulated HOV lane flows. Samples of simulated and field-measured HOV lane flows are shown in Figure 7.5. Recall that the upstream boundary flows (at Vernon) are an input to the model. The differences at other locations may reflect modeling errors, such as errors in the provided percentage of HOV vehicles at on-ramps, and/or errors in the modeling of route choice by HOV drivers (Section 6.7). In general, the result is considered a sufficiently good match for the control purposes of this model. However, this aspect of the model might be improved with a refinement of the HOV input percentages, and/or more iterations of VISSIM's dynamic assignment routine.

### 7.4.3 Mixed-flow lane response

The bulk of the calibration effort was dedicated to matching the response of the mixed-flow lanes, in terms of the start time, end time, and extent of the queue generated by each of the
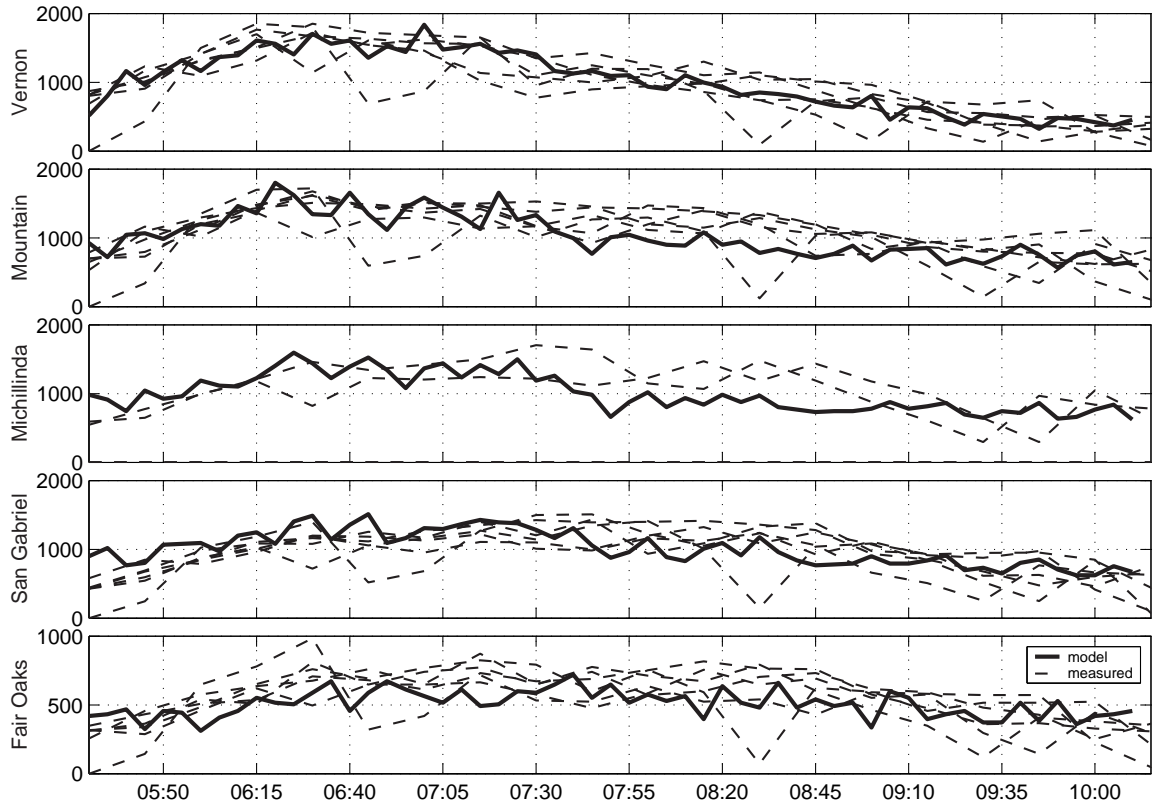
Figure 7.5: Measured and simulated HOV flows (in [vph]).

three major bottlenecks. The iterative procedure was stopped when all of the 9 indicators for the mixed-flow response fell within their target ranges. Target and simulated values for these 9 indicators are given in Table 7.3. The resulting speed contour plot, shown in Figure 7.6, is compared to the typical PeMS contour of Appendix B.1. Notice that the model has approximately matched the period of activation and queue length for the three bottlenecks. This was accomplished with a few global changes to the default parameter values, and with a couple local changes that were based on the analysis of field data and freeway geometry.

## 7.5   Random seed variations

The calibrated parameter set was run with 10 different *random seeds*. The random seed affects the realization of the stochastic quantities in VISSIM, such as inlet flows and vehicle capabilities. Contour plots for three examples are shown in Appendix C.2. Average percent variations in several simulation input and output variables resulting from random seed variations are shown in Table 7.5 and Figure 7.7.
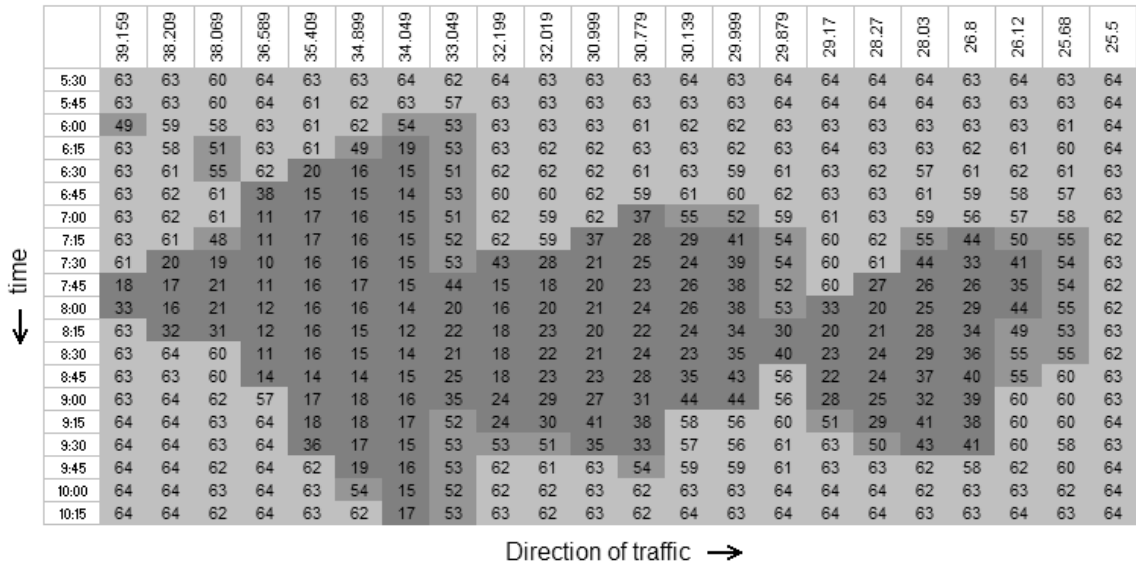
Figure 7.6: Contour plot with final parameters selection.

| Quantity | Average Value | % Variation |
|---|---|---|
| On-ramp flow | e.g. Figure 7.7 | 12.20% |
| Offramp flow | e.g. Figure 7.7 | 14.04% |
| Average speed | Figure 7.7 | 2.26% |
| Average volume | Figure 7.7 | 1.05% |
| Total Passenger Hours | 22,482 veh.hr | 1.56% |
| Total Passenger Kilometers | 1,539,700 veh.km | 0.09% |

Table 7.5: Variation in model output due to changes in the random seed.
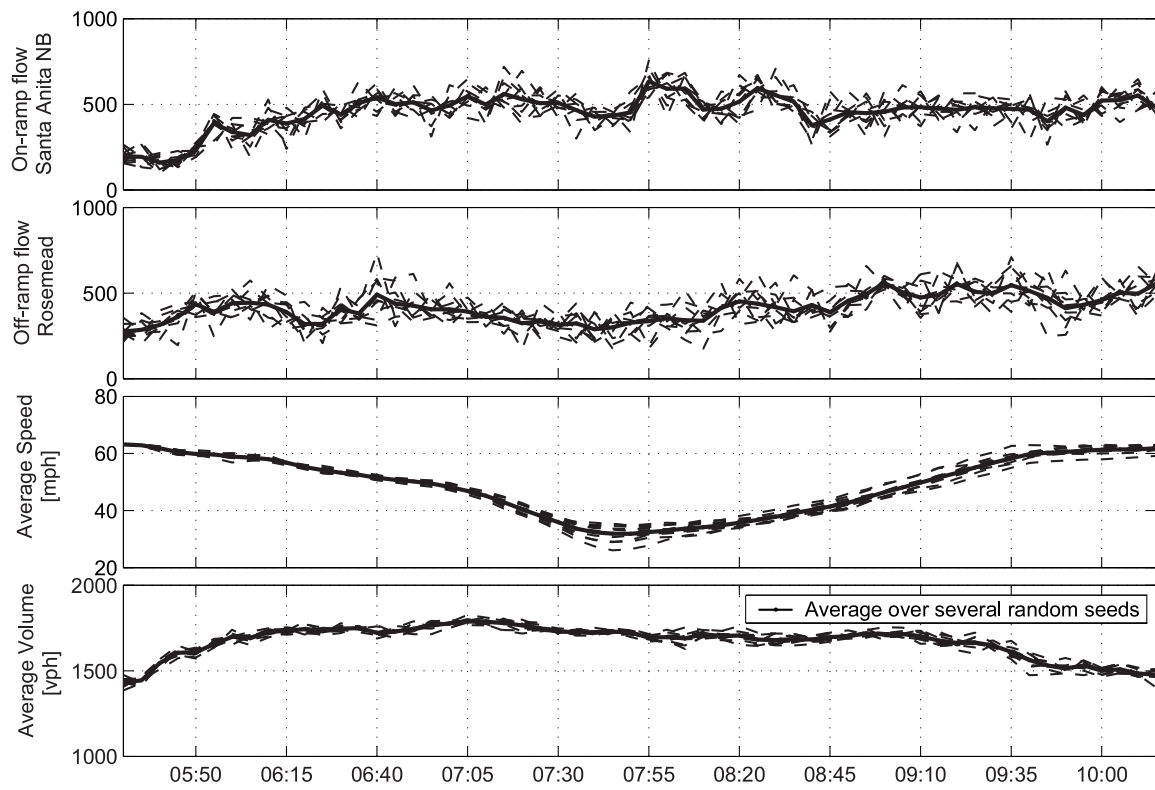
Figure 7.7: Mainline speed and flow measurements with several random seeds.

# Chapter 8

# On-ramp Metering Tests with VISSIM

This chapter reports a series of controlled simulation runs conducted with the VISSIM model of I-210 described in Chapters 6 and 7. The objective is to compare several options of on-ramp metering, and to tune their parameters for this particular test site. The candidate metering strategies are fixed-rate control, Alinea, and %-Occ. One of the issues pertaining to the potential use of Alinea is whether it can be employed when the feedback sensor is placed upstream of the on-ramp, and not downstream as envisioned by its creators. Chapter 4 partially answered this question with the finding that the upstream placement is actually preferred during congestion. The next question is how then to tune Alinea when the only upstream detectors are available.

An issue for any candidate controller on I-210 is that its commanded rate may be overridden if the on-ramp queue threatens to spill onto the streets (queue override), or if the freeway is relatively empty (green ball override). These experiments also attempt to determine the effect of these rules on controller performance and tuning.

## 8.1   Performance Measures

The performance measures used in this chapter are defined here in terms of variables exported by VISSIM's *link evaluation* output file and *DDE trace file*. The Total Travel Time and Total Travel Distance defined in Section 5.1 are referred to here as the Total Vehicle Hours (`TVH`) and Total Vehicle Kilometers (`TVK`), to distinguish them from the the Total Passenger Hours and Total Passenger Kilometers. The latter are the total trip times and distances incurred by individual *passengers*, instead of *vehicles*. VISSIM's *link evaluation* file contains 5-minute average densities, flows, and speeds for every link in the model, including on-ramp, mainline, and offramp links. Separate variables are exported for each vehicle type (LOV, HOV, and trucks). LOV and truck type densities, flows, and speeds are aggregated and denoted with $\rho_{ik}^l$, $f_{ik}^l$, and $v_{ik}^l$ respectively. Variables for the HOV type are denoted $\rho_{ik}^h$, $f_{ik}^h$, and $v_{ik}^h$. The sub-indices $i$ and $k$ in these variables indicate the link and 5-minute time interval respectively.

**On-ramp, Freeway, and Total Passenger Hours** (`OPH`, `FPH`, `TPH`) :

The Total Passenger Hours (`TPH` in [pass.hr]) is the sum of the On-ramp Passenger Hours (`OPH`) and the Freeway Passenger Hours (`FPH`). The formulas for these three quantities are:

$$\texttt{OPH} \;=\; \sum_{i}^{\text{on-ramps}} \sum_{k}^{\text{time}} (\, \alpha^h \rho_{ik}^h \;+\; \alpha^l \rho_{ik}^l \,) \, L_i \, \Delta t$$

$$\texttt{FPH} \;=\; \sum_{i}^{!\text{on-ramps}} \sum_{k}^{\text{time}} (\, \alpha^h \rho_{ik}^h \;+\; \alpha^l \rho_{ik}^l \,) \, L_i \, \Delta t$$

$$\texttt{TPH} \;=\; \texttt{OPH} \;+\; \texttt{FPH}$$

$L_i$ is the length in kilometers of section $i$. $\Delta t$ is the length of the data collection time interval in hours; $\Delta t = 5$ min $= 1/12$ hr. Coefficients $\alpha^h$ and $\alpha^l$ are the average number of passengers in HOV and LOV (including truck) type vehicles. They were assigned values of $\alpha^h = 2.5$ and $\alpha^l = 1.2$. %`TPH` is the percent improvement in passenger hours over the uncontrolled scenario:

$$\%\texttt{TPH} = 100 \times \frac{\texttt{TPH} - \texttt{TPH}_g}{\texttt{TPH}_g}$$

where $\texttt{TPH}_g$ is the passenger hours without on-ramp control.

**On-ramp, Freeway, and Total Passenger Kilometers** (`OPK`, `FPK`, `TPK`) :

The Total Passenger Kilometers (`TPK` in [pass.km]) is the sum of the On-ramp Passenger Kilometers (`OPK`) and the Freeway Passenger Kilometers (`FPK`):

$$\texttt{OPK} \;=\; \sum_{i}^{\text{on-ramps}} \sum_{k}^{\text{time}} (\, \alpha^h f_{ik}^h \;+\; \alpha^l f_{ik}^l \,) \, L_i \, \Delta t$$

$$\texttt{FPK} \;=\; \sum_{i}^{!\text{on-ramps}} \sum_{k}^{\text{time}} (\, \alpha^h f_{ik}^h \;+\; \alpha^l f_{ik}^l \,) \, L_i \, \Delta t$$

$$\texttt{TPK} \;=\; \texttt{OPK} \;+\; \texttt{FPK}$$

**Average Mainline Speed** (`AMS`) :

The average mainline speed (`AMS` in [mile/hr]) is computed as the mean of the per link speeds over time and space:

$$\texttt{AMS} \;=\; \frac{1}{n_k} \sum_{k}^{\text{time}} \frac{\sum_i L_i v_{ik}}{\sum_i L_i} \tag{8.1}$$

Both internal summations in Eq. (8.1) are over all mainline links. $n_k$ is the number of 5-minute time intervals in the summation. As explained below, only measurements from the 5-hour period from 5:30 am to 10:30 am were used to compute `AMS`.

**Average Throughput** (`ATh`) :

The `TPH`, `TPK`, and `AMS` are computed with *link* data contained in the *link evaluations* file. Point-wise detector data printed in the *DDE trace file* is used to compute the throughput,

or average mainline flow over the 5-hour period:

$$\texttt{ATh} \ = \ \frac{1}{n_{md}} \frac{1}{n_k} \sum_{k}^{\text{time}} \sum_{i}^{\text{mainline}} \frac{\text{vol}(i,k)}{\Delta t} \tag{8.2}$$

$n_{md} = 100$ is the number of mainline loop detectors in the test site. $\text{vol}(i,k)$ is the number if vehicles registered by detector $i$ during time interval $k$. $\Delta t = 5$ min.

The simulation time for all of the runs reported in this chapter is 60,000 seconds. The first 900 seconds are *warmup time*, used to create a non-empty initial condition. The demands for the 5:30-5:45 am period were applied during the warmup time. The period of interest, from 5:30 am to 10:30 am, covers simulation seconds 900 to 18,900. All source flows were turned off after second 18,900 in order to allow the freeway to discharge. This period, from 18,900 to 60,000, is referred to as the *cool down* period. Output data from the full 60,000 seconds was used to compute the travel time and travel distance performance measures: OPH, FPH, TPH, OPK, FPK, and TPK. Only measurements collected during the period of interest, i.e. excluding warmup and cool-down periods, were used to find AMS and ATh.

Measures for the uncontrolled freeway with a number of different random seed values are provided in Table 8.1.

| Random seed | OPH [pass.hr] | FPH [pass.hr] | TPH [pass.hr] | TVH [veh.hr] | TPK [pass.km] | TVK [veh.km] | AMS [mph] | ATh [vphpl] |
|---|---|---|---|---|---|---|---|---|
| 11 | 5,300 | 24,779 | 30,079 | 22,731 | 2,161,429 | 1,597,989 | 49.6 | 1,662 |
| 15 | 5,313 | 24,199 | 29,512 | 22,254 | 2,165,870 | 1,598,657 | 50.4 | 1,659 |
| 22 | 5,236 | 24,253 | 29,489 | 22,232 | 2,170,573 | 1,600,880 | 50.2 | 1,662 |
| 28 | 5,265 | 25,001 | 30,266 | 22,883 | 2,164,348 | 1,600,268 | 49.0 | 1,666 |
| 35 | 5,516 | 25,061 | 30,577 | 23,120 | 2,167,644 | 1,602,424 | 49.0 | 1,666 |
| 41 | 5,305 | 24,064 | 29,369 | 22,167 | 2,163,746 | 1,599,573 | 50.4 | 1,663 |
| 42 | 5,412 | 25,282 | 30,693 | 23,227 | 2,161,107 | 1,598,105 | 48.8 | 1,660 |
| 52 | 5,215 | 23,779 | 28,994 | 21,873 | 2,158,866 | 1,595,752 | 50.9 | 1,656 |
| 66 | 5,729 | 26,175 | 31,904 | 24,184 | 2,162,793 | 1,600,595 | 47.8 | 1,659 |
| 73 | 5,531 | 25,723 | 31,254 | 23,648 | 2,165,615 | 1,600,901 | 48.0 | 1,665 |
| mean | 5,382 | 24,832 | 30,214 | 22,832 | 2,164,199 | 1,599,514 | 49.4 | 1,662 |
| % dev. | 2.45 | 2.48 | 2.40 | 2.54 | 0.12 | 0.09 | 2.35 | 1.02 |

Table 8.1: Performance measures with 9 different random seeds.

## 8.2 The TMC and field controllers

The Traffic Management Center (TMC) is the "brain" of the freeway management system. It is in the TMC that freeway loop detector measurements are collected and processed, and the control commands for each of the on-ramps are calculated. These commands are

updated and distributed to the on-ramp control boxes every 30 seconds. In the case of I-210, a chip inside each box processes the TMC command and makes the decision to either implement or *override* it, depending on local conditions. This section describes the important features of the on-ramp control system currently installed on I-210 West. The control logic is schematically depicted in Figure 8.1. It was implemented in VISSIM using the software's C-based scripting feature.

The task of the field controller chip is to adjust and enforce the control rate issued by the TMC. In Figure 8.1, the TMC commanded rate for period `k` is denoted `TMCrate[k]`. It is computed according to whatever control algorithm is being used – %-Occ, Alinea, etc. The control loop starts at point `A` at the top of the left side of the diagram, and ends at point `A` at the bottom right corner. Each on-ramp operates in one of three modes: *normal*, *queue override*, or *green ball*. In the normal mode, the TMC commanded rate is applied. The other two modes are described below.

The policy of the *queue override* is shown in the left-hand side of the diagram. The goal of this mode is to prevent the on-ramp queue from spilling onto the surface streets. It works by monitoring a *queue detector* placed near the entrance to the ramp (shown in Figure 6.11, page 95). The controller enters the queue override mode whenever the queue detector becomes "hot", while the the speed on the mainline is above 35 mph. The queue detectors on I-210 are considered to be "hot" if they are continuously occupied for more than 3.2 seconds. In VISSIM a detector is "hot" if it reports a *smoothed occupancy rate* of 40% or higher[1]. However, if the average mainline speed is less than 35 mph, the on-ramp will remain in the normal mode, and the on-ramp queue will be allowed to invade the surface streets. This exception to the queue override is meant to avoid the on-ramp queue from being flushed into an already congested freeway. Once in the queue override mode, the TMC commanded rate is discarded. Instead, the metering rate applied during the previous interval is increased by 120 vphpl, without exceeding the maximum metering rate of 900 vphpl.

The second override, termed the *green ball override*, occurs when the local conditions on the mainline fall below an average flow of 1500 vphpl, at an occupancy of 14% or lower. In this situation, on-ramp metering is suspended and the signal is set to green. The green ball condition has a minimum duration of 1 minute, or until either of the thresholds are exceeded. This override is intended to reduce the on-ramp waiting times when the freeway is relatively empty. However, it has the drawback of possibly interfering with system-wide control.

If neither of the two overrides are active, then the per-lane metering rate is computed from the TMC command with:

$$\texttt{rate[k]} \;=\; \frac{\texttt{TMCrate[k]} - \texttt{HOVflow[k-1]}}{\text{\# metered lanes}} \qquad\qquad \text{in [vphpl]}$$

where `HOVflow[k-1]` is the average flow measured by the *HOV bypass detector*, shown in Figure 6.11, during the previous time interval. In the queue override mode, `rate[k]` is found with:

$$\texttt{rate[k]} \;=\; \texttt{rate[k-1]} \,+\, 120 \qquad\qquad \text{in [vphpl]}$$

---

[1]The smoothed occupancy rate is an output of VISSIM's loop detector model.

Both are subject to minimum and maximum bounds:

$$\texttt{rate[k]} \; = \; \max\{ \; \min\{ \; \texttt{rate[k]} \; ; \; 900 \; \} \; ; \; 180 \; \} \qquad\qquad \text{in [vphpl]}$$

The duration of the red phase is then found with:

$$\text{red phase} \; = \; \frac{3600}{\texttt{rate[k]}} \; - \; \text{green phase} \qquad\qquad \text{in [seconds]}$$

where the green phase lasts a fixed 2 seconds. The metering lights on I-210 "rest on red", meaning that they show a red light if there is no vehicle waiting at the stop line (i.e. detected by the *presence detector* in Figure 6.11). The red phase clock advances whether or not there is a vehicle waiting. Hence, a vehicle arriving to the stopline at second 8 of a 10 second red phase will have to wait for only 2 seconds.

## 8.3   Fixed-rate metering

Fixed-rate metering, where the TMC commands to each of the on-ramps is a *constant*, is the simplest form of metering possible. This section investigates the use of fixed-rate metering using different constant rates, and activating or deactivating different portions of the controller logic described in the previous section. These experiments can be used to gain a better understanding of the effects of the two controller overrides and of the limited on-ramp storage space. For example, the longest possible queue lengths and on-ramp waiting times occur when the meters are held at 180 vphpl, and the queue override policy is deactivated. Maximum freeway delay occurs in the 900 vphpl experiment. These experiments also serve to identify ramps that trigger the queue override even when metered at the maximum rate of 900 vphpl. These on-ramps are likely to be dominated by the queue override policy, regardless of the control strategy implemented by the TMC.

Four sets of runs were conducted with fixed-rate metering. The experimental setup is given in Table 8.2. Each set contains five runs, with constant metering rates of 180, 300, 450, 600, and 900 vphpl. These rates correspond to integer-valued cycle durations of 20, 12, 8, 6, and 4 seconds respectively.

In batch FR-1, no overrides were applied, and the on-ramp queues were allowed to grow without limit. Figure 8.2 shows flows from the Vernon St. on-ramp for the 180 vph and 900 vph experiments of FR-1. The solid gray line in both plots is the number of vehicles waiting in the on-ramp queue. In the 180 vph case the queue reaches a peak value of about 500 vehicles, or 3,461 meters in Table 8.3. Peak queue lengths for all of batch FR-1 are provided in Table 8.3. Peak queue lengths that exceed the on-ramp storage capacity (right-most column) have been highlighted in bold type. Notice in the table that four on-ramps (Myrtle, Huntington, San Gabriel, and Lake St.) develop excessive queues even when the metering rate is kept at its maximum value. This suggests that these on-ramps will be affected and possibly even driven by the queue override policy, regardless of the control strategy. Ramps other than these can withstand more restrictive metering rates.

Batches FR-2, FR-3, and FR-4 respectively incorporate the queue override, the green ball override, and the mainline speed exception to the queue override. The peak queue lengths for

## Green Ball Override

B →

ML flow < 1500 vphpl AND ML occ < 14% ? — NO → GBclock > 1min AND GB Override ON ? — NO (to top line)

ML flow < 1500 vphpl AND ML occ < 14% ? — YES → If GB Override OFF then reset and start GBclock → GB Override ON

GBclock > 1min AND GB Override ON ? — YES → GB Override OFF

GB Override ON ? — NO → Implement: green phase = 2 seconds, red phase = $\frac{3600}{\text{rate}[k]}$ - 2 sec.

GB Override ON ? — YES → Set light to green

wait 30 seconds → A

---

## Queue Override

A →

queue detector HOT ? — NO → Queue Override OFF

queue detector HOT ? — YES → Queue Override ON

mainline speed < 35mph ? — NO (continue)

mainline speed < 35mph ? — YES → Queue Override OFF

Queue Override ON ? — NO → rate[k] = $\frac{(\text{TMCrate - HOVflow})}{\text{\# metered lanes}}$

Queue Override ON ? — YES → rate[k] = min{ rate[k-1] + 120, 900 }

→ B

Figure 8.1: I-210 controller logic.

these three sets are shown in Tables 8.4, 8.5, and 8.6. Figures 8.3 through 8.5 show control related variables for the Rosemead NB on-ramp. These figures are intended to give some insight into the effect of each of the overrides. Each figure has 4 plots; the top plot shows the evolution of the on-ramp queue for each of the five fixed metering rates. The bottom three

116

| Batch | Number of trials | Fixed rate | Queue override | Green Ball override | Mainline speed exception |
|---|---|---|---|---|---|
| FR-1 | 5 | 180 - 900 | - | - | - |
| FR-2 | 5 | 180 - 900 | ✓ | - | - |
| FR-3 | 5 | 180 - 900 | ✓ | ✓ | - |
| FR-4 | 5 | 180 - 900 | ✓ | ✓ | ✓ |
| FR-5 | 4 | non-uniform | - | - | - |
| | | | ✓ | - | - |
| | | | ✓ | ✓ | - |
| | | | ✓ | ✓ | ✓ |

Table 8.2: Experiments with fixed-rate control.
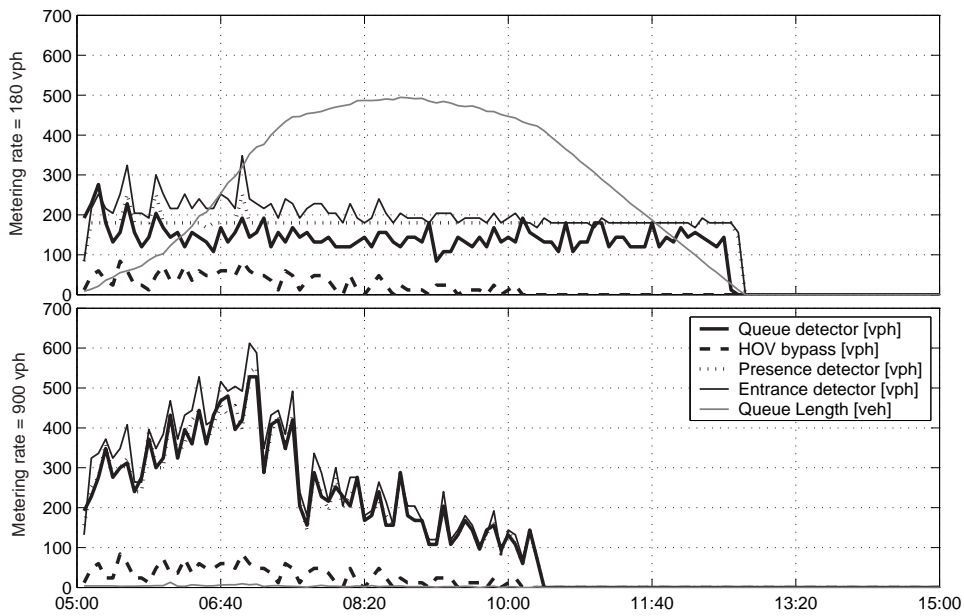


Figure 8.2: On-ramp detector measurements from the Vernon St. on-ramp.

windows show a number between 0 and 1 indicating how often the queue override, green ball override, and mainline speed exception were active. This number was calculated every five minutes for each of the on-ramp controllers as the number of 30 second cycles in which the override was active divided by the total number of cycles in the previous 5 minutes (i.e. 10). Figure 8.3 shows that the queue override was active at Rosemead NB in the 180 vphpl experiment during about 80% of the period between 7:00 am and 9:00 am.

Also evident in Figure 8.3 is the influence of the queue override on the maximum queue length: the peak queue length for the 180 vphpl experiment of 178 m is contrasted with the 8,990 m shown in Table 8.3 when the queue override was deactivated. Figure 8.4 shows the situation when both queue and green ball overrides were enforced, but without the mainline speed exception. It shows that the green ball override is only triggered before and after the peak period (6:30 am to 9:00 am). It therefore acts more or less like an automated

| Street Name | Ramp metering rate [vph] | | | | | Storage [veh] | Max. Queue Length [m] |
|---|---|---|---|---|---|---|---|
| | 180 | 300 | 450 | 600 | 900 | | |
| Vernon | **3,461** | **1,985** | **408** | 107 | 91 | 28 | 190 |
| Irwindale NB | **8,294** | **8,293** | **4,336** | **302** | 88 | 21 | 147 |
| Irwindale SB | **7,263** | **4,833** | **600** | 184 | 67 | 27 | 189 |
| Mount Olive | **480** | 102 | 48 | 30 | 27 | 33 | 231 |
| Buena Vista | **1,324** | **313** | 85 | 44 | 37 | 19 | 133 |
| Mountain | **2,394** | **1,014** | 145 | 56 | 52 | 25 | 175 |
| Myrtle | **7,755** | **7,753** | **5,241** | **1,663** | **203** | 12 | 84 |
| Huntington | **10,788** | **9,401** | **5,634** | **2,454** | **215** | 26 | 182 |
| Santa Anita NB | **3,974** | **82** | 39 | 35 | 22 | 45 | 155 |
| Santa Anita SB | **3,111** | **931** | 107 | 68 | 68 | 21 | 147 |
| Baldwin NB | **3,853** | **1,779** | **243** | 96 | 70 | 14 | 98 |
| Baldwin SB | **3,290** | **79** | 28 | 29 | 27 | 30 | 105 |
| Michillinda | **5,449** | **1,927** | 47 | 34 | 29 | 19 | 65 |
| Rosemead NB | **8,990** | **5,288** | **444** | 55 | 38 | 17 | 60 |
| Rosemead SB | 53 | 41 | 33 | 28 | 28 | 10 | 70 |
| Sierra Madre | **5,114** | **4,281** | **318** | 118 | 74 | 18 | 126 |
| San Gabriel | **10,375** | **9,024** | **5,791** | **2,669** | **207** | 23 | 161 |
| Altadena | **2,719** | **2,719** | **559** | 96 | 69 | 26 | 182 |
| Hill | **14,523** | **14,526** | **7,396** | **1,821** | 61 | 28 | 98 |
| Lake | **19,211** | **16,522** | **6,735** | **772** | **178** | 42 | 145 |

Table 8.3: Peak queue lengths for `FR-1`.

on/off switch for the metering system. Figure 8.5 shows that the mainline speed exception, although sometimes active during the peak period (7:00 am to 9:00 am), does not have an important effect on the frequency of occurrence of the queue override at Rosemead NB. This is because these two rules respond to complementary situations: the queue override is more likely in the 180 vphpl experiment, whereas the mainline speed exception is more likely in the 900 vphpl experiment.

| Street Name | Ramp metering rate [vph] | | | | | Storage [veh] | Max. Queue Length [m] |
|---|---|---|---|---|---|---|---|
| | 180 | 300 | 450 | 600 | 900 | | |
| Vernon | **266** | **267** | 96 | 95 | 91 | 28 | 190 |
| Irwindale NB | **345** | **314** | **259** | **161** | 88 | 21 | 147 |
| Irwindale SB | **280** | **325** | **209** | 158 | 69 | 27 | 189 |
| Mount Olive | 65 | 37 | 33 | 30 | 28 | 33 | 231 |
| Buena Vista | 85 | 59 | 37 | 38 | 34 | 19 | 133 |
| Mountain | **192** | 104 | 50 | 56 | 52 | 25 | 175 |
| Myrtle | **322** | **327** | **173** | **150** | 98 | 12 | 84 |
| Huntington | **690** | **450** | **360** | **431** | 134 | 26 | 182 |
| Santa Anita NB | 148 | 83 | 39 | 30 | 21 | 45 | 155 |
| Santa Anita SB | **205** | **188** | 88 | 68 | 68 | 21 | 147 |
| Baldwin NB | **174** | **174** | **124** | 83 | 70 | 14 | 98 |
| Baldwin SB | **156** | 79 | 28 | 29 | 27 | 30 | 105 |
| Michillinda | **200** | **166** | 47 | 34 | 29 | 19 | 65 |
| Rosemead NB | **178** | **123** | 91 | 55 | 38 | 17 | 60 |
| Rosemead SB | 53 | 41 | 33 | 28 | 28 | 10 | 70 |
| Sierra Madre | **295** | **281** | **150** | 91 | 74 | 18 | 126 |
| San Gabriel | **960** | **618** | **509** | **350** | **185** | 23 | 161 |
| Altadena | **302** | **297** | **242** | 91 | 69 | 26 | 182 |
| Hill | **523** | **484** | **239** | **231** | 56 | 28 | 98 |
| Lake | **788** | **460** | **329** | **314** | **179** | 42 | 145 |

Table 8.4: Peak queue lengths for `FR-2`.
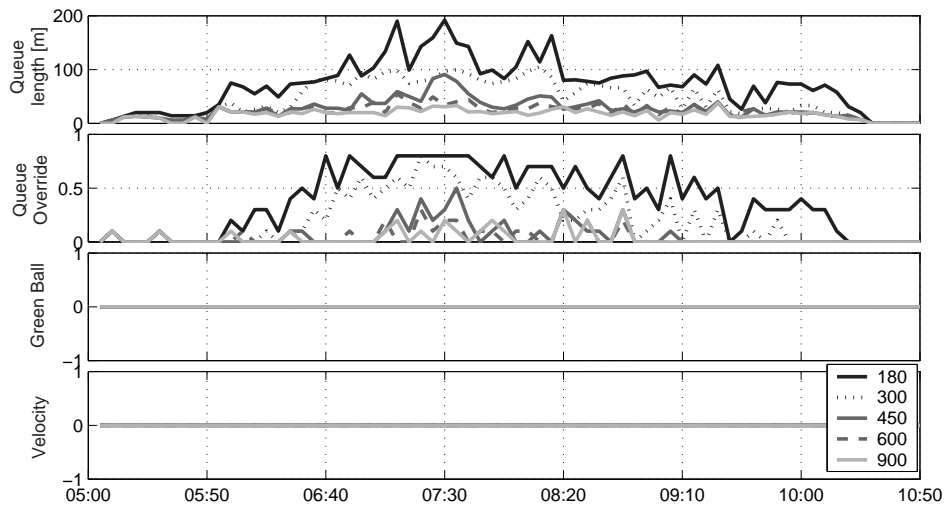


Figure 8.3: Queue length and overrides for `FR-2` (Rosemead NB on-ramp).

119

| Street Name | Ramp metering rate [vph] | | | | | Storage [veh] | Max. Queue Length [m] |
|---|---|---|---|---|---|---|---|
| | 180 | 300 | 450 | 600 | 900 | | |
| Vernon | **266** | **267** | 96 | 95 | 91 | 28 | 190 |
| Irwindale NB | **637** | **626** | **298** | **161** | 88 | 21 | 147 |
| Irwindale SB | **1,502** | **1,186** | **484** | 184 | 67 | 27 | 189 |
| Mount Olive | 54 | 63 | 33 | 30 | 28 | 33 | 231 |
| Buena Vista | 90 | 72 | 68 | 36 | 34 | 19 | 133 |
| Mountain | **327** | **256** | 56 | 56 | 52 | 25 | 175 |
| Myrtle | **1,691** | **1,155** | **395** | **204** | 81 | 12 | 84 |
| Huntington | **706** | **843** | **571** | **339** | 147 | 26 | 182 |
| Santa Anita NB | **453** | 82 | 38 | 35 | 22 | 45 | 155 |
| Santa Anita SB | **272** | **226** | 87 | 68 | 68 | 21 | 147 |
| Baldwin NB | **513** | **512** | **198** | 96 | 70 | 14 | 98 |
| Baldwin SB | **350** | 79 | 28 | 29 | 27 | 30 | 105 |
| Michillinda | **1,109** | **240** | 47 | 34 | 29 | 19 | 65 |
| Rosemead NB | **439** | **145** | **86** | 55 | 38 | 17 | 60 |
| Rosemead SB | 53 | 41 | 33 | 29 | 28 | 10 | 70 |
| Sierra Madre | **279** | **235** | **204** | 91 | 74 | 18 | 126 |
| San Gabriel | **2,889** | **4,508** | **1,758** | **609** | **178** | 23 | 161 |
| Altadena | **451** | **647** | **256** | 96 | 69 | 26 | 182 |
| Hill | **3,020** | **2,638** | **476** | **476** | 59 | 28 | 98 |
| Lake | **983** | **925** | **333** | **318** | **284** | 42 | 145 |

Table 8.5: Peak queue lengths for `FR-3`.
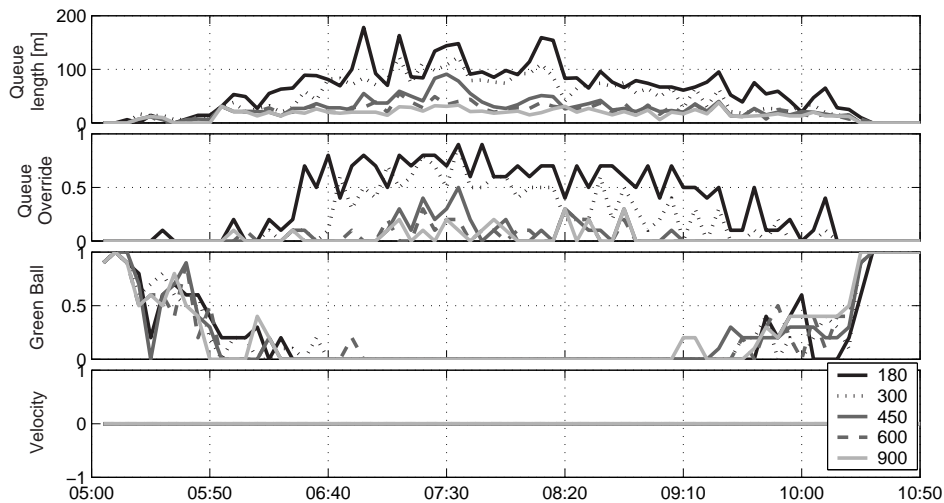


Figure 8.4: Queue length and overrides for `FR-3` (Rosemead NB on-ramp).

120

| Street Name | Ramp metering rate [vph] | | | | | Storage [veh] | Max. Queue Length [m] |
|---|---|---|---|---|---|---|---|
| | 180 | 300 | 450 | 600 | 900 | | |
| Vernon | **309** | **307** | **277** | 108 | 91 | 28 | 190 |
| Irwindale NB | **650** | **484** | **343** | **178** | 88 | 21 | 147 |
| Irwindale SB | **377** | **392** | **269** | 174 | 69 | 27 | 189 |
| Mount Olive | **243** | 102 | 48 | 30 | 27 | 33 | 231 |
| Buena Vista | **264** | **173** | 85 | 44 | 37 | 19 | 133 |
| Mountain | **241** | **216** | 147 | 56 | 52 | 25 | 175 |
| Myrtle | **617** | **393** | **338** | **281** | **203** | 12 | 84 |
| Huntington | **2,223** | **1,488** | **953** | **568** | **191** | 26 | 182 |
| Santa Anita NB | **218** | 82 | 39 | 35 | 22 | 45 | 155 |
| Santa Anita SB | **219** | **232** | 117 | 68 | 68 | 21 | 147 |
| Baldwin NB | **203** | **199** | **124** | 83 | 70 | 14 | 98 |
| Baldwin SB | **148** | **79** | 28 | 29 | 27 | 30 | 105 |
| Michillinda | **200** | **177** | 47 | 34 | 29 | 19 | 65 |
| Rosemead NB | **192** | **105** | **91** | 55 | 38 | 17 | 60 |
| Rosemead SB | 53 | 41 | 33 | 28 | 28 | 10 | 70 |
| Sierra Madre | **342** | **319** | **204** | 118 | 74 | 18 | 126 |
| San Gabriel | **2,092** | **1,378** | **1,161** | **470** | **198** | 23 | 161 |
| Altadena | **301** | **317** | **242** | 91 | 69 | 26 | 182 |
| Hill | **519** | **443** | **344** | **255** | 59 | 28 | 98 |
| Lake | **811** | **522** | **342** | **297** | **167** | 42 | 145 |

Table 8.6: Peak queue lengths for `FR-4`.
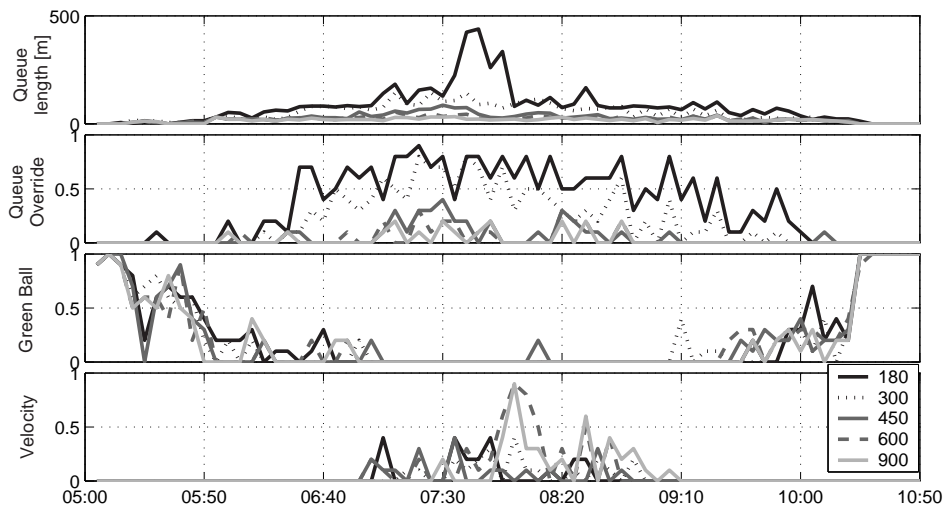


Figure 8.5: Queue length and overrides for `FR-4` (Rosemead NB on-ramp).

121

Performance measures for all of the fixed-rate tests are provided in Table 8.7. The values in the table are percent changes with respect to the uncontrolled freeway. Thus, improvements over no control are indicated by negative values in the %TPH and %TVH columns and positive values in %AMS and %TVH.

Batch FR-1 exhibits some expected tendencies: the %OPH and %AMS decreased and %ATh increased with increasing metering rate. The travel time measures (%TPH and %TVH) on the other hand, did not behave monotonically, but instead reached a minimum value in the 600 vphpl experiment. In fact, 600 vphpl was the only level of on-ramp flow that performed favorably as compared to no control. The reductions in travel time of -4.4% TPH and -4.3% TVH deteriorated quickly to 25.7% and 29.1% increases at 450 vphpl. This suggests that selecting a single fixed metering rate for a large and complex site such as I-210 is a delicate task, and probably not a wise approach.

Similar patterns in terms of %OPH and %AMS were observed when the queue override was added in batch FR-2. ATh remained more or less constant for all metering rates (also in batches FR-3 and FR-4) because the queue override ensured that relatively few vehicles remained in the on-ramp queues at 10:30 am. The %TPH and %TVH are less sensitive to the fixed metering rate in batches FR-2, FR-3, FR-4 than in FR-1. Thus, adding the queue override helped to avoid the huge travel time losses observed in batch FR-1 at low metering rates (e.g. 327% TPH in FR-1 became 4.1% in FR-4), but also reduced the potential benefits (-4.4% TPH became -1.5% at 600 vphpl).

| Batch | metering rate [vph] | %OPH | %FPH | %TPH | %TVH | %TPK | %TVK | %AMS | %ATh |
|-------|---------------------|------|------|------|------|------|------|------|------|
|       | 180 | 2,019.9 | -35.4 | 327.0 | 358.5 | -2.7 | -3.1 | 28.6 | -15.9 |
|       | 300 | 967.2 | -30.5 | 145.4 | 160.1 | -0.3 | -0.5 | 24.4 | -8.4 |
| FR-1  | 450 | 256.4 | -23.6 | 25.7 | 29.1 | 0.2 | 0.3 | 18.4 | -2.8 |
|       | 600 | 41.8 | -14.2 | **-4.4** | **-4.3** | 0.1 | 0.0 | 10.2 | -0.2 |
|       | 900 | 3.2 | 4.4 | 4.2 | 4.3 | 0.4 | 0.3 | -2.2 | -0.2 |
|       | 180 | 18.8 | -2.9 | 0.9 | 0.9 | 0.2 | 0.1 | 2.1 | 0.0 |
|       | 300 | 11.8 | -0.8 | 1.5 | 1.5 | 0.1 | 0.1 | 0.4 | 0.1 |
| FR-2  | 450 | 5.1 | 0.5 | 1.3 | 1.4 | 0.1 | 0.1 | -0.3 | 0.1 |
|       | 600 | 5.8 | -2.6 | **-1.1** | **-1.3** | 0.3 | 0.2 | 1.8 | 0.2 |
|       | 900 | -1.6 | -3.4 | **-3.0** | **-3.2** | 0.0 | 0.0 | 2.4 | 0.0 |
|       | 180 | 40.0 | -8.7 | **-0.1** | **-0.1** | 0.2 | 0.2 | 6.0 | 0.4 |
|       | 300 | 33.2 | -5.8 | 1.1 | 1.2 | 0.3 | 0.2 | 3.7 | 0.4 |
| FR-3  | 450 | 10.4 | -7.5 | **-4.4** | **-4.6** | 0.0 | 0.0 | 5.5 | 0.2 |
|       | 600 | 4.4 | 0.4 | 1.1 | 1.0 | 0.3 | 0.2 | 0.3 | 0.0 |
|       | 900 | 0.6 | -0.2 | **-0.1** | **-0.1** | 0.3 | 0.3 | 0.0 | 0.4 |
|       | 180 | 50.7 | -5.8 | 4.1 | 4.4 | 0.2 | 0.2 | 3.9 | 0.2 |
|       | 300 | 30.7 | -5.6 | 0.8 | 1.0 | 0.1 | 0.1 | 3.6 | 0.4 |
| FR-4  | 450 | 16.3 | -1.6 | 1.6 | 1.6 | 0.3 | 0.3 | 1.1 | 0.5 |
|       | 600 | 6.8 | -3.3 | **-1.5** | **-1.7** | 0.2 | 0.2 | 2.2 | 0.3 |
|       | 900 | 1.5 | -1.9 | **-1.3** | **-1.4** | 0.2 | 0.2 | 1.3 | 0.3 |

Table 8.7: Performance measures with fixed-rate metering.

## 8.4 Alinea control

The Alinea control law was introduced in Section 4.4. Its control law is expressed in the notation of this chapter as:

$$\texttt{TMCrate[k]} = \texttt{ENTflw[k-1]} + K_r \left( \hat{o} - \texttt{occ[k]} \right) \qquad (8.3)$$

The two tunable parameters in this control law are the target occupancy ($\hat{o}$) and a positive gain ($K_r$). The desired metering rate for the upcoming time interval, $\texttt{TMCrate[k]}$, is computed as the measured flow entering the freeway during the previous interval, $\texttt{ENTflw[k-1]}$, adjusted by an amount proportional to the difference between the measured mainline occupancy and the target mainline occupancy ($\hat{o} - \texttt{occ[k]}$). Thus, if the mainline occupancy is less than the desired value, Alinea will increase the number of vehicles being released onto the freeway. The quickness with which Alinea reacts is determined by $K_r$.
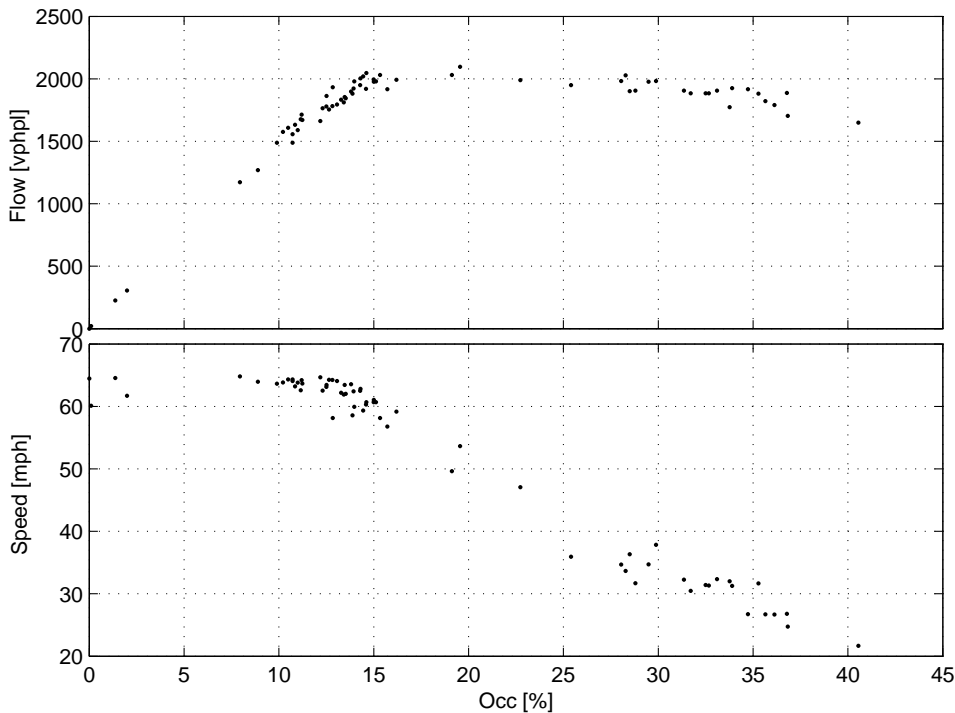


Figure 8.6: Simulated scatter plot for the Rosemead NB mainline detector station.

It is suggested in [77] that the target occupancy be set a little lower than the estimated *critical occupancy* at which the freeway transitions from free-flow into congestion. Scatter plots from the VISSIM model such as the one shown in Figure 8.6 suggest that its critical occupancy is somewhere between 15% and 20%. The recommended value for the gain is 7,000. Alinea was originally conceived for freeways with mainline loop detectors placed *downstream* of the on-ramp.

The goals of the experiments with Alinea presented here are 1) to understand the influence of $\hat{o}$ and $K_r$ on performance, 2) to determine a best choice of controller parameters, 3) to see whether Alinea can be successfully applied to a freeway such as I-210, equipped with

upstream rather than downstream detectors, and 4) to study the effects of the controller overrides on Alinea's performance.

Table 8.8 provides the experimental setup. The first three batches are a study of Alinea without overrides, using upstream detectors. `AL-1` tests 13 different values of $\hat{o}$ between 8% and 40%, using the recommended value of $K_r = 7,000$. The gain is varied in `AL-2` and `AL-3` using a good and a bad choice of $\hat{o}$ from `AL-1`. The study is repeated with detectors placed downstream of the on-ramps in `AL-4` through `AL-6`. Finally, `AL-7`, `AL-8`, and `AL-9` investigate the effects of the overrides.

| Batch | Number of trials | $K_r$ | $\hat{o}$ | Detector location | Queue override | Green Ball override | Speed exception |
|-------|------------------|-------|-----------|-------------------|----------------|---------------------|-----------------|
| AL-1 | 13 | 7,000 | 8.0% - 40.0% | UP | - | - | - |
| AL-2 | 5 | 70 - 20,000 | 14.4% | UP | - | - | - |
| AL-3 | 5 | 70 - 20,000 | 27.2% | UP | - | - | - |
| AL-4 | 13 | 7,000 | 8.0% - 40.0% | DOWN | - | - | - |
| AL-5 | 5 | 70 - 20,000 | 14.4% | DOWN | - | - | - |
| AL-6 | 5 | 70 - 20,000 | 27.2% | DOWN | - | - | - |
| AL-7 | 13 | 7,000 | 8.0% - 40.0% | UP | ✓ | - | - |
| AL-8 | 13 | 7,000 | 8.0% - 40.0% | UP | ✓ | ✓ | - |
| AL-9 | 13 | 7,000 | 8.0% - 40.0% | UP | ✓ | ✓ | ✓ |

Table 8.8: Experiments with Alinea.
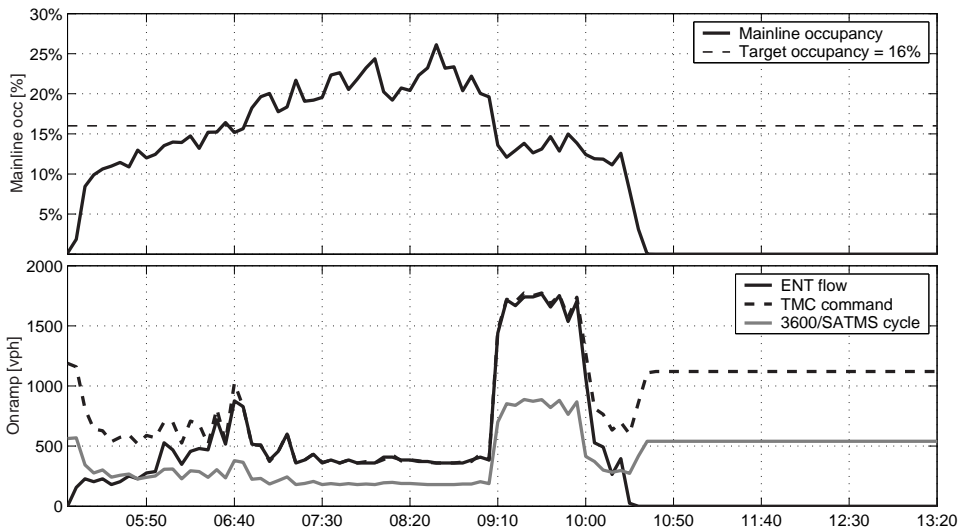


Figure 8.7: Control variables for Rosemead NB (`AL-1`).

First, however, Figure 8.7 illustrates how Alinea works. The top window in the figure shows the average mainline occupancy (`occ[k]`). Alinea reacts to "congestion", i.e. `occ[k]` > $\hat{o}$, by reducing the metering rate. In the figure, congestion starts at around 6:40 am. At this point, Alinea starts to reduce its commanded rate, which quickly reaches its minimum value of 360 vph (= 180 vphpl $\times$ 2 metered lanes). The actual flow onto the freeway

closely follows the commanded rate because there is a queue of vehicles waiting on the ramp. After congestion dissipates at around 9:00 am, and mainline occupancy decreases below the threshold, Alinea responds by increasing the on-ramp flow to its maximum value of 1,800 vph, until 10:30 am when simulated demand stops. After the freeway has emptied, the TMC commanded rate settles at 1,120 vph ($= K_r \hat{o}$), which is the value obtained by replacing `ENTflw[k-1] = occ[k] = 0` in Eq. (8.3).
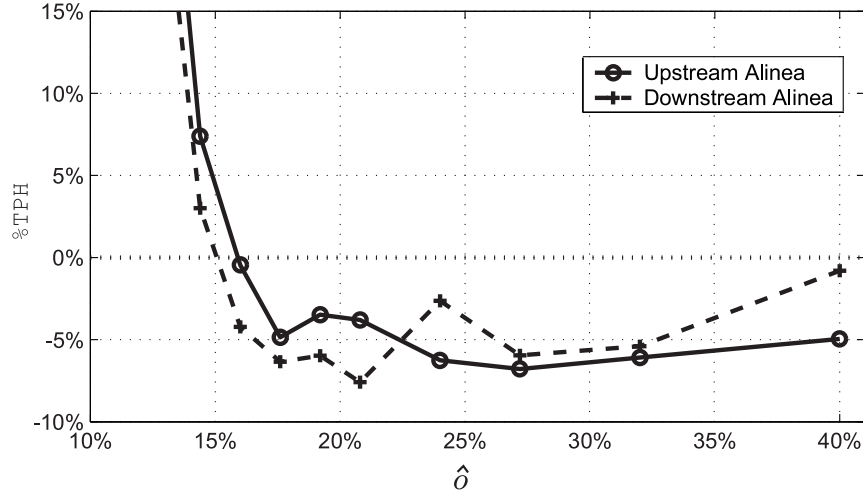


Figure 8.8: Upstream vs. Downstream Alinea (`AL-1` vs. `AL-3`).

Tables 8.9 and 8.10 at the end of this section provide performance measures for all of the experiments with Alinea. Figure 8.8 compares the system performance when the mainline detectors are placed either immediately upstream or downstream of the on-ramp junction. Recall that the loop detectors on I-210 are located upstream of the on-ramps, and that the recommended placement according to [77] is downstream. The reasoning behind this recommendation is that, in the free-flow state, which is the state that Alinea attempts to preserve, changes in on-ramp flow are reflected in the downstream and not in the upstream measurement. Therefore, in the free-flow state, feedback control is only possible with a downstream measurement. However, one of the conclusions of Chapter 4 was that the opposite is true when the freeway is congested: observability is achieved with the upstream measurement. Figure 8.8 confirms this observation. Better performance was obtained with downstream detectors at low target occupancies, which lead to shorter congestion periods, whereas the upstream detectors performed better at high (i.e. super-critical) occupancies. Also, the best case performance was achieved with downstream detectors and a near critical target occupancy, as originally claimed by [77].

Figure 8.9 compares the performance of several variations of Alinea using upstream detectors. The four lines represent the four levels of override implementation (`AL-1`, `AL-7`, `AL-8`, `AL-9`), with target occupancy as the independent variable. Best case performance without overrides and using upstream detectors, is achieved at relatively high target occupancy values of 24% to 32%. This is also true when the queue override and later the green ball override are enforced in `AL-7` and `AL-8`. As with fixed-rate control, the pattern becomes more irregular when all of the overrides are activated.
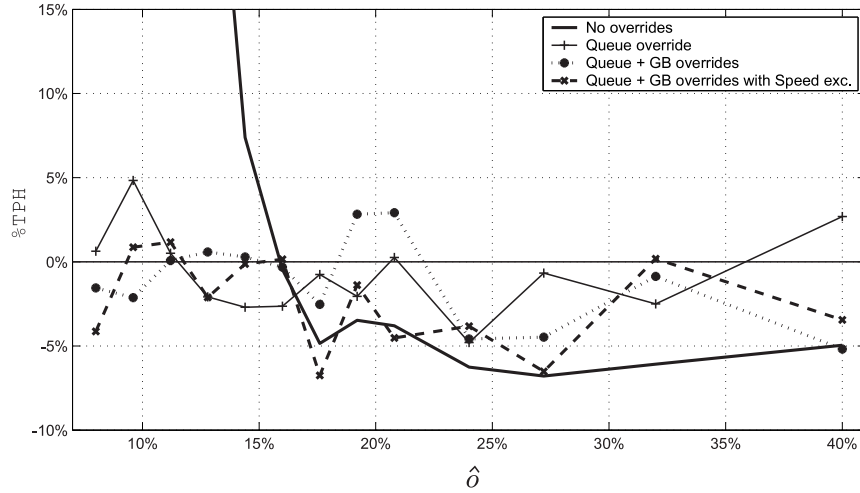
Figure 8.9: Performance with upstream Alinea (`AL-1`, `AL-7`, `AL-8`, `AL-9`).
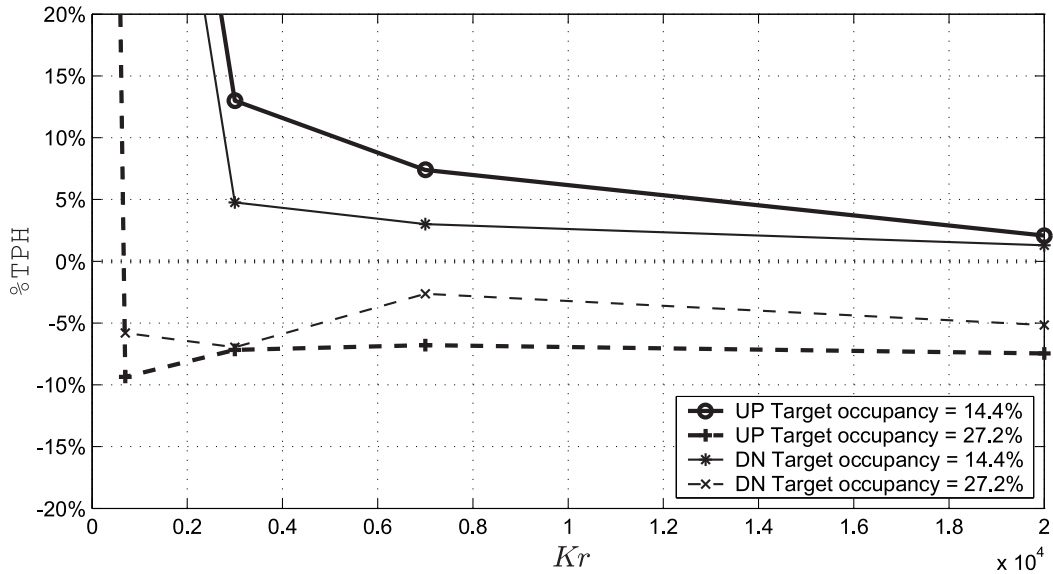


Figure 8.10: Alinea - Gain variations (`AL-2`, `AL-4`).

Batches `AL-2` and `AL-4` test the sensitivity of Alinea to changes in the value of $K_r$, for $\hat{o} = 14.4\%$ and $27.2\%$, and using either the upstream or downstream detectors. $\hat{o} = 27.2\%$ was the best choice in `AL-1`, when upstream detectors were used, and a near-best choice in `AL-3` with downstream detectors. $\hat{o} = 14.4\%$ was a bad choice in both cases. The results are illustrated in Figure 8.10. With $\hat{o} = 14.4\%$, the controller never performs better than in the no control case, for any values of the gain, although the performance with the downstream detectors is consistently better than with the upstream detectors. For $\hat{o} = 27.2\%$, the best performance was recorded with relatively small values of the gain: $K_r = 3,000$ with downstream detectors and $K_r = 700$ with upstream detectors. The figure also reconfirms the observation from Figure 8.8, that upstream detectors perform better at

high target occupancies (27.2%), and downstream perform better at low target occupancies (14.4%). Another interesting observation is that, for both super-critical and sub-critical target occupancies and for upstream and downstream detectors, $K_r = 20,000$ was a near-optimal choice. Although not tested, this suggests that a "bang-bang" controller, might be an effective strategy.

| Batch | $\hat{o}$ | $K_r$ | %OPH | %FPH | %TPH | %TVH | %TPK | %TVK | %AMS | %ATh |
|---|---|---|---|---|---|---|---|---|---|---|
|  | 8.0 |  | 1170.0 | -34.6 | 177.8 | 195.0 | -1.8 | -2.1 | 28.0 | -13.3 |
|  | 9.6 |  | 893.6 | -33.5 | 130.0 | 143.0 | -0.7 | -0.8 | 27.0 | -8.6 |
|  | 11.2 |  | 569.7 | -31.9 | 74.1 | 82.1 | -0.3 | -0.4 | 25.3 | -4.0 |
|  | 12.8 |  | 318.2 | -28.4 | 32.7 | 36.7 | 0.1 | 0.1 | 21.8 | -1.6 |
|  | 14.4 | ↑ | 163.8 | -26.1 | 7.4 | 8.9 | 0.3 | 0.3 | 19.6 | 0.1 |
|  | 16.0 |  | 118.8 | -26.0 | **-0.5** | 0.3 | 0.0 | 0.0 | 18.9 | 0.0 |
| AL-1 | 17.6 | 7,000 | 94.8 | -26.2 | **-4.9** | **-4.5** | 0.2 | 0.1 | 19.2 | 0.1 |
|  | 19.2 |  | 90.7 | -23.6 | **-3.5** | **-3.0** | 0.3 | 0.3 | 17.0 | 0.3 |
|  | 20.8 | ↓ | 86.4 | -23.1 | **-3.8** | **-3.4** | 0.2 | 0.2 | 16.3 | 0.2 |
|  | 24.0 |  | 70.0 | -22.6 | **-6.3** | **-6.0** | 0.1 | 0.1 | 15.8 | 0.0 |
|  | 27.2 |  | 60.3 | -21.2 | **-6.8** | **-6.7** | 0.4 | 0.3 | 14.5 | 0.4 |
|  | 32.0 |  | 53.5 | -18.8 | **-6.1** | **-5.9** | 0.3 | 0.2 | 12.7 | 0.3 |
|  | 40.0 |  | 32.7 | -13.0 | **-5.0** | **-4.9** | 0.3 | 0.2 | 8.7 | 0.2 |
|  | ↑ | 70 | 1102.0 | -30.1 | 169.5 | 185.7 | -0.6 | -0.7 | 24.2 | -8.3 |
|  |  | 700 | 521.7 | -28.8 | 68.2 | 75.5 | -0.5 | -0.6 | 22.4 | -2.7 |
| AL-2 | 14.4 | 3,000 | 199.7 | -27.0 | 13.0 | 15.1 | 0.0 | 0.0 | 20.2 | -0.6 |
|  |  | 7,000 | 163.8 | -26.1 | 7.4 | 8.9 | 0.3 | 0.3 | 19.6 | 0.1 |
|  | ↓ | 20,000 | 136.4 | -26.7 | 2.1 | 3.1 | 0.2 | 0.2 | 19.9 | 0.3 |
|  | ↑ | 70 | 1089.2 | -29.7 | 167.6 | 183.6 | -0.5 | -0.7 | 23.9 | -8.1 |
|  |  | 700 | 60.8 | -24.4 | **-9.4** | **-9.4** | 0.1 | 0.0 | 17.2 | -0.2 |
| AL-3 | 27.2 | 3,000 | 60.0 | -21.5 | **-7.2** | **-7.0** | 0.1 | 0.1 | 14.7 | 0.1 |
|  |  | 7,000 | 60.3 | -21.2 | **-6.8** | **-6.7** | 0.4 | 0.3 | 14.5 | 0.4 |
|  | ↓ | 20,000 | 55.1 | -20.9 | **-7.5** | **-7.4** | 0.3 | 0.2 | 14.3 | 0.1 |
|  | 8.0 |  | 1109.6 | -34.7 | 167.1 | 183.7 | -1.5 | -1.8 | 28.2 | -15.2 |
|  | 9.6 |  | 789.9 | -32.9 | 112.1 | 124.0 | -0.2 | -0.3 | 26.7 | -10.3 |
|  | 11.2 |  | 505.3 | -31.4 | 63.2 | 70.3 | 0.4 | 0.3 | 25.0 | -6.5 |
|  | 12.8 |  | 276.6 | -28.6 | 25.2 | 28.5 | 0.2 | 0.2 | 22.0 | -3.2 |
|  | 14.4 | ↑ | 141.5 | -26.6 | 3.0 | 4.2 | 0.1 | 0.1 | 19.9 | -2.4 |
|  | 16.0 |  | 101.3 | -26.8 | **-4.2** | **-3.7** | 0.3 | 0.3 | 19.9 | -1.9 |
| AL-4 | 17.6 | 7,000 | 82.6 | -25.4 | **-6.3** | **-6.1** | 0.1 | 0.1 | 18.4 | -2.1 |
|  | 19.2 |  | 77.1 | -23.8 | **-6.0** | **-5.8** | 0.4 | 0.3 | 17.1 | -2.1 |
|  | 20.8 | ↓ | 73.2 | -24.9 | **-7.6** | **-7.4** | 0.3 | 0.3 | 17.8 | -1.9 |
|  | 24.0 |  | 75.0 | -19.2 | **-2.6** | **-2.1** | 0.2 | 0.2 | 13.0 | -2.0 |
|  | 27.2 |  | 60.2 | -20.1 | **-5.9** | **-5.7** | 0.2 | 0.2 | 13.7 | -2.1 |
|  | 32.0 |  | 46.5 | -16.5 | **-5.4** | **-5.2** | 0.3 | 0.3 | 10.9 | -1.9 |
|  | 40.0 |  | 12.2 | -3.6 | **-0.8** | **-0.6** | 0.2 | 0.2 | 2.0 | -1.6 |
|  | ↑ | 70 | 1091.4 | -30.7 | 167.1 | 183.3 | -0.7 | -0.8 | 24.6 | -11.0 |
|  |  | 700 | 483.8 | -30.3 | 60.3 | 66.9 | -0.4 | -0.5 | 23.7 | -4.8 |
| AL-5 | 14.4 | 3,000 | 160.8 | -28.6 | 4.8 | 6.2 | 0.2 | 0.2 | 21.5 | -2.3 |
|  |  | 7,000 | 141.5 | -26.6 | 3.0 | 4.2 | 0.1 | 0.1 | 19.9 | -2.4 |
|  | ↓ | 20,000 | 129.7 | -26.2 | 1.3 | 2.3 | 0.3 | 0.2 | 19.5 | -2.3 |
|  | ↑ | 70 | 1091.4 | -30.7 | 167.1 | 183.3 | -0.7 | -0.8 | 24.6 | -11.0 |
|  |  | 700 | 79.5 | -24.1 | **-5.8** | **-5.5** | 0.1 | 0.1 | 17.0 | -2.2 |
| AL-6 | 27.2 | 3,000 | 65.7 | -22.5 | **-7.0** | **-6.9** | 0.2 | 0.1 | 15.9 | -2.2 |
|  |  | 7,000 | 75.0 | -19.2 | **-2.6** | **-2.1** | 0.2 | 0.2 | 13.0 | -2.0 |
|  | ↓ | 20,000 | 67.4 | -20.7 | **-5.2** | **-4.9** | 0.4 | 0.4 | 14.3 | -2.1 |

Table 8.9: Performance with Alinea.

| Batch | $\hat{o}$ | $K_r$ | %OPH | %FPH | %TPH | %TVH | %TPK | %TVK | %AMS | %ATh |
|---|---|---|---|---|---|---|---|---|---|---|
| | 8.0 | | 36.5 | -7.0 | 0.6 | 0.7 | -0.1 | -0.1 | 4.8 | -0.1 |
| | 9.6 | | 34.2 | -1.5 | 4.8 | 5.0 | 0.3 | 0.3 | 1.0 | 0.5 |
| | 11.2 | | 26.2 | -5.0 | 0.5 | 0.5 | 0.1 | 0.0 | 3.6 | 0.0 |
| | 12.8 | | 22.5 | -7.4 | **-2.1** | **-2.2** | 0.0 | 0.0 | 4.9 | 0.1 |
| | 14.4 | ↑ | 19.2 | -7.4 | **-2.7** | **-2.8** | 0.2 | 0.2 | 5.0 | 0.4 |
| | 16.0 | | 14.3 | -6.3 | **-2.6** | **-2.8** | 0.1 | 0.1 | 4.2 | 0.2 |
| AL-7 | 17.6 | 7,000 | 16.3 | -4.4 | **-0.8** | **-0.8** | 0.1 | 0.1 | 2.8 | 0.2 |
| | 19.2 | | 15.5 | -5.8 | **-2.1** | **-2.1** | 0.3 | 0.3 | 4.0 | 0.7 |
| | 20.8 | ↓ | 16.7 | -3.3 | 0.3 | 0.3 | 0.2 | 0.2 | 1.7 | 0.5 |
| | 24.0 | | 10.9 | -8.2 | **-4.8** | **-5.0** | 0.1 | 0.1 | 5.5 | 0.3 |
| | 27.2 | | 10.6 | -3.1 | **-0.7** | **-0.7** | 0.3 | 0.2 | 2.0 | 0.3 |
| | 32.0 | | 6.2 | -4.4 | **-2.5** | **-2.6** | 0.3 | 0.2 | 2.9 | 0.5 |
| | 40.0 | | 7.2 | 1.7 | 2.7 | 2.8 | 0.3 | 0.3 | -1.0 | 0.2 |
| | 8.0 | | 14.2 | -4.9 | **-1.6** | **-1.7** | 0.2 | 0.1 | 3.2 | 0.3 |
| | 9.6 | | 15.7 | -5.9 | **-2.1** | **-2.3** | 0.1 | 0.0 | 3.9 | 0.1 |
| | 11.2 | | 13.2 | -2.7 | 0.1 | 0.0 | 0.0 | 0.0 | 2.2 | 0.0 |
| | 12.8 | | 13.1 | -2.1 | 0.6 | 0.6 | -0.1 | -0.1 | 1.5 | -0.1 |
| | 14.4 | ↑ | 12.3 | -2.3 | 0.3 | 0.2 | 0.1 | 0.1 | 1.9 | 0.1 |
| | 16.0 | | 12.1 | -3.0 | **-0.3** | **-0.4** | 0.2 | 0.2 | 2.0 | 0.5 |
| AL-8 | 17.6 | 7,000 | 6.3 | -4.4 | **-2.5** | **-2.7** | 0.2 | 0.1 | 3.2 | 0.2 |
| | 19.2 | | 10.3 | 1.2 | 2.8 | 2.9 | 0.3 | 0.2 | -0.2 | 0.0 |
| | 20.8 | ↓ | 11.7 | 1.0 | 2.9 | 3.0 | 0.3 | 0.2 | -0.1 | 0.0 |
| | 24.0 | | 5.3 | -6.7 | **-4.6** | **-4.9** | 0.1 | 0.0 | 4.6 | 0.1 |
| | 27.2 | | 2.6 | -6.0 | **-4.5** | **-4.7** | 0.1 | 0.1 | 4.4 | 0.2 |
| | 32.0 | | 6.5 | -2.4 | **-0.9** | **-1.0** | 0.1 | 0.1 | 1.7 | 0.1 |
| | 40.0 | | -2.0 | -5.9 | **-5.2** | **-5.4** | -0.1 | -0.1 | 3.8 | -0.1 |
| | 8.0 | | 33.8 | -12.3 | **-4.1** | **-4.3** | 0.1 | 0.1 | 8.3 | 0.3 |
| | 9.6 | | 42.7 | -8.1 | 0.9 | 1.0 | 0.4 | 0.3 | 5.6 | 0.6 |
| | 11.2 | | 45.7 | -8.4 | 1.2 | 1.4 | 0.2 | 0.2 | 5.5 | 0.3 |
| | 12.8 | | 37.0 | -10.5 | **-2.1** | **-2.1** | 0.1 | 0.1 | 7.0 | 0.4 |
| | 14.4 | ↑ | 42.7 | -9.3 | **-0.1** | **-0.1** | 0.1 | 0.1 | 6.3 | 0.1 |
| | 16.0 | | 32.0 | -6.7 | 0.1 | 0.2 | 0.3 | 0.2 | 4.5 | 0.5 |
| AL-9 | 17.6 | 7,000 | 21.0 | -12.7 | **-6.8** | **-7.0** | 0.0 | 0.0 | 8.7 | 0.1 |
| | 19.2 | | 30.3 | -8.2 | **-1.4** | **-1.5** | 0.2 | 0.1 | 5.8 | 0.3 |
| | 20.8 | ↓ | 23.1 | -10.4 | **-4.5** | **-4.7** | 0.2 | 0.2 | 7.3 | 0.3 |
| | 24.0 | | 23.7 | -9.7 | **-3.8** | **-4.0** | 0.3 | 0.3 | 6.7 | 0.6 |
| | 27.2 | | 14.3 | -11.0 | **-6.5** | **-6.8** | 0.1 | 0.1 | 7.7 | 0.3 |
| | 32.0 | | 16.3 | -3.3 | 0.2 | 0.2 | 0.1 | 0.1 | 1.8 | 0.2 |
| | 40.0 | | -1.6 | -3.8 | **-3.4** | **-3.6** | 0.1 | 0.1 | 2.4 | 0.1 |

Table 8.10: Performance with Alinea, continued.

## 8.5 Percent occupancy control

The Percent Occupancy (%-Occ) control law was introduced in Section 4.3. It is expressed in the notation of this chapter as,

$$\texttt{TMCrate}[\texttt{k}] = 900 - \frac{900 - 180}{o^h - o^l}\left(\texttt{occ}[\texttt{k}] - o^l\right) \tag{8.4}$$

Under %-Occ control, the commanded rate is a decreasing linear function of the mainline occupancy, as plotted in Figure 8.11. The two parameters in the %-Occ control law are the low and high occupancy thresholds, $o^l$ and $o^h$, at which the commanded rate is assigned the maximum and the minimum rate respectively. For values below $o^l$, the value computed by Eq. (8.4) exceeds 900 vphpl, but it is reduced to 900 vphpl by the control boxes. Similarly, the implemented rate is never less than 180 vphpl. The author knows of no published guidelines for tuning $o^l$ and $o^h$.
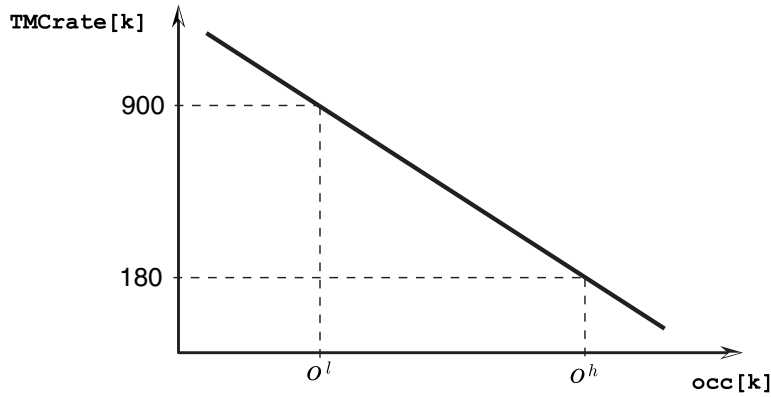


Figure 8.11: The %-Occ control function.

Three sets of simulation experiments were conducted, as shown in Table 8.11. Each set consists of 26 parameter selections with $o^l$ ranging from 10% to 30%, and $o^h$ from 18% to 30%. The resulting changes in `TPH` are shown in Table 8.11 and Figure 8.12.

| Batch | Number of trials | occ_low | occ_high | Detector location | Queue override | Green Ball override | Speed exception |
|-------|------------------|---------|----------|-------------------|----------------|---------------------|-----------------|
| PO-1  | 26 | 10% - 30% | 18% - 30% | UP | - | - | - |
| PO-2  | 26 | 10% - 30% | 18% - 30% | UP | ✓ | - | - |
| PO-3  | 26 | 10% - 30% | 18% - 30% | UP | ✓ | ✓ | ✓ |

Table 8.11: Experiments with %-Occ.

The plots in Figure 8.12 are a bit erratic and difficult to interpret. Even in `PO-1`, when no overrides were enforced, there is no recognizable trend and no obvious best choice. All

131

but one of the tests in this batch reduced travel time. A maximum reduction of -7.4% was recorded with $o^l = 14\%$ and $o^h = 30\%$. One consistent trend in PO-1 is in the performance of "bang-bang" controllers, whose performance worsened with increasing $o^l = o^h$. As with Alinea, enforcing the overrides in PO-2 and PO-3 has the general effect of diminishing travel time reductions. In general, no concrete recommendation for tuning the %-Occ controller can be extracted from the results.

| $o^l$ \ $o^h$ | PO-1 | | | | PO-2 | | | | PO-3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 18% | 22% | 26% | 30% | 18% | 22% | 26% | 30% | 18% | 22% | 26% | 30% |
| 10% | 5.7 | -4.0 | -3.8 | -5.4 | -4.4 | -3.2 | -3.2 | -2.4 | -4.9 | -5.8 | -2.6 | -4.4 |
| 12% | - | - | -5.5 | -6.1 | - | - | -2.6 | 0.2 | - | - | -3.2 | -2.4 |
| 13% | - | - | -7.1 | -7.2 | - | - | 1.1 | -2.1 | - | - | -2.0 | -3.3 |
| 14% | -3.5 | -4.3 | -4.5 | -7.4 | -7.6 | -2.7 | -2.5 | -0.9 | -3.0 | -4.0 | -2.4 | -2.5 |
| 15% | - | - | -5.6 | -3.0 | - | - | -1.9 | -2.5 | - | - | -4.6 | -0.3 |
| 17% | - | - | -5.3 | -5.7 | - | - | 0.2 | -4.9 | - | - | 0.8 | -3.1 |
| 18% | -7.2 | -4.8 | -3.7 | -6.9 | -5.6 | -2.4 | -1.4 | -4.4 | -2.8 | -1.7 | -5.0 | -3.2 |
| 22% | - | -6.5 | -6.1 | -2.8 | - | -4.6 | -2.3 | -3.2 | - | -3.3 | -5.0 | -3.6 |
| 26% | - | - | -6.4 | -7.0 | - | - | -2.4 | -3.2 | - | - | -3.5 | 1.5 |
| 30% | - | - | - | -2.9 | - | - | - | 0.2 | - | - | - | -6.1 |

Table 8.12: Percent change in TPH with %-Occ.
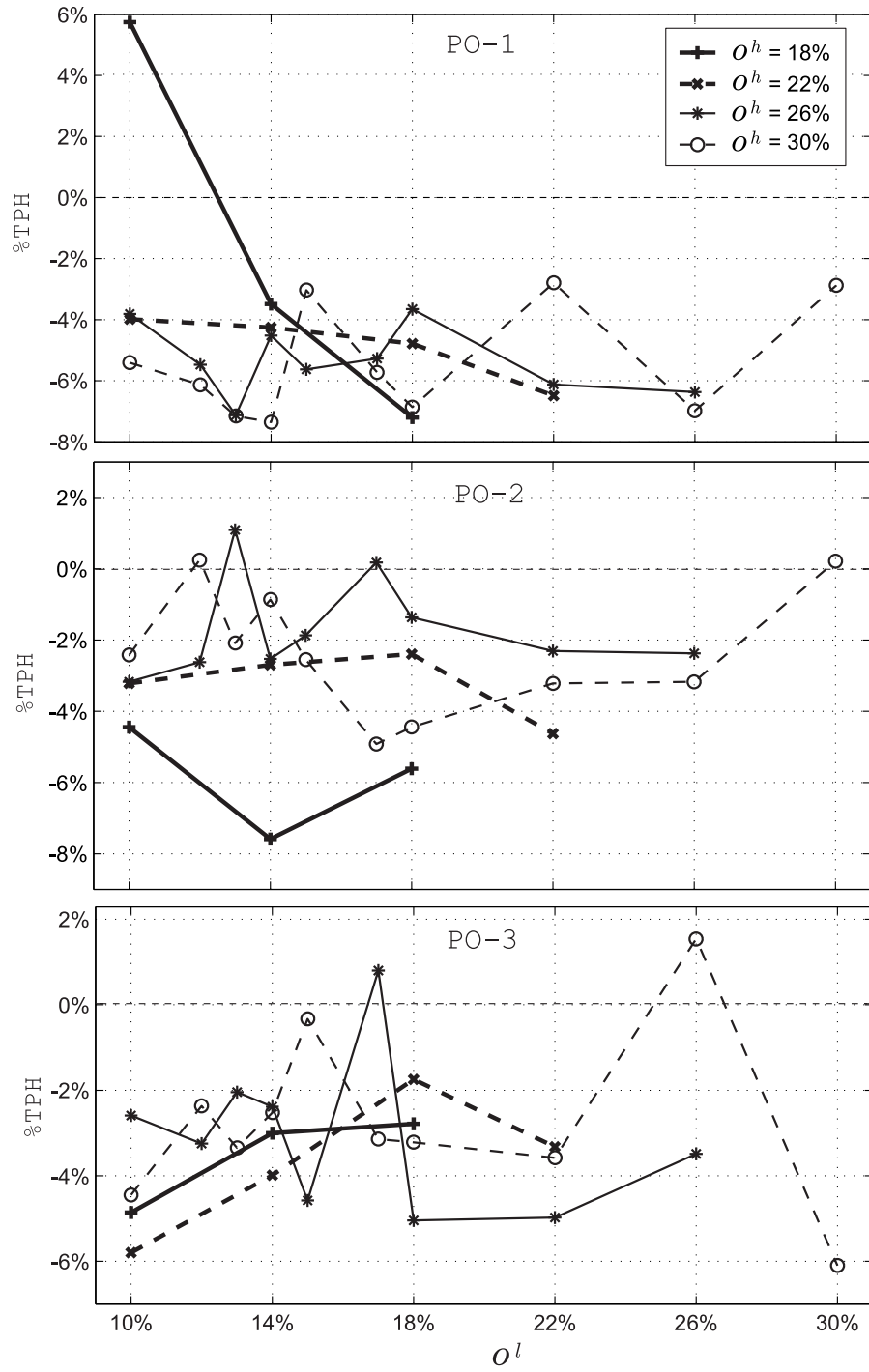
Figure 8.12: Performance with %-Occ.

# Chapter 9

# Summary and Conclusions

This dissertation has investigated several aspects of the design and evaluation of on-ramp metering control systems for congested freeways. A survey of the literature on macroscopic models and optimization-based ramp metering schemes was presented in Chapter 2. A discussion of the rationale for regulating the on-ramp flows identified three mechanisms which can be exploited by on-ramp metering controllers in order to reduce travel time: the prevention of a *capacity drop*, the prevention of *offramp blockage*, and traffic *diversion*. It was concluded that the optimal policy depends on which of these mechanisms is prevalent. For freeways in which the prevention of the capacity drop is the only way to reduce travel time, the optimal policy is to keep the freeway in a free-flow state at all times. On the other hand, for freeways without a capacity drop, where the only available mechanism is the prevention of offramp blockage, the best strategy is to reduce congestion only sufficiently to unclog the offramps.

The limited amount of storage space on the ramps was identified in Chapter 2 as an important consideration for on-ramp metering design, since it restricts the ability of the controller to reduce congestion. In many cases, a scarcity of on-ramp storage space may make the complete elimination of congestion impossible. The I-210 test site described in Chapter 6 was found to be one such case. In this situation, the mechanism of capacity drop is lost in the congested portions of the freeway, and offramp blockage becomes the main concern. The goal of ramp metering is then to distribute the unavoidable congestion so that the offramps remain unblocked whenever possible, and to dispel the congestion quickly in order to recover the lost capacity. The amount and distribution of on-ramp storage space becomes an especially important factor for control design when congestion is unavoidable.

Chapter 3 described a macroscopic model of freeway traffic which was used in Chapter 5 in the design of a new optimization-based ramp metering strategy. The proposed design has the following attributes:

- *It is dynamic.* The underlying macroscopic model, the Asymmetric Cell Transmission Model, is a discrete form of the dynamic LWR model, and related to the original Cell Transmission Model of Daganzo [21, 22].

- *It considers both free-flow and congested traffic states.* The simplifying assumption of persistent free-flow traffic is not made. This is important because, as noted before, it

is often not possible to avoid congestion when on-ramp queue constraints are imposed.

- *It requires only to solve a single linear program.* The exact solution to the full nonlinear problem can be found via a linear program. This transformation is possible when the objective being minimized is the Generalized Total Travel Time (gTTT), as opposed to the standard Total Travel Time (TTT), and the cost weights of the gTTT are obtained from the Cost Weights Synthesis procedure described in Section 5.2. Although the use of gTTT has the disadvantage that TTT may not be minimized, it has the significant advantage that *globally* optimal solutions can be found using very *efficient* LP solvers.

- *It observes on-ramp queue length constraints.* It was found that hard constraints on the on-ramp queue lengths can be imposed if none of the on-ramp flows are restricted because of insufficient space on the mainline (Assumption #1, page 73). Thus, as opposed to many other ramp metering designs ([3, 28, 48, 49] excluded), the proposed technique considers the distribution of on-ramp storage space in the system.

- *It provides a near global optimum solution.* The globally optimal solution to the non-linear problem is found with the solution to the equivalent linear program. This is in contrast to most other techniques that seek only local minima. The solution is said to be a "near global" optimum, because it is global with respect to gTTT, and not necessarily TTT. These two objective functions were found to be *qualitatively* similar, in that both are minimized by maximizing the mainline flows with linearly decreasing cost weights, whenever Assumptions #1 and #2 (page 74) are made. A second reason for "near global" optimality is that the optimal plan assumes that ramp metering rates can be set to zero, whereas in reality the meters on I-210 can be set to no less than 180 vph. Hence, the optimal plan must be modified before implementation by increasing all rates to at least 180 vph (Eq. (5.49)). The example of Section 5.5 found that this extra step produced only a 1.12% sacrifice in travel time savings.

The optimization technique was tested in Section 5.5 using data from the I-210 test site. It was found to produce an 11.74% reduction in TTT, when the on-ramp queues were left unrestricted, and 8.44% when they were constrained.

Chapters 4 and 8 respectively presented macroscopic and microscopic studies of Percent-Occupancy (%-Occ) and Alinea control; two widely used local traffic responsive ramp metering strategies. These studies investigated several questions of interest to freeway analysts, such as: Are either %-Occ or Alinea capable of reducing travel time? Can Alinea be used on freeways where the mainline detectors are placed upstream of the on-ramps? How should the controller parameters be chosen in either case?

The macroscopic study of Chapter 4 focused on a single on-ramp and incorporated the effect of capacity drop by introducing a discontinuity in the original ACTM model. The discontinuous ACTM was cast as a hybrid system with 7 finite states, each of which evolved according to a discrete-time linear state equation. A set of criteria defining a "good strategy" was then delineated. These criteria required the controller to eliminate congestion by driving the system to the UU-I finite state, in fairly direct way. This was expressed in terms of conditions on the closed-loop steady-state values in Table 4.6. Furthermore, after reaching the UU-I finite state, the mainline flow was required to converge to some value exceeding

the queue discharge rate. It was found that neither %-Occ nor Alinea meet all of the stated criteria, although Alinea failed in only one relatively unimportant condition (marginal instead of asymptotic stability of mode `IV`). The main conclusions from this study are:

- With respect to detector placement, both upstream and downstream positions have their advantages. The system is observable from the downstream measurement when the freeway section is in free-flow (mode `I`), or lightly congested (mode `IV`). However it is never observable from the upstream measurement. On the other hand, when the freeway is completely congested, in modes `V` and `VI`, the downstream density can be deduced from upstream measurements, whereas the upstream density cannot be deduced from the downstream measurement. In this situation, the upstream detector location is preferred.

- No selection of parameters for the %-Occ strategy exists that both rapidly dissipates congestions and increases throughput during free-flow. This was concluded due to a conflict in the parameter requirements for modes `I`, `IV`, and `V`. Sacrificing the behavior during congestion, the range of parameters that increases uncongested throughput was depicted in Figure 4.8. Exhaustive simulation lead to the conclusion that no selection of parameters in this range eliminated congestion for more than 84% of all initial conditions.

- Recommendations for Alinea's target occupancy and gain were provided in Eqs. (4.13) and (4.14). Simulations were conducted with gain values ranging from zero to the maximum stable value of 2.6, according to Eq. (4.13). Alinea's response was found to be slow for gains below 0.3. An infinite loop between modes `UC-IV` and `UU-I` was observed for gains larger than 1.7. Behavior between these two extremes was relatively uniform, with the fastest average convergence being registered with $K_r = 0.65$.

The microscopic study of local traffic responsive ramp metering was covered in Chapters 6 through 8. The test site for the study was a 15-mile stretch of Westbound Interstate 210. The methodology that was followed for building the microscopic model was described in Chapter 6. This procedure included the gathering and processing of field data from the PeMS database and a District 7 survey of ramps, the estimation of OD matrices using FREQ, and the coding of the different elements in VISSIM. It was found in Section 6.4 that some of the loop-detectors on I-210, especially those on its offramps, do not provide consistent and reliable measurements. Gaps in the data were filled by using ramp counts from the D07 survey, and by combining data from several *typical* days.

Chapter 7 described the coding and calibration procedure. The calibration was aided by speed contour plots generated using PeMS data. From these contour plots, three separate but interacting bottlenecks were identified. Presumed causes behind the breakdown of traffic at each of the bottlenecks were suggested in Section 6.8 by comparing the demands with the nominal capacities. It was concluded that two of the bottlenecks are geometry-induced, while a third is at least partially caused by weaving. VISSIM's driver behavior parameters provided more than sufficient flexibility for replicating these behaviors. In conclusion, VISSIM was found to be a suitable tool for testing complex control strategies on medium-sized freeway

systems. Our simulations ran at approximately 2 times real time on a Pentium 4, 512 Mb, 2.6 GHz laptop computer, when the freeway was fully loaded.

The simulation runs with local on-ramp metering control lead to the following observations:

- Tests with several different levels of fixed rate ramp metering were conducted, with only the 600 vph experiment producing any reduction in total travel time, when no overrides were applied (4.4%). When the complete set of overrides was enforced, both the 600 vph and 900 vph resulted in small travel time savings (1.5% and 1.3%).

- Figure 8.8 on page 126 showed that comparable reductions in travel time can be achieved with either upstream or downstream mainline detector banks. The plot also supports the notion that upstream detectors may be preferred during congestion, while downstream detectors work better in free-flow conditions. The optimal target occupancy when downstream detectors are used is near the critical occupancy, as suggested by [77]. With upstream detectors, the best choice may be larger.

- The choice of gain with Alinea is less important than the target occupancy. Using downstream detectors and the optimal choice of target occupancy, Alinea's performance was practically uniform over a large range of gains. Only very small values should be avoided. Contrary to Chapter 4, the microscopic study did not conclude that Alinea's gain should not be set too high.

- A number of experiments using %-Occ control and with a range of parameter values were conducted. Almost all resulted in moderate reductions in travel time. However, unlike Alinea, it was difficult to identify a trend in the results that could be used to make a concrete recommendation on parameter selection. It is concluded therefore that some savings can be expected with %-Occ, but that optimizing its performance for a given freeway may be difficult.

The methods of optimal ramp metering and microscopic testing presented in this dissertation provide ample opportunity for future research. In order to implement the proposed technique, it must first be made traffic-responsive by embedding it in a control structure including real-time measurements, model parameter estimation, and demands prediction.

# Bibliography

[1] J. Acha-Daza and F. Hall. Application of catastrophe theory to traffic flow variables. *Transportation Research. Part B*, 28(3), 1994.

[2] K. Agyemang-Duah and F. Hall. Some issues regarding the numerical value of freeway capacity. *International Symposium on Highway Capacity*, 1991.

[3] A. Alessandri, A. Di Febraro, A. Ferrara, and E. Punta. Optimal control of freeways via speed signalling and ramp metering. *Control Engineering Practice*, 6, 1998.

[4] R. Ansorge. What does the entropy condition mean in traffic flow theory? *Transportation Research. Part B*, 24(2), 1990.

[5] J. Banks. Two-capacity phenomenon at freeway bottlenecks: A basis for ramp metering? *Transportation Research Record*, 1320, 1991.

[6] J. Banks. Effect of response limitations on traffic-responsive ramp metering. *Transportation Research Record*, 1394, 1993.

[7] J. Banks. Are minimization of delay and minimization of freeway congestion compatible ramp metering objectives? *Transportation Research Record*, 1727, 2000.

[8] Transportation Research Board. *Highway Capacity Manual*. National Research Council, Washington, D.C., 2000.

[9] K. Bogenberger and A. D. May. Advanced coordinated traffic responsive ramp metering strategies. Technical Report UCB-ITS-PWP-99-19, Institute of Transportation Studies, U.C. Berkeley, 1999.

[10] M. Cassidy. Critique of a freeway on-ramp metering scheme and broader related issues. Technical Report UCB-ITS-RR-2002-4, Institute of Transportation Studies, U.C. Berkeley, 2002.

[11] M. Cassidy. Freeway onramp metering, delay savings, and the diverge bottleneck. *Intellimotion*, 10(3), 2003.

[12] M. Cassidy, S. Anani, and J. Haigwood. Study of freeway traffic near an off-ramp. Technical Report UCB-ITS-PWP-2000-10, Institute of Transportation Studies, U.C. Berkeley, 2000.

[13] M. Cassidy and R. Bertini. Some traffic features at freeway bottlenecks. *Transportation Research. Part B*, 33(1), 1999.

[14] M. Cassidy and J. Rudjanakanoknad. Study of traffic at a freeway merge and roles for ramp metering. Technical Report UCB-ITS-PWP-2002-2, Institute of Transportation Studies, U.C. Berkeley, 2002.

[15] G. Chang, J. Wu, and H. Lieu. Real-time incident-responsive corridor control: A successive linear programming approach. In *Proceedings of the 1994 Annual Meeting of IVHS America*, volume 2, 1994.

[16] T. Chang and Z. Li. Optimization of mainline traffic via an adaptive coordinated ramp metering control model with dynamic OD estimation. *Transportation Research. Part C*, 10, 2002.

[17] C. Chen, J. Cruz, and J. Paquet. Entrance ramp control for travel-rate maximization in expressways. *Transportation Research*, 8, 1974.

[18] R. Crayford, W.H. Lin, and C. Daganzo. The NETCELL simulation package: Technical description. Technical Report UCB-ITS-PRR-97-23, Institute of Transportation Studies, U.C. Berkeley, 1997.

[19] M. Cremer and A. D. May. An extended traffic model for freeway control. Technical Report UCB-ITS-RR-85-7, Institute of Transportation Studies, U.C. Berkeley, 1985.

[20] M. Cremer and M. Papageorgiou. Parameter identification for a traffic flow model. *Automatica*, 17(6), 1981.

[21] C. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research. Part B*, 28(4), 1994.

[22] C. Daganzo. The cell transmission model, Part II: Network traffic. *Transportation Research. Part B*, 29(2), 1995.

[23] C. Daganzo. A finite difference approximation of the kinematic wave model of traffic flow. *Transportation Research. Part B*, 29(4), 1995.

[24] C. Daganzo. Requiem for second-order fluid approximations of traffic flow. *Transportation Research. Part B*, 20(3), 1995.

[25] C. Daganzo, editor. *Fundamentals of Transportation and Traffic Operations*. Pergamon, 1st edition, 1997.

[26] J. Del Castillo, P. Pintado, and G. Benitez. The reaction time of drivers and the stability of traffic flow. *Transportation Research. Part B*, 28(1), 1994.

[27] N. Derzko, A. J. Ugge, and E. R. Case. Evaluation of dynamic freeway flow model by using field data. *Transportation Research Record*, 905, 1983.

[28] A. Di Febbraro, D. Di Noto, T. Parisini, V. Recagno, V. Sacone, and R. Zoppoli. Neural networks for minimization of travelling time on freeway systems. *Transportation systems : Theory and application of advanced technology*, 2, 1995.

[29] M. Eldor. *Integrated real-time freeway on-ramp control strategies*. PhD thesis, University of California at Berkeley, 1976.

[30] A. Erera, C. Daganzo, and D. Lovell. The access control problem on capacitated FIFO networks with unique OD paths is hard. Technical Report UCB-ITS-PRR-99-35, Institute of Transportation Studies, 1999.

[31] M. Fellendorf and P. Vortisch. Validation of the microscopic traffic flow model VISSIM in different real-world situations. *National Research Council. 80th Meeting of the Transportation Research Board*, Washington, D.C., January, 2001.

[32] D. Gazis, R. Herman, and G. Weiss. Density oscillations between lanes of a multilane highway. *Operations Research*, 10(5), 1962.

[33] B. Greenshields. A study of traffic capacity. *Proceedings of the 14th annual meeting of the Highway Research Board*, 1934.

[34] M. Grewal and H. Payne. Identification of parameters in a freeway traffic model. *IEEE Transactions on systems, man, and cybernetics*, 6(3), 1976.

[35] H. Hadj-Salem and J. M. Blosseville. Alinea: A local feedback law for on-ramp metering; A real-life study. In *Third International Conference on Road Traffic Control*, 1990.

[36] F. Hall and K. Agyemang-Duah. Freeway capacity drop and the definition of capacity. *Transportation Research Record*, (1320), 1991.

[37] F. Hall, B. Allen, and M. Gunter. Empirical Analysis of freeway flow-density relationships. *Transportation Research. Part A*, 20(3), 1986.

[38] A. Hegyi, B. De Schutter, H. Hellendoorn, and T. van den Boom. Optimal coordination of ramp metering and variable speed control - An MPC approach. In *Proceedings of the American Control Conference*, 2002.

[39] Y. Iida, T. Hasegawa, Y. Asakura, and C. Shao. A formulation of on-ramp traffic control system with route guidance for urban expressway. *Communications in Transportation*, 1989.

[40] Texas Transportation Institute. Advanced freeway system ramp metering strategies for Texas. Technical Report RR-1232-23, Texas Department of Transportation, 1993.

[41] A. Kaya. *Optimization of traffic flow for an urban freeway transportation system*. PhD thesis, University of Minnesota, 1970.

[42] B. Kerner. Congested traffic flow: Observations and theory. *Transportation Research Record*, (1678), 1999.

[43] B. Kerner. Theory of breakdown phenomenon at highway bottlenecks. *Transportation Research Record*, (1710), 2000.

[44] B. Kerner. Tracing and forecasting of congested patterns for highway traffic management. *IEEE Intelligent Transportation Systems Conference Proceedings*, 2001.

[45] A. Kotsialos and M. Papageorgiou. Integrated optimal control of motorway traffic networks. In *Proceedings of the American Controls Conference*, 1999.

[46] A. Kotsialos, M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham. Traffic flow modeling of large-scale motorway network using the macroscopic modeling tool METANET. *IEEE Transaction on Intelligent Transportation Systems*, 3(4), 2002.

[47] A. Kotsialos, M. Papageorgiou, M. Mangeas, and H. Hadj-Salem. Coordinated and integrated control of motorway networks via nonlinear optimal control. *Transportation Research. Part C*, 10(1), 2002.

[48] A. Kotsialos, M. Papageorgiou, and F. Middelham. Optimal coordinated ramp metering with advanced motorway optimal control. *Transportation Research Record*, 1748, 2001.

[49] A. Kotsialos and M. Papgeorgiou. Efficiency versus fairness in network-wide ramp metering. In *IEEE Intelligent Transportation Systems Conference Proceedings*, 2001.

[50] R. Kuhne. Macroscopic freeway model for dense traffic - Stop-start waves and incident detection. *Ninth International Symposium on Transportation and Traffic Theory*, 1984.

[51] J. Laval and J.C. Muñoz. System optimum diversion of congested freeway traffic. Technical Report UCB-ITS-RR-2002-6, Institute of Transportation Studies, 2002.

[52] M. Lighthill and G. Whitham. On kinematic waves II. A theory of traffic flow on long crowded roads. *Proceedings Royal Society of London, Part A*, 229(1178), 1955.

[53] W.H. Lin and D. Ahanota. Validating the basic cell transmission model on a single freeway link. *PATH Technical Note*, (3), 1995.

[54] W.H. Lin and C. Daganzo. Technical description of NETCELL. General framework and data structure. *PATH Technical Note*, (7), 1994.

[55] G. Liu, A. Lyrintzis, and P. Michalopoulos. Improved high-order model for freeway traffic flow. *Transportation Research Record*, (1644), 1998.

[56] H. Lo, N. Zhang, and W. Lam. Estimation of an origin-destination matrix with random link choice proportions: A statistical approach. *Transportation Research. Part B*, 30(4), 1996.

[57] H. Lo, N. Zhang, and W. Lam. Decomposition algorithm for statistical estimation of OD matrix with random link choice proportions from traffc counts. *Transportation Research. Part B*, 33(5), 1999.

[58] D. Lovell. *Traffic Control on Metered Networks without Route Control*. PhD thesis, U.C. Berkeley, 1997.

[59] D. Lovell and C. Daganzo. Access control on networks with unique origin-destination paths. *Transportation research. Part B*, 34(3), 2000.

[60] A. Lyrintzis, G. Liu, and P. Michalopoulos. Development and comparative evaluation of high-order traffic flow models. *Transportation Research Record*, (1457), 1994.

[61] A. D. May. Experimentation with manual and automatic ramp control. *Highway Research Record*, (59), 1964.

[62] A. D. May. Models for freeway corridor analysis. *Transportation Research Board Special Report*, (194), 1981.

[63] P. Michalopoulos and D. Beskos. Improved continuum models of freeway flow. In *9th International Symposium on Transportation and Traffic Theory*, 1984.

[64] P. Michalopoulos, D. Beskos, and Y. Yamamuchi. Multilane traffic flow dynamics: some macroscopic considerations. *Transportation Research. Part B*, 18(4), 1984.

[65] L. Muñoz, X. Sun, D. Sun, G. Gomes, and R. Horowitz. Methodological calibration of the cell transmission model. In *Proceedings of the American Controls Conference*, 2004.

[66] P. Munjal, Y. Hsu, and R. Lawrence. Analysis and validation of lane-drop effects on multi-lane freeways. *Transportation Research*, 5, 1971.

[67] P. Munjal and L. Pipes. Propagation of on-ramp density perturbations on unidirectional two- and three-lane freeways. *Transportation Research*, 5, 1971.

[68] G.F. Newell. Comments on traffic dynamics. *Transportation Research. Part B*, 23(5), 1989.

[69] G.F. Newell. A simplified theory of kinematic waves in highway traffic, Part I: General theory. *Transportation Research. Part B*, 27(4), 1993.

[70] G.F. Newell. A simplified theory of kinematic waves in highway traffic, Part II: Queueing at freeway bottlenecks. *Transportation Research. Part B*, 27(4), 1993.

[71] G.F. Newell. A simplified theory of kinematic waves in highway traffic, Part III: Multi-destination flows. *Transportation Research. Part B*, 27(4), 1993.

[72] Office of Freeway Operations. SWARM study final report, on W/B Foothill Freeway (W/B LA-210). Technical report, Caltrans District 07, 2002.

[73] M. Papageorgiou. A integrated control approach for traffic corridors. *Transportation Research. Part C*, 3(1), 1995.

[74] M. Papageorgiou. Some remarks on macroscopic traffic flow modelling. *Transportation Research. Part A*, 32(5), 1998.

[75] M. Papageorgiou, J.M. Blosseville, and H. Hadj-Salem. Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris. *Transportation Research. Part B*, 23(1), 1989.

[76] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang. Review of road traffic control strategies. In *Proceedings of the IEEE*, volume 91, 2003.

[77] M. Papageorgiou, H. Hadj-Salem, and F. Middelham. Alinea local ramp metering. Summary of field results. *Transportation Research Record*, (1603), 1997.

[78] M. Papageorgiou, B. Posch, and G. Schmidt. Comparison of macroscopic models for control of freeway traffic. *Transportation Research. Part B*, 17(2), 1983.

[79] M. Papgeorgiou, H. Hadj-Salem, and H. M. Blosseville. Alinea: A local feedaback control law for onramp metering. *Transportation Research Record*, 1320, 1991.

[80] H. Payne. Models of freeway traffic and control. *Mathematical Models of Public Systems*, 1(1), 1971.

[81] H. Payne. FREFLO: A macroscopic simulation model of freeway traffic. *Transportation Research Record*, (722), 1979.

[82] H. Payne and W. Thompson. Allocation of freeway ramp metering volumes to optimize corridor performance. In *IEEE Transactions on Automatic Control*, volume 19, 1974.

[83] W. Phillips. A kinetic model for traffic flow with continuum implications. *Transportation Planning and Technology*, 5, 1979.

[84] G. Piotowicz and J. Robinson. Ramp Metering Satus in North America, 1995 Update. Technical Report DOT-T-95-17, U.S. Department of Transportation.

[85] L. Pipes. Vehicle accelerations in the hydrodynamic theory of traffic flow. *Transportation Research*, 3, 1969.

[86] Traffic Operations Program. Ramp Metering Design Manual. Technical report, www.dot.ca.gov/hq/traffops/systemops/ramp_meter.pdf, 2000.

[87] PTV AG. Vissim User Manual - V.3.70. 2003.

[88] P. Richards. Shock waves on the highway. *Operations Research*, 4(1), 1956.

[89] J. Robinson and M. Doctor. Ramp metering: Does it really work? In *ITE Compendium of Technical Papers*, 1989.

[90] P. Ross. Traffic dynamics. *Transportation Research. Part B*, 22(6), 1988.

[91] K. Sanwal, K. Petty, and J. Walrand. An extended macroscopic model for traffic flow. *Transportation Research. Part B*, 30(1), 1996.

[92] H. Sherali and T Park. Estimation of dynamic origin-destination trip tables for a general network. *Transportation Research. Part B*, 35(3), 2001.

[93] B. Son. *A study of G.F. Newell's "Simplified theory of kinematic waves in highway traffic"*. PhD thesis, University of Toronto, 1996.

[94] J. Tan. Calibration of cell transmission model to morning commute traffic on I-210W. Master's thesis, University of California at Berkeley, 2003.

[95] J. Treiterer and J. Myers. The hysteresis phenomenon in traffic flow. *Transportation and Traffic Theory*, 6, 1974.

[96] J. Wang and A. D. May. Computer model for optimal freeway on-ramp control. *Highway Research Record*, 469, 1973.

[97] J. G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings, Institution of Civil Engineers*, 1, 1952.

[98] J. Wattleworth. Peak period analysis and control of a freeeway system. *Highway Research Record*, 157, 1967.

[99] G.B. Whitham. *Linear and nonlinear waves*. John Wiley & Sons Inc., New York, 1974.

[100] R. Wiedemann. Simulation des Straßenverkehrsflusses. *Schtiftenreihe des Instituts fur Verkehrswesen der Universitat Karlsruhe, Heft 8*, 1974.

[101] R. Wiedemann. Modelling of RTI-Elements on multi-lane roads. *Advanced Telematics in Road Transport*, 1991.

[102] M. Wills. A flexible gravity-opportunities model for trip distribution. *Transportation Research. Part B*, 20(2), 1986.

[103] H. Yang, Y. Iida, and T. Sasaki. An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts. *Transportation Research. Part B*, 25(5), 1991.

[104] H. Yang and S. Yagar. Traffic assignment and traffic control in general freeway-arterial corridor systems. *Transportation Research. Part B*, 28(6), 1994.

[105] H. Yang, S. Yagar, Y. Iida, and Y. Asakura. An algorithm for the inflow control problem of urban freeway networks with user optimal flows. *Transportation Research. Part B*, 28(2), 1994.

[106] L. Yuan and J. Kreer. Adjustment of freeway ramp metering rates to balance entrance ramp queues. *Transportation Research*, 5, 1971.

[107] M. Zhang. A theory of nonequilibrium traffic flow. *Transportation Research. Part B*, 32(7), 1998.

[108] M. Zhang. A mathematical theory of traffic hysteresis. *Transportation Research. Part B*, (33), 1999.

[109] M. Zhang. A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Research. Part B*, 36, 2002.

[110] M. Zhang, R. Jayakrishnan, and W. Recker. An optimization algorithm for freeway traffic control. In *IEEE Intelligent Transportation Systems Conference*, 2001.

[111] M. Zhang and W. Recker. On optimal freeway ramp control policies for congested traffic corridors. *Transportation Reseach Part B*, 33, 1999.

[112] M. Zhang, S. Ritchie, and Z. Lo. Macroscopic modeling of freeway traffic using an artificial neural network. *Transportation Research Record*, (1588), 1997.

[113] M. Zhang, S. Ritchie, and W. Recker. Some general results on the optimal ramp control problem. *Transportation Research. Part C*, 4(2), 1996.

[114] A. Ziliaskopoulos. A linear programming model for the single destination system optimum dynamic traffic assignment problem. *Transportation Science*, 34(1), 2000.

# Appendices

# A. Appendix for Chapter 4

## A.1 Open-loop observability and controlability

| Dynamic mode | $O_1^x$ | $O_2^x$ |
|---|---|---|
| I | $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & v \\ 0 & 1 & v(2-v) \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 & 0 \\ 1-v & v & 0 \\ (1-v)^2 & 2v(1-v) & v^2 \end{bmatrix}$ |
| II | $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & v \\ 0 & 1 & v(2-v) \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 & 0 \\ 1-v & 0 & 0 \\ (1-v)^2 & 0 & 0 \end{bmatrix}$ |
| III | $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1-w & 0 \\ 0 & (1-w)^2 & 0 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 & 0 \\ 1-v & 0 & 0 \\ (1-v)^2 & 0 & 0 \end{bmatrix}$ |
| IV | $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1-v & v \\ 0 & (1-v)^2 & 2v(1-v) \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 & 0 \\ 1 & v & 0 \\ 1 & v(2-v) & v^2 \end{bmatrix}$ |
| V | $\begin{bmatrix} 0 & 1 & 0 \\ w & 1 & v \\ w(2-w) & 1 & v(2-v) \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 & 0 \\ 1-w & 0 & 0 \\ (1-w)^2 & 0 & 0 \end{bmatrix}$ |
| VI | $\begin{bmatrix} 0 & 1 & 0 \\ w & 1-w & 0 \\ 2w(1-w) & (1-w)^2 & 0 \end{bmatrix}$ | $\begin{bmatrix} 1 & 0 & 0 \\ 1-w & 0 & 0 \\ (1-w)^2 & 0 & 0 \end{bmatrix}$ |

Table 9.1: Observability matrices.

Rank of $P^V$: $P^V$ has rank 2 unless its $(3,2)$ and $(3,3)$ minors are zero:

$$\begin{cases} (1-\alpha)(\alpha + w(1-\alpha)) - \alpha(1-\alpha)(1-w) = 0 \\ (1-\alpha)(w(1-\alpha)(2-w)+\alpha) - \alpha(1-\alpha)(1-w)^2 = 0 \end{cases}$$

The first line can never be satisfied for any $\alpha \in [0,1)$ and $w > 0$. Therefore, $P^V$ has rank 2 for the range of valid $\alpha$ and $w$.

| Dynamic mode | $P^x$ |
|---|---|
| I | $\begin{bmatrix} 1 & 1-v & (1-v)^2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ |
| II | $\begin{bmatrix} 1 & 1-v & (1-v)^2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ |
| III | $\begin{bmatrix} 1 & 1-v & (1-v)^2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ |
| IV | $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ |
| V | $\begin{bmatrix} 1-\alpha & (1-\alpha)(1-w) & (1-\alpha)(1-w)^2 \\ \alpha & w(1-\alpha)+\alpha & w(1-\alpha)(2-w)+\alpha \\ 0 & 0 & 0 \end{bmatrix}$ |
| VI | $\begin{bmatrix} 1-\alpha & (1-w)(1-\alpha) & (1-w)^2(1-\alpha) \\ \alpha & w(1-\alpha)+\alpha(1-w) & 2w(1-w)(1-\alpha)+\alpha(1-w)^2 \\ 0 & w\alpha & w^2+2\alpha w(1-w) \end{bmatrix}$ |

Table 9.2: Controllability matrices.

Rank of $P^{VI}$: We compute the determinant of $P^{VI}$:

$$|P^{VI}| = (1-w)^2(1-\alpha)w\alpha^2 - \left[ 2w(1-\alpha)(1-2) + \alpha(1-w)^2 \right] w\alpha(1-\alpha)$$

$$\left[ w^2(1-\alpha) + w\alpha(2-w) \right] w(1-\alpha)^2$$

$$\vdots$$

$$= w^3(1-\alpha)^2$$

$P^{VI}$ is therefore full rank in the range of valid $w$ and $\alpha$.

## A.2  %-Occ closed-loop dynamics

| Dynamic mode | $A^x_{occ}$ | $B^x_{occ}$ |
|---|---|---|
| I | $\begin{bmatrix} 1-v & v-K_2 & 0 \\ 0 & 1-v & v \\ 0 & 0 & 1-v \end{bmatrix}$ | $\begin{bmatrix} K_1 \\ 0 \\ q \end{bmatrix}$ |
| II | $\begin{bmatrix} 1-v & -K_2 & 0 \\ 0 & 1 & v \\ 0 & 0 & 1-v \end{bmatrix}$ | $\begin{bmatrix} K_1+f_d \\ -f_d \\ q \end{bmatrix}$ |
| III | $\begin{bmatrix} 1-v & -K_2 & 0 \\ 0 & 1-w & 0 \\ 0 & w & 1 \end{bmatrix}$ | $\begin{bmatrix} K_1+f_d \\ w\bar{\rho}-f_d \\ q-w\bar{\rho} \end{bmatrix}$ |
| IV | $\begin{bmatrix} 1 & v-K_2 & 0 \\ 0 & 1-v & v \\ 0 & 0 & 1-v \end{bmatrix}$ | $\begin{bmatrix} K_1-f_d \\ 0 \\ q \end{bmatrix}$ |
| V | $\begin{bmatrix} 1-w & -K_2(1-\alpha) & 0 \\ w & 1-\alpha K_2 & v \\ 0 & 0 & 1-v \end{bmatrix}$ | $\begin{bmatrix} (1-\alpha)K_1+w\bar{\rho}-f_d \\ \alpha K_1-w\bar{\rho} \\ q \end{bmatrix}$ |
| VI | $\begin{bmatrix} 1-w & -K_2(1-\alpha) & 0 \\ w & -\alpha K_2+1-w & 0 \\ 0 & w & 1 \end{bmatrix}$ | $\begin{bmatrix} K_1(1-\alpha)+w\bar{\rho}-f_d \\ \alpha K_1 \\ q-w\bar{\rho} \end{bmatrix}$ |

Table 9.3: Closed-loop %-Occ.

<u>Maximum $K_2$ in mode V</u>

Characteristic polynomial : $\lambda^2 + \underbrace{(\alpha K_2+w-2)}_{b}\lambda + \underbrace{(1-w+(w-\alpha)K_2)}_{c}$

Definition of $m$:

$$\begin{cases} -2m \;=\; b \;=\; \alpha K_2+w-2 \\[2ex] m^2 \;=\; c \;=\; 1-w+(w-\alpha)K_2 \end{cases}$$

$$\therefore \quad \alpha m^2 + 2(w-\alpha)m + (w^2-2w+\alpha) = 0$$

$$\therefore \quad m > -1 \quad \Leftrightarrow \quad 4\alpha - 2w + w^2 > 0$$

The condition for $\lambda$ in the unit circle is:

$$\| \lambda \| < 1 \quad \Leftrightarrow \quad \begin{cases} c < 1 & \text{if } b^2 < 4c \\[2mm] \begin{cases} b + c + 1 > 0 \\[1mm] -b + c + 1 > 0 \end{cases} & \text{if } b^2 > 4c \end{cases}$$

Case $m > -1$: The crossing point is on the real axis. Therefore, we are in the case $b^2 > 4c$, and the condition for stability is $b + c + 1 > 0$ and $-b + c + 1 > 0$. Inserting the values of $b$ and $c$, and after some manipulation, this becomes:

$$K_2 > 0$$

$$K_2(2\alpha - w) < 2(2 - w)$$

There are two cases to consider: $(2\alpha - w) > 0$ and $(2\alpha - w) < 0$. Assuming $(2\alpha - w) < 0$, the condition becomes $K_2 > -2(2 - w)/(w - 2\alpha)$, which is already covered by $K_2 > 0$. Therefore, the condition for stability in the case $m > -1$ is $K_2 < 2(2 - w)/(2\alpha - w)$.

Case $m < -1$: In this case the condition for stability is $c < 1$, which translates into $K_2 < w/(w - \alpha)$. Putting both cases together we get the maximum bound for $K_2$:

$$\bar{K}_2 = \begin{cases} \frac{2(2-w)}{2\alpha-w} & \text{if } 4\alpha - 2w + w^2 > 0 \\[3mm] \frac{w}{w-\alpha} & \text{else} \end{cases}$$

Maximum $K_2$ in mode $\mathtt{VI}$

Characteristic polynomial : $\lambda^2 + \underbrace{(\alpha K_2 + 2w - 2)}_{b}\lambda + \underbrace{(1 - w)^2 + (w - \alpha)K_2}_{c}$

Definition of $m$:

$$\alpha m^2 + 2(w - \alpha)m - (1 - w)(2w - \alpha - w\alpha) = 0$$

$$\therefore \quad m > -1 \quad \Leftrightarrow \quad 2\alpha - 2w + w\alpha > 0$$

Case $m > -1$, in a similar way as mode V, reduces to $(2 - w)^2/(2\alpha - w)$. The maximum gain in the case $m < -1$ is $w(2 - w)/(w - \alpha)$:

$$\bar{K}_2 = \begin{cases} \frac{(2-w)^2}{2\alpha-w} & \text{if } 2\alpha - 2w + w\alpha > 0 \\ \frac{w(2-w)}{w-\alpha} & \text{else} \end{cases}$$

## A.3  Alinea closed-loop dynamics

Dynamic mode IV    $\rho_{2[k]}$ and $r$ behave as an unforced oscillator:

$$\begin{bmatrix} \rho_{2[k+1]} \\ r_{[k+1]} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -K_r & 1 - K_r \end{bmatrix} \begin{bmatrix} \rho_{2[k]} \\ r \end{bmatrix} + \begin{bmatrix} -f_d + q \\ K_r(\hat{\rho}_2 + f_d - q) \end{bmatrix}$$

Substitute $\zeta_{[k]} = r + q - f_d$ and $\gamma_{[k]} = \zeta_{[k]} + K_r(\rho_{2[k]} - \hat{\rho}_2)$. Then:

$$\gamma_{[k+1]} = \zeta_{[k]}$$

$$\zeta_{[k+1]} = (2 - K_r)\zeta_{[k]} - \gamma_{[k]}$$

$$\therefore \quad \zeta_{[k+2]} + (K_r - 2)\zeta_{[k+1]} + \zeta_{[k]} = 0$$

$\Rightarrow \zeta_{[k]}$ oscillates about $0 \Rightarrow \gamma_{[k]}$ oscillates about $0 \Rightarrow \rho_{2[k]}$ oscillates about $\hat{\rho}_2$.

| Dynamic mode | $A_{occ}^x$ | $B_{occ}^x$ |
|---|---|---|
| I | $\begin{bmatrix} 1-v & v & 0 & 1 \\ 0 & 1-v & v & 0 \\ 0 & 0 & 1-v & 0 \\ -K_r(1-v) & -K_rv & 0 & 1-K_r \end{bmatrix}$ | $\begin{bmatrix} 0 \\ 0 \\ q \\ K_r\hat{\rho}_2 \end{bmatrix}$ |
| II | $\begin{bmatrix} 1-v & 0 & 0 & 1 \\ 0 & 1 & v & 0 \\ 0 & 0 & 1-v & 0 \\ -K_r(1-v) & 0 & 0 & 1-K_r \end{bmatrix}$ | $\begin{bmatrix} f_d \\ -f_d \\ q \\ K_r(\hat{\rho}_2 - f_d) \end{bmatrix}$ |
| III | $\begin{bmatrix} 1-v & 0 & 0 & 1 \\ 0 & 1-w & 0 & 0 \\ 0 & w & 1 & 0 \\ -K_r(1-v) & 0 & 0 & 1-K_r \end{bmatrix}$ | $\begin{bmatrix} f_d \\ w\bar{\rho} - f_d \\ q - w\bar{\rho} \\ K_r(\hat{\rho}_2 - f_d) \end{bmatrix}$ |
| IV | $\begin{bmatrix} 1 & v & 0 & 1 \\ 0 & 1-v & v & 0 \\ 0 & 0 & 1-v & 0 \\ -K_r & -K_rv & 0 & 1-K_r \end{bmatrix}$ | $\begin{bmatrix} -f_d \\ 0 \\ q \\ K_r(\hat{\rho}_2 + f_d) \end{bmatrix}$ |
| V | $\begin{bmatrix} 1-w & 0 & 0 & 1-\alpha \\ w & 1 & v & \alpha \\ 0 & 0 & 1-v & 0 \\ -K_r(1-w) & 0 & 0 & 1-K_r(1-\alpha) \end{bmatrix}$ | $\begin{bmatrix} w\bar{\rho} - f_d \\ -w\bar{\rho} \\ q \\ K_r(\hat{\rho}_2 - w\bar{\rho} + f_d) \end{bmatrix}$ |
| VI | $\begin{bmatrix} 1-w & 0 & 0 & 1-\alpha \\ w & 1-w & 0 & \alpha \\ 0 & w & 1 & 0 \\ -K_r(1-w) & 0 & 0 & 1-K_r(1-\alpha) \end{bmatrix}$ | $\begin{bmatrix} w\bar{\rho} \\ 0 \\ q - w\bar{\rho} \\ K_r(\hat{\rho}_2 - w\bar{\rho} + f_d) \end{bmatrix}$ |

Table 9.4: Closed-loop Alinea.

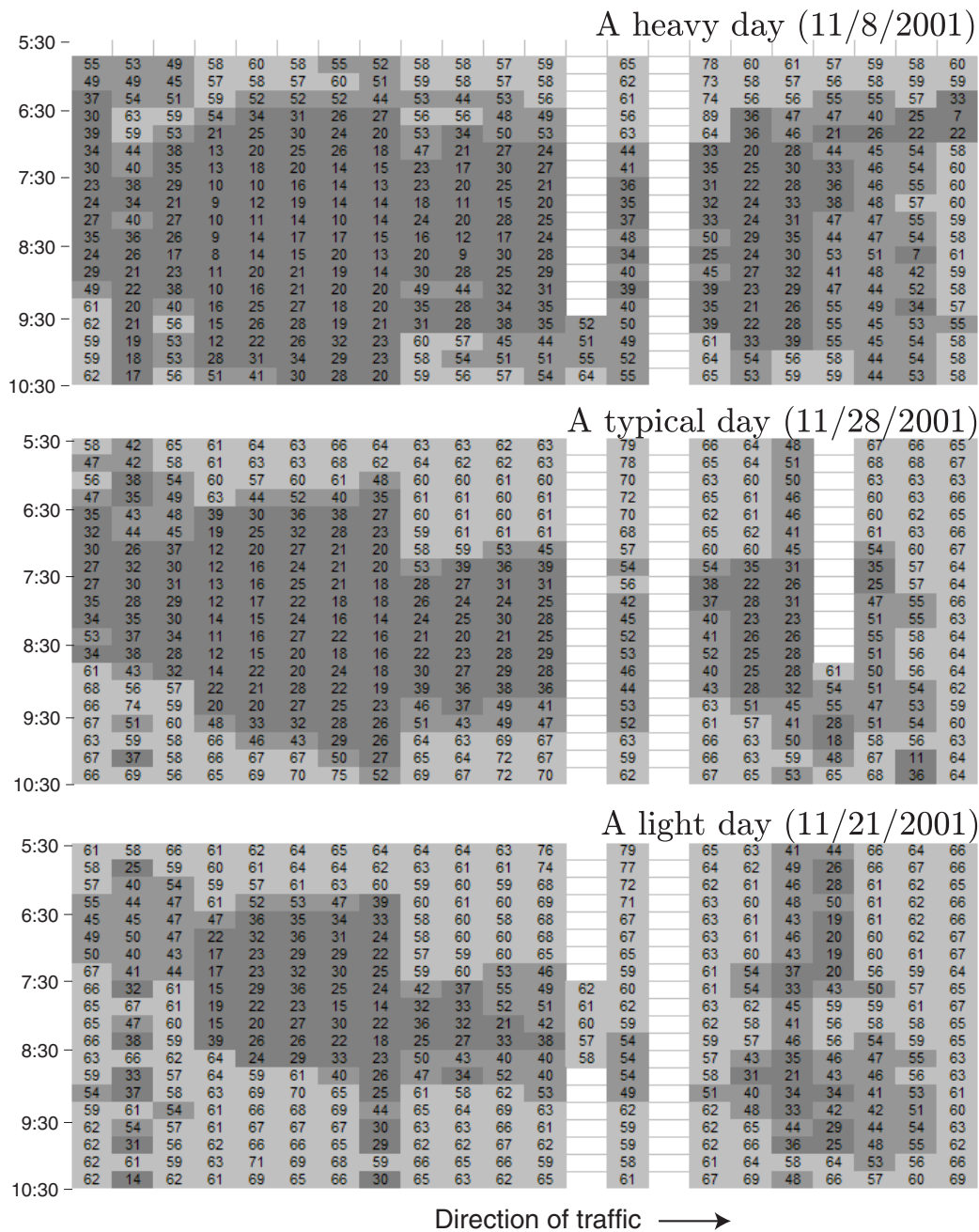# B. Appendix for Chapter 6

## B.1   PeMS speed contours



Figure 9.1: PeMS speed contour plots.

## B.2 Good days from the D07 survey
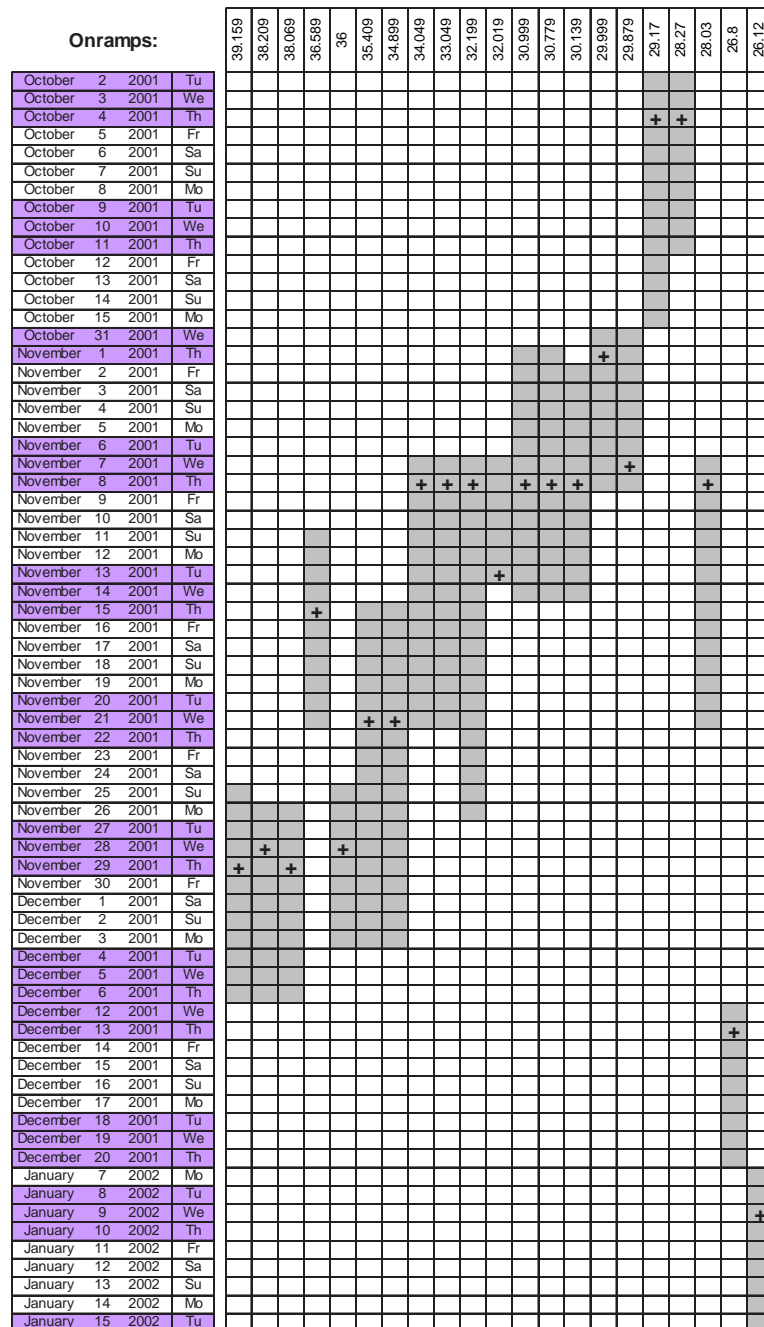


Figure 9.2: On-ramp days.

Figure 9.3: Offramp days.

| Offramps: | | | | 38.209 | 38 | 37 | 36 | 35.409 | 34.049 | 33.049 | 32.019 | 30.779 | 30.5 | 29.17 | 28.27 | 27.64 | 26.8 | 26.12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| October | 2 | 2001 | Tu | | | | | | | | | | | | | | | |
| October | 3 | 2001 | We | | | | | | | | | | | | | | | |
| October | 4 | 2001 | Th | | | | | | | | | | | | + | | | |
| October | 5 | 2001 | Fr | | | | | | | | | | | | | | | |
| October | 6 | 2001 | Sa | | | | | | | | | | | | | | | |
| October | 7 | 2001 | Su | | | | | | | | | | | | | | | |
| October | 8 | 2001 | Mo | | | | | | | | | | | | | | | |
| October | 9 | 2001 | Tu | | | | | | | | | | | | | | | |
| October | 10 | 2001 | We | | | | | | | | | | | | | | | |
| October | 11 | 2001 | Th | | | | | | | | | | | | | | | |
| October | 12 | 2001 | Fr | | | | | | | | | | | | | | | |
| October | 13 | 2001 | Sa | | | | | | | | | | | | | | | |
| October | 14 | 2001 | Su | | | | | | | | | | | | | | | |
| October | 15 | 2001 | Mo | | | | | | | | | | | | | | | |
| October | 31 | 2001 | We | | | | | | | | | | | | | | | |
| November | 1 | 2001 | Th | | | | | | | | | | | + | | | | |
| November | 2 | 2001 | Fr | | | | | | | | | | | | | | | |
| November | 3 | 2001 | Sa | | | | | | | | | | | | | | | |
| November | 4 | 2001 | Su | | | | | | | | | | | | | | | |
| November | 5 | 2001 | Mo | | | | | | | | | | | | | | | |
| November | 6 | 2001 | Tu | | | | | | | | | | | + | | | | |
| November | 7 | 2001 | We | | | | | | | | | | | | | | | |
| November | 8 | 2001 | Th | | | | | + | + | + | + | | | | | | | |
| November | 9 | 2001 | Fr | | | | | | | | | | | | | | | |
| November | 10 | 2001 | Sa | | | | | | | | | | | | | | | |
| November | 11 | 2001 | Su | | | | | | | | | | | | | | | |
| November | 12 | 2001 | Mo | | | | | | | | | | | | | | | |
| November | 13 | 2001 | Tu | | | + | | | | | | | | | | | | |
| November | 14 | 2001 | We | | + | | | | | | | | | | | | | |
| November | 15 | 2001 | Th | | | | | | | | | | | | | | | |
| November | 16 | 2001 | Fr | | | | | | | | | | | | | | | |
| November | 17 | 2001 | Sa | | | | | | | | | | | | | | | |
| November | 18 | 2001 | Su | | | | | | | | | | | | | | | |
| November | 19 | 2001 | Mo | | | | | | | | | | | | | | | |
| November | 20 | 2001 | Tu | | | | | | | | | | | | | | | |
| November | 21 | 2001 | We | | | | + | + | | | | | | | | | | |
| November | 22 | 2001 | Th | | | | | | | | | | | | | | | |
| November | 23 | 2001 | Fr | | | | | | | | | | | | | | | |
| November | 24 | 2001 | Sa | | | | | | | | | | | | | | | |
| November | 25 | 2001 | Su | | | | | | | | | | | | | | | |
| November | 26 | 2001 | Mo | | | | | | | | | | | | | | | |
| November | 27 | 2001 | Tu | | | | | | | | | | | | | | | |
| November | 28 | 2001 | We | + | | | | | | | | | | | | | | |
| November | 29 | 2001 | Th | | | | | | | | | | | | | | | |
| November | 30 | 2001 | Fr | | | | | | | | | | | | | | | |
| December | 1 | 2001 | Sa | | | | | | | | | | | | | | | |
| December | 2 | 2001 | Su | | | | | | | | | | | | | | | |
| December | 3 | 2001 | Mo | | | | | | | | | | | | | | | |
| December | 4 | 2001 | Tu | | | | | | | | | | | | | | | |
| December | 12 | 2001 | We | | | | | | | | | | | | | | | |
| December | 13 | 2001 | Th | | | | | | | | | | | | | | + | |
| December | 14 | 2001 | Fr | | | | | | | | | | | | | | | |
| December | 15 | 2001 | Sa | | | | | | | | | | | | | | | |
| December | 16 | 2001 | Su | | | | | | | | | | | | | | | |
| December | 17 | 2001 | Mo | | | | | | | | | | | | | | | |
| December | 18 | 2001 | Tu | | | | | | | | | | | | | + | | |
| December | 19 | 2001 | We | | | | | | | | | | | | | | | |
| December | 20 | 2001 | Th | | | | | | | | | | | | | | | |
| January | 7 | 2002 | Mo | | | | | | | | | | | | | | | |
| January | 8 | 2002 | Tu | | | | | | | | | | | | | | | |
| January | 9 | 2002 | We | | | | | | | | | | | | | | | + |
| January | 10 | 2002 | Th | | | | | | | | | | | | | | | |
| January | 11 | 2002 | Fr | | | | | | | | | | | | | | | |
| January | 12 | 2002 | Sa | | | | | | | | | | | | | | | |
| January | 13 | 2002 | Su | | | | | | | | | | | | | | | |
| January | 14 | 2002 | Mo | | | | | | | | | | | | | | | |
| January | 15 | 2002 | Tu | | | | | | | | | | | | | | | |

155

# B.3 D07 survey ramp counts



Figure 9.4: District 07 ramp survey (OR=on-ramp, FR=offramp).
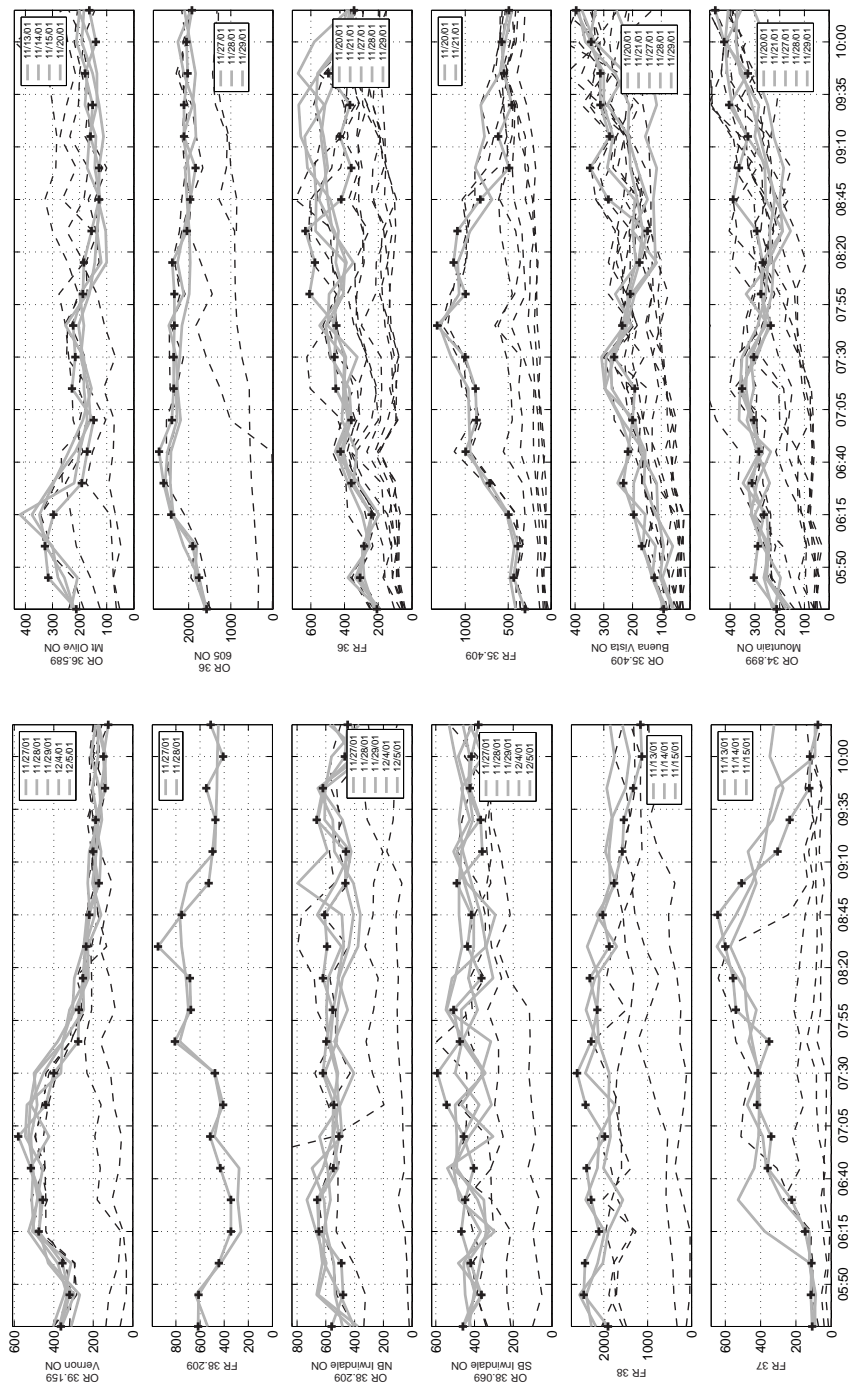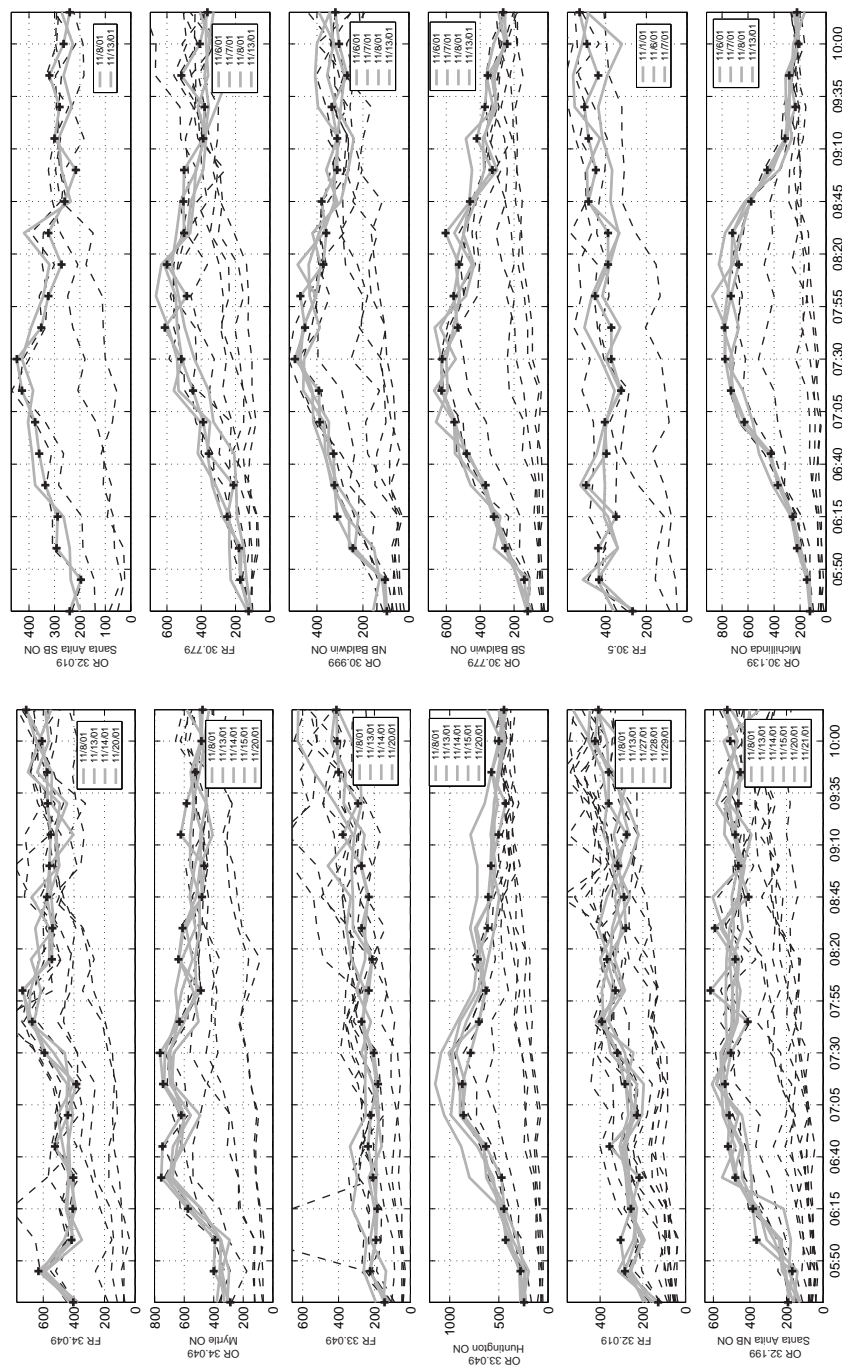
156

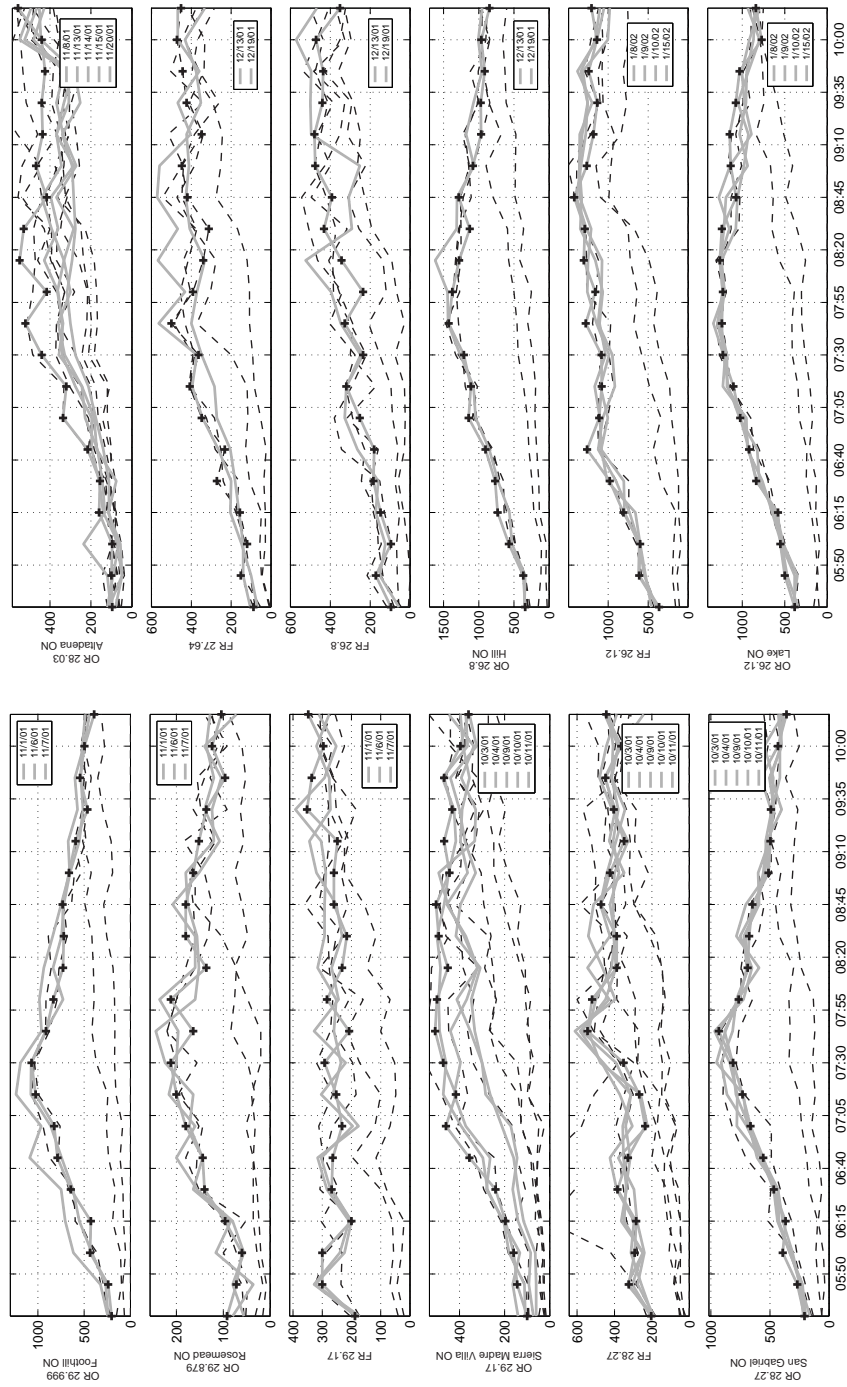Figure 9.5: District 07 ramp survey (OR=on-ramp, FR=offramp).

157

Figure 9.6: District 07 ramp survey (OR=on-ramp, FR=offramp).

158

| Onramps(vph) | 39.159 | 39.159 | 38.209 | 38.069 | 36.589 | 36 | 35.409 | 34.899 | 34.049 | 33.049 | 32.199 | 32.019 | 30.999 | 30.779 | 30.139 | 29.999 | 29.879 | 29.17 | 28.27 | 28.03 | 26.8 | 26.12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5:30 | 7632 | 364 | 560 | 460 | 212 | 1576 | 92 | 212 | 288 | 244 | 192 | 240 | 96 | 120 | 124 | 204 | 92 | 100 | 208 | 96 | 340 | 384 |
| 5:45 | 9521 | 320 | 480 | 364 | 316 | 1752 | 124 | 304 | 400 | 280 | 168 | 196 | 104 | 140 | 148 | 240 | 72 | 144 | 268 | 100 | 368 | 500 |
| 6:00 | 8948 | 356 | 492 | 420 | 328 | 1908 | 168 | 288 | 392 | 432 | 364 | 292 | 244 | 252 | 224 | 436 | 60 | 160 | 392 | 96 | 572 | 552 |
| 6:15 | 9152 | 476 | 648 | 468 | 296 | 2416 | 196 | 264 | 576 | 448 | 384 | 288 | 312 | 320 | 256 | 428 | 96 | 196 | 368 | 160 | 732 | 580 |
| 6:30 | 8608 | 456 | 660 | 448 | 192 | 2596 | 232 | 312 | 756 | 472 | 480 | 336 | 324 | 368 | 372 | 644 | 140 | 240 | 468 | 156 | 768 | 836 |
| 6:45 | 8552 | 516 | 548 | 404 | 172 | 2704 | 216 | 284 | 748 | 632 | 520 | 360 | 328 | 480 | 424 | 788 | 144 | 356 | 556 | 216 | 904 | 920 |
| 7:00 | 8820 | 580 | 508 | 456 | 148 | 2400 | 200 | 304 | 620 | 860 | 512 | 376 | 388 | 552 | 628 | 824 | 180 | 460 | 664 | 336 | 1144 | 1024 |
| 7:15 | 9028 | 440 | 544 | 544 | 228 | 2356 | 192 | 352 | 744 | 872 | 536 | 428 | 392 | 628 | 732 | 1024 | 200 | 416 | 732 | 320 | 1112 | 1104 |
| 7:30 | 8744 | 400 | 620 | 592 | 216 | 2356 | 264 | 304 | 764 | 788 | 504 | 448 | 496 | 624 | 776 | 1068 | 212 | 472 | 812 | 440 | 1212 | 1228 |
| 7:45 | 7936 | 276 | 596 | 476 | 224 | 2340 | 236 | 236 | 632 | 704 | 412 | 352 | 452 | 532 | 780 | 912 | 164 | 508 | 932 | 520 | 1428 | 1240 |
| 8:00 | 7252 | 272 | 552 | 508 | 188 | 2340 | 208 | 276 | 488 | 632 | 616 | 324 | 472 | 556 | 732 | 832 | 212 | 500 | 764 | 416 | 1376 | 1228 |
| 8:15 | 7968 | 252 | 620 | 364 | 184 | 2388 | 176 | 268 | 640 | 716 | 480 | 272 | 372 | 524 | 672 | 728 | 136 | 452 | 688 | 548 | 1280 | 1264 |
| 8:30 | 7100 | 236 | 592 | 436 | 156 | 2040 | 148 | 292 | 612 | 616 | 592 | 324 | 360 | 604 | 720 | 720 | 180 | 492 | 676 | 528 | 1128 | 1240 |
| 8:45 | 6672 | 220 | 608 | 416 | 128 | 1964 | 284 | 388 | 480 | 608 | 408 | 260 | 380 | 460 | 576 | 732 | 180 | 504 | 648 | 416 | 1284 | 1072 |
| 9:00 | 6648 | 172 | 464 | 492 | 128 | 1836 | 348 | 364 | 464 | 580 | 464 | 216 | 312 | 328 | 452 | 664 | 164 | 444 | 512 | 468 | 1084 | 1136 |
| 9:15 | 6432 | 200 | 460 | 360 | 160 | 2112 | 280 | 328 | 624 | 504 | 480 | 300 | 312 | 420 | 316 | 592 | 152 | 468 | 496 | 436 | 968 | 1148 |
| 9:30 | 6298 | 188 | 664 | 368 | 152 | 2112 | 312 | 404 | 584 | 432 | 464 | 280 | 336 | 372 | 236 | 464 | 136 | 432 | 492 | 440 | 972 | 1076 |
| 9:45 | 6072 | 140 | 620 | 424 | 180 | 2024 | 312 | 328 | 524 | 576 | 452 | 320 | 268 | 356 | 284 | 544 | 96 | 468 | 576 | 424 | 916 | 1032 |
| 10:00 | 6192 | 148 | 468 | 416 | 140 | 2052 | 344 | 424 | 484 | 500 | 508 | 264 | 304 | 240 | 212 | 496 | 124 | 396 | 432 | 440 | 964 | 772 |
| 10:30 | 6368 | 124 | 448 | 380 | 164 | 1924 | 396 | 460 | 476 | 448 | 524 | 240 | 320 | 264 | 224 | 392 | 104 | 360 | 360 | 556 | 848 | 840 |

Figure 9.7: On-ramp flows (first column is upstream mainline).

| Offramps(vph) | 38.209 | 38 | 37 | 36 | 35.409 | 34.049 | 33.049 | 32.019 | 30.779 | 30.5 | 29.17 | 28.27 | 27.64 | 26.8 | 26.12 | 25.68 | 25.6 | 25.5 | 25.4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5:30 | 616 | 2200 | 108 | 208 | 304 | 404 | 144 | 128 | 124 | 268 | 188 | 204 | 88 | 96 | 364 | 304 | 1372 | 1372 | 3568 |
| 5:45 | 612 | 2456 | 116 | 308 | 436 | 632 | 224 | 284 | 172 | 432 | 300 | 324 | 152 | 172 | 612 | 436 | 1692 | 1692 | 4088 |
| 6:00 | 444 | 2428 | 112 | 284 | 388 | 416 | 192 | 304 | 180 | 436 | 300 | 292 | 120 | 96 | 600 | 300 | 1974 | 1974 | 4456 |
| 6:15 | 344 | 2108 | 148 | 240 | 496 | 408 | 184 | 256 | 248 | 348 | 200 | 284 | 156 | 148 | 808 | 432 | 2202 | 2202 | 5376 |
| 6:30 | 344 | 2288 | 224 | 360 | 716 | 404 | 208 | 216 | 212 | 496 | 268 | 384 | 272 | 184 | 980 | 440 | 2176 | 2176 | 6044 |
| 6:45 | 432 | 2392 | 360 | 424 | 996 | 524 | 236 | 356 | 356 | 396 | 264 | 328 | 232 | 180 | 1264 | 592 | 1930 | 1930 | 6176 |
| 7:00 | 516 | 1976 | 340 | 360 | 868 | 440 | 220 | 228 | 388 | 404 | 232 | 236 | 348 | 252 | 1112 | 648 | 2534 | 2534 | 6472 |
| 7:15 | 408 | 2416 | 420 | 452 | 880 | 384 | 180 | 284 | 448 | 324 | 252 | 268 | 408 | 320 | 1080 | 600 | 2752 | 2752 | 6560 |
| 7:30 | 476 | 2608 | 416 | 460 | 1000 | 592 | 204 | 320 | 516 | 372 | 292 | 352 | 364 | 236 | 1084 | 656 | 2636 | 2636 | 6556 |
| 7:45 | 808 | 2288 | 352 | 448 | 1328 | 672 | 272 | 392 | 612 | 372 | 208 | 544 | 500 | 328 | 1280 | 764 | 2584 | 2584 | 6648 |
| 8:00 | 676 | 2148 | 540 | 608 | 996 | 736 | 232 | 328 | 484 | 452 | 284 | 520 | 392 | 236 | 1160 | 776 | 2372 | 2372 | 6720 |
| 8:15 | 684 | 2320 | 556 | 576 | 1136 | 544 | 212 | 368 | 600 | 388 | 232 | 388 | 340 | 344 | 1304 | 748 | 2358 | 2358 | 6420 |
| 8:30 | 948 | 1872 | 600 | 632 | 1092 | 540 | 272 | 280 | 500 | 388 | 216 | 388 | 312 | 432 | 1292 | 640 | 2304 | 2304 | 6540 |
| 8:45 | 752 | 2024 | 644 | 420 | 824 | 576 | 232 | 288 | 504 | 484 | 260 | 472 | 420 | 392 | 1424 | 776 | 2222 | 2222 | 6052 |
| 9:00 | 528 | 1764 | 508 | 360 | 488 | 560 | 272 | 316 | 500 | 448 | 260 | 424 | 448 | 476 | 1268 | 628 | 2120 | 2120 | 5536 |
| 9:15 | 496 | 1576 | 304 | 428 | 616 | 552 | 376 | 276 | 388 | 484 | 248 | 348 | 348 | 480 | 1184 | 640 | 2122 | 2122 | 5280 |
| 9:30 | 472 | 1544 | 236 | 368 | 456 | 572 | 292 | 360 | 380 | 504 | 352 | 404 | 424 | 440 | 1136 | 512 | 1926 | 1926 | 4732 |
| 9:45 | 548 | 1328 | 124 | 496 | 548 | 576 | 400 | 360 | 516 | 436 | 336 | 448 | 444 | 436 | 1244 | 672 | 2040 | 2040 | 4664 |
| 10:00 | 408 | 1132 | 120 | 380 | 576 | 608 | 408 | 424 | 408 | 492 | 296 | 364 | 472 | 472 | 1140 | 472 | 1806 | 1806 | 4216 |
| 10:30 | 512 | 1164 | 76 | 344 | 492 | 712 | 412 | 408 | 364 | 528 | 348 | 444 | 452 | 352 | 1208 | 456 | 1588 | 1588 | 3808 |

Figure 9.8: Offramp flows (last column is downstream mainline).

# C. Appendix for Chapter 7

## C.1 Intermediate VISSIM speed contours

| | 39.159 | 38.209 | 38.069 | 36.589 | 35.409 | 34.899 | 34.049 | 33.049 | 32.199 | 32.019 | 30.999 | 30.779 | 30.139 | 29.999 | 29.879 | 29.17 | 28.27 | 28.03 | 26.8 | 26.12 | 25.68 | 25.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5:30 | 63 | 63 | 62 | 63 | 62 | 63 | 63 | 64 | 63 | 63 | 63 | 62 | 63 | 62 | 63 | 63 | 64 | 63 | 64 | 63 | 63 | 64 |
| 5:45 | 63 | 62 | 61 | 63 | 61 | 62 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 64 | 63 | 64 | 64 | 64 | 61 | 64 |
| 6:00 | 63 | 62 | 60 | 63 | 62 | 62 | 63 | 63 | 63 | 62 | 63 | 62 | 62 | 61 | 62 | 63 | 63 | 63 | 63 | 63 | 60 | 64 |
| 6:15 | 63 | 61 | 57 | 63 | 57 | 60 | 62 | 61 | 61 | 60 | 61 | 60 | 62 | 62 | 62 | 63 | 62 | 61 | 63 | 63 | 58 | 63 |
| 6:30 | 63 | 63 | 57 | 63 | 49 | 57 | 60 | 61 | 59 | 57 | 47 | 38 | 57 | 54 | 58 | 61 | 61 | 61 | 62 | 61 | 54 | 62 |
| 6:45 | 63 | 62 | 60 | 63 | 57 | 58 | 61 | 60 | 48 | 37 | 29 | 26 | 56 | 55 | 58 | 61 | 62 | 59 | 61 | 61 | 55 | 62 |
| 7:00 | 63 | 60 | 58 | 63 | 60 | 61 | 62 | 63 | 33 | 24 | 26 | 25 | 56 | 50 | 56 | 61 | 60 | 56 | 60 | 58 | 54 | 62 |
| 7:15 | 63 | 63 | 57 | 63 | 61 | 62 | 63 | 62 | 22 | 28 | 27 | 27 | 47 | 44 | 54 | 59 | 58 | 50 | 61 | 38 | 40 | 62 |
| 7:30 | 63 | 62 | 61 | 63 | 59 | 61 | 62 | 62 | 21 | 23 | 25 | 27 | 29 | 40 | 54 | 60 | 60 | 43 | 29 | 22 | 40 | 61 |
| 7:45 | 63 | 63 | 61 | 64 | 59 | 62 | 63 | 63 | 25 | 26 | 22 | 25 | 36 | 38 | 52 | 58 | 26 | 23 | 21 | 25 | 40 | 62 |
| 8:00 | 63 | 63 | 62 | 64 | 63 | 63 | 64 | 63 | 63 | 52 | 40 | 32 | 26 | 41 | 54 | 33 | 19 | 28 | 27 | 29 | 42 | 61 |
| 8:15 | 63 | 63 | 62 | 64 | 63 | 63 | 64 | 64 | 63 | 63 | 63 | 61 | 58 | 53 | 58 | 43 | 37 | 41 | 35 | 36 | 45 | 61 |
| 8:30 | 63 | 63 | 61 | 64 | 63 | 63 | 64 | 63 | 63 | 61 | 63 | 62 | 63 | 60 | 62 | 63 | 63 | 62 | 59 | 44 | 42 | 61 |
| 8:45 | 63 | 64 | 61 | 64 | 64 | 63 | 64 | 63 | 63 | 63 | 63 | 63 | 63 | 60 | 62 | 63 | 63 | 62 | 63 | 62 | 61 | 64 |
| 9:00 | 64 | 64 | 63 | 64 | 62 | 63 | 63 | 64 | 63 | 62 | 63 | 62 | 63 | 62 | 63 | 63 | 63 | 63 | 63 | 63 | 62 | 64 |
| 9:15 | 63 | 64 | 63 | 64 | 62 | 63 | 63 | 63 | 63 | 63 | 63 | 62 | 63 | 63 | 63 | 63 | 63 | 62 | 63 | 63 | 58 | 64 |
| 9:30 | 64 | 64 | 62 | 64 | 63 | 62 | 63 | 63 | 63 | 63 | 63 | 63 | 62 | 62 | 62 | 63 | 63 | 63 | 63 | 63 | 61 | 63 |
| 9:45 | 64 | 64 | 62 | 64 | 63 | 63 | 64 | 64 | 63 | 63 | 63 | 63 | 62 | 62 | 63 | 64 | 63 | 62 | 63 | 63 | 58 | 64 |
| 10:00 | 63 | 64 | 63 | 64 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 64 | 63 | 63 | 64 | 64 | 63 | 63 | 63 | 60 | 64 |
| 10:15 | 63 | 64 | 62 | 64 | 62 | 61 | 63 | 62 | 62 | 61 | 63 | 62 | 63 | 62 | 63 | 63 | 64 | 63 | 63 | 63 | 61 | 64 |

Figure 9.9: Contour plot after adjusting the look back distances.

| | 39.159 | 38.209 | 38.069 | 36.589 | 35.409 | 34.899 | 34.049 | 33.049 | 32.199 | 32.019 | 30.999 | 30.779 | 30.139 | 29.999 | 29.879 | 29.17 | 28.27 | 28.03 | 26.8 | 26.12 | 25.68 | 25.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5:30 | 63 | 63 | 62 | 64 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 62 | 63 | 64 | 64 | 63 | 64 | 64 | 63 | 64 |
| 5:45 | 63 | 63 | 62 | 63 | 61 | 62 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 64 | 64 | 64 | 64 | 64 | 62 | 64 |
| 6:00 | 63 | 63 | 61 | 64 | 62 | 63 | 63 | 63 | 62 | 62 | 63 | 62 | 61 | 61 | 61 | 63 | 63 | 63 | 63 | 63 | 60 | 64 |
| 6:15 | 63 | 61 | 59 | 63 | 58 | 60 | 62 | 62 | 61 | 62 | 60 | 58 | 62 | 62 | 62 | 63 | 63 | 63 | 63 | 63 | 60 | 64 |
| 6:30 | 63 | 63 | 60 | 63 | 56 | 60 | 61 | 61 | 55 | 50 | 49 | 48 | 48 | 47 | 56 | 61 | 62 | 60 | 61 | 61 | 58 | 63 |
| 6:45 | 63 | 63 | 60 | 63 | 60 | 61 | 62 | 61 | 27 | 33 | 27 | 30 | 55 | 53 | 57 | 61 | 61 | 58 | 62 | 62 | 58 | 63 |
| 7:00 | 63 | 63 | 61 | 63 | 61 | 61 | 62 | 63 | 31 | 26 | 28 | 28 | 55 | 44 | 55 | 60 | 62 | 59 | 60 | 61 | 57 | 63 |
| 7:15 | 63 | 63 | 59 | 63 | 60 | 61 | 63 | 59 | 20 | 33 | 23 | 24 | 22 | 41 | 53 | 59 | 61 | 59 | 62 | 52 | 45 | 62 |
| 7:30 | 63 | 63 | 62 | 63 | 61 | 61 | 62 | 62 | 31 | 15 | 19 | 19 | 23 | 23 | 41 | 55 | 60 | 61 | 57 | 47 | 23 | 43 |
| 7:45 | 63 | 63 | 61 | 64 | 63 | 63 | 64 | 16 | 20 | 27 | 23 | 26 | 26 | 41 | 54 | 60 | 58 | 46 | 25 | 28 | 49 | 62 |
| 8:00 | 63 | 63 | 63 | 64 | 63 | 63 | 64 | 52 | 18 | 29 | 23 | 29 | 26 | 42 | 55 | 59 | 58 | 50 | 26 | 29 | 55 | 62 |
| 8:15 | 63 | 63 | 62 | 64 | 63 | 63 | 64 | 64 | 63 | 55 | 45 | 36 | 23 | 42 | 53 | 58 | 61 | 60 | 35 | 30 | 53 | 61 |
| 8:30 | 63 | 63 | 62 | 64 | 63 | 64 | 64 | 63 | 62 | 63 | 63 | 54 | 55 | 60 | 62 | 62 | 61 | 61 | 37 | 50 | 60 | 63 |
| 8:45 | 63 | 64 | 62 | 64 | 63 | 63 | 64 | 64 | 63 | 63 | 63 | 63 | 63 | 61 | 62 | 63 | 63 | 61 | 63 | 62 | 60 | 63 |
| 9:00 | 64 | 64 | 62 | 64 | 63 | 63 | 63 | 64 | 63 | 63 | 63 | 63 | 63 | 62 | 63 | 63 | 63 | 62 | 63 | 63 | 62 | 63 |
| 9:15 | 63 | 64 | 63 | 64 | 63 | 63 | 63 | 63 | 63 | 63 | 62 | 63 | 63 | 62 | 62 | 63 | 63 | 62 | 63 | 63 | 59 | 63 |
| 9:30 | 64 | 64 | 62 | 64 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 62 | 63 | 63 | 63 | 63 | 63 | 63 | 61 | 64 |
| 9:45 | 64 | 64 | 62 | 64 | 63 | 63 | 64 | 64 | 63 | 62 | 63 | 63 | 63 | 62 | 63 | 64 | 63 | 62 | 63 | 63 | 62 | 64 |
| 10:00 | 63 | 64 | 63 | 64 | 63 | 63 | 63 | 63 | 63 | 62 | 63 | 63 | 63 | 63 | 63 | 64 | 64 | 63 | 63 | 63 | 61 | 64 |
| 10:15 | 63 | 64 | 62 | 64 | 63 | 62 | 63 | 63 | 63 | 63 | 63 | 63 | 61 | 61 | 63 | 63 | 63 | 62 | 63 | 63 | 62 | 64 |

Figure 9.10: Contour plot after adjusting the waiting time.

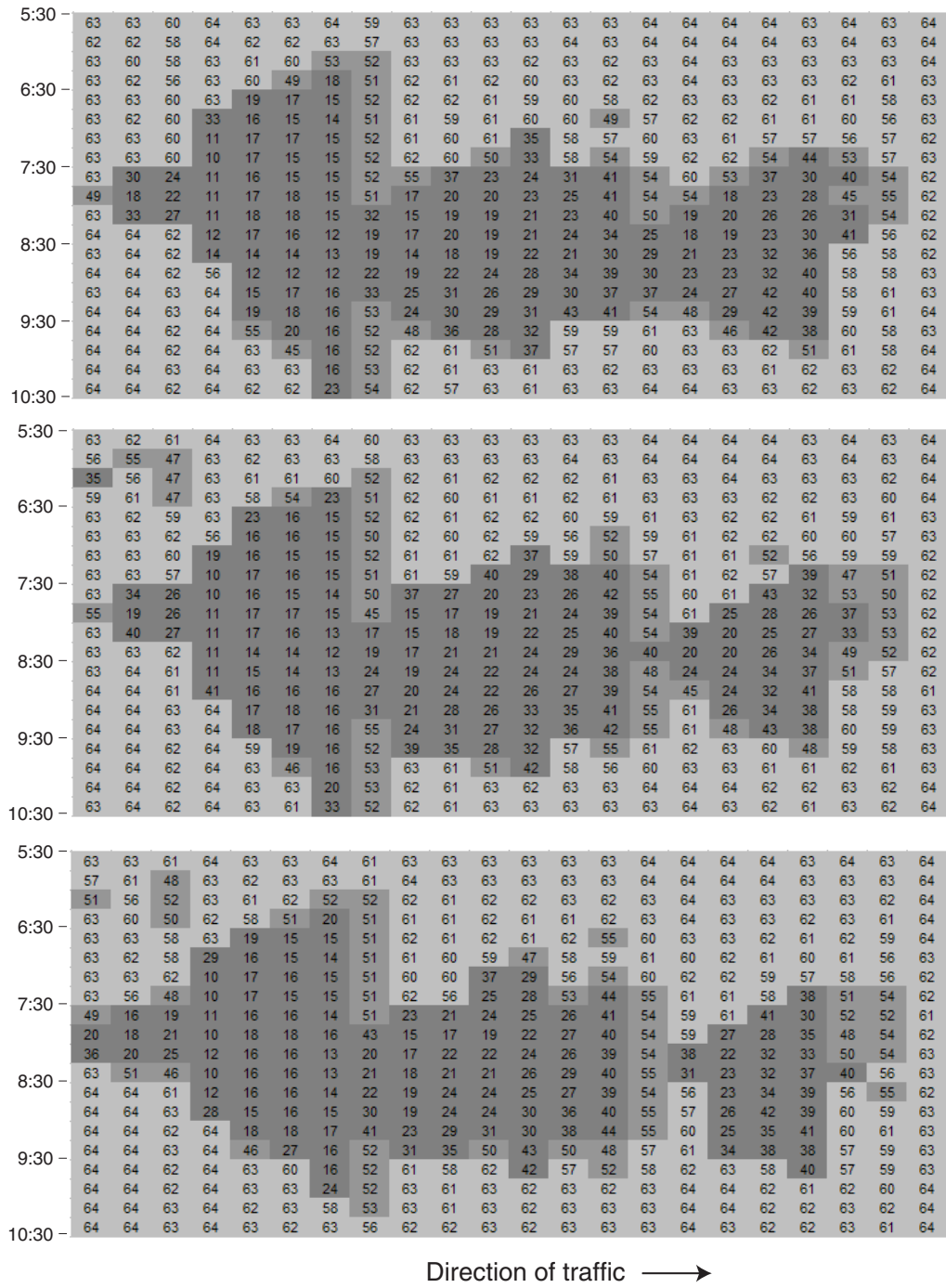## C.2 Random seed variations



Figure 9.11: Three different random seed values.

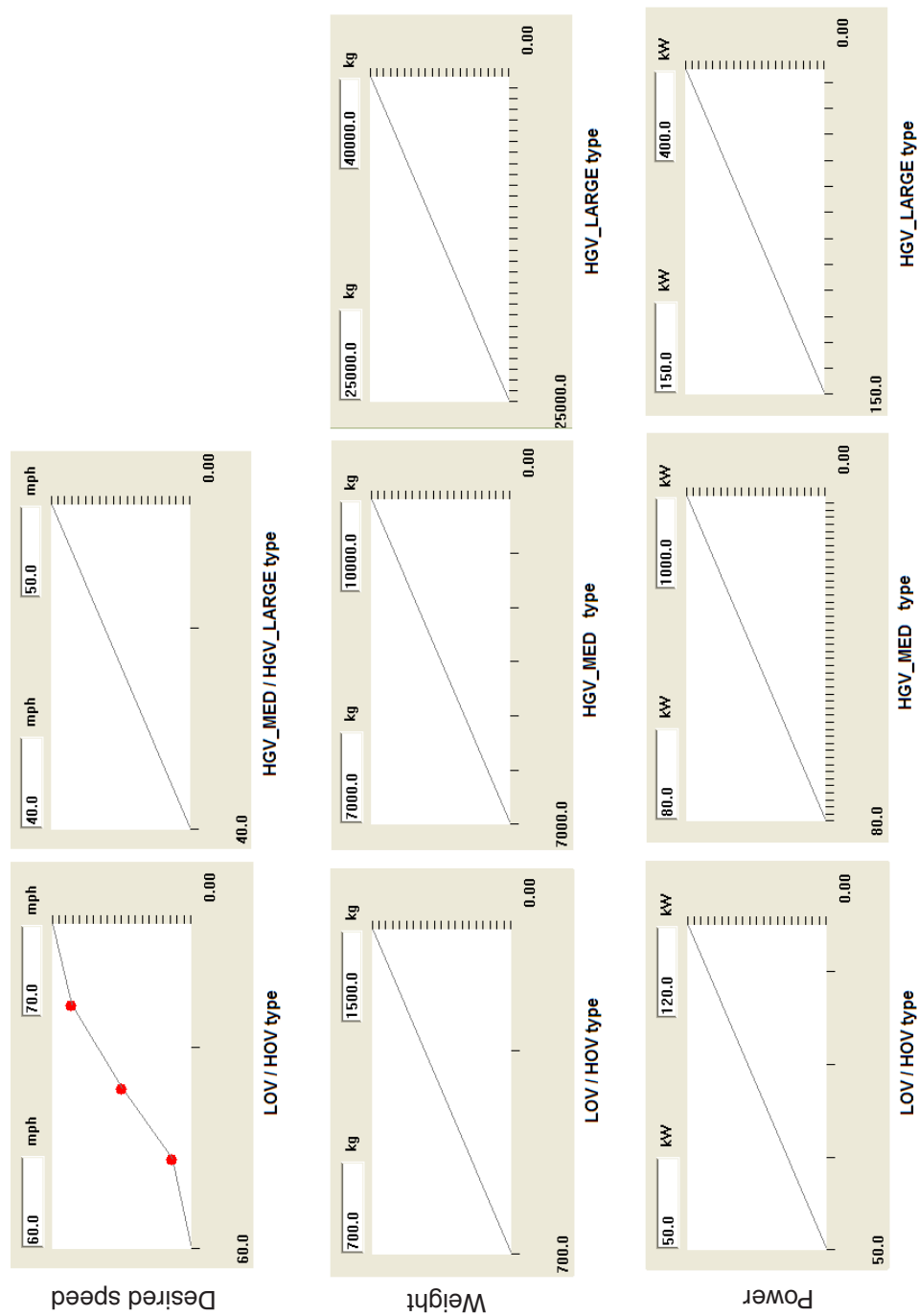## C.3    Vehicle characteristics



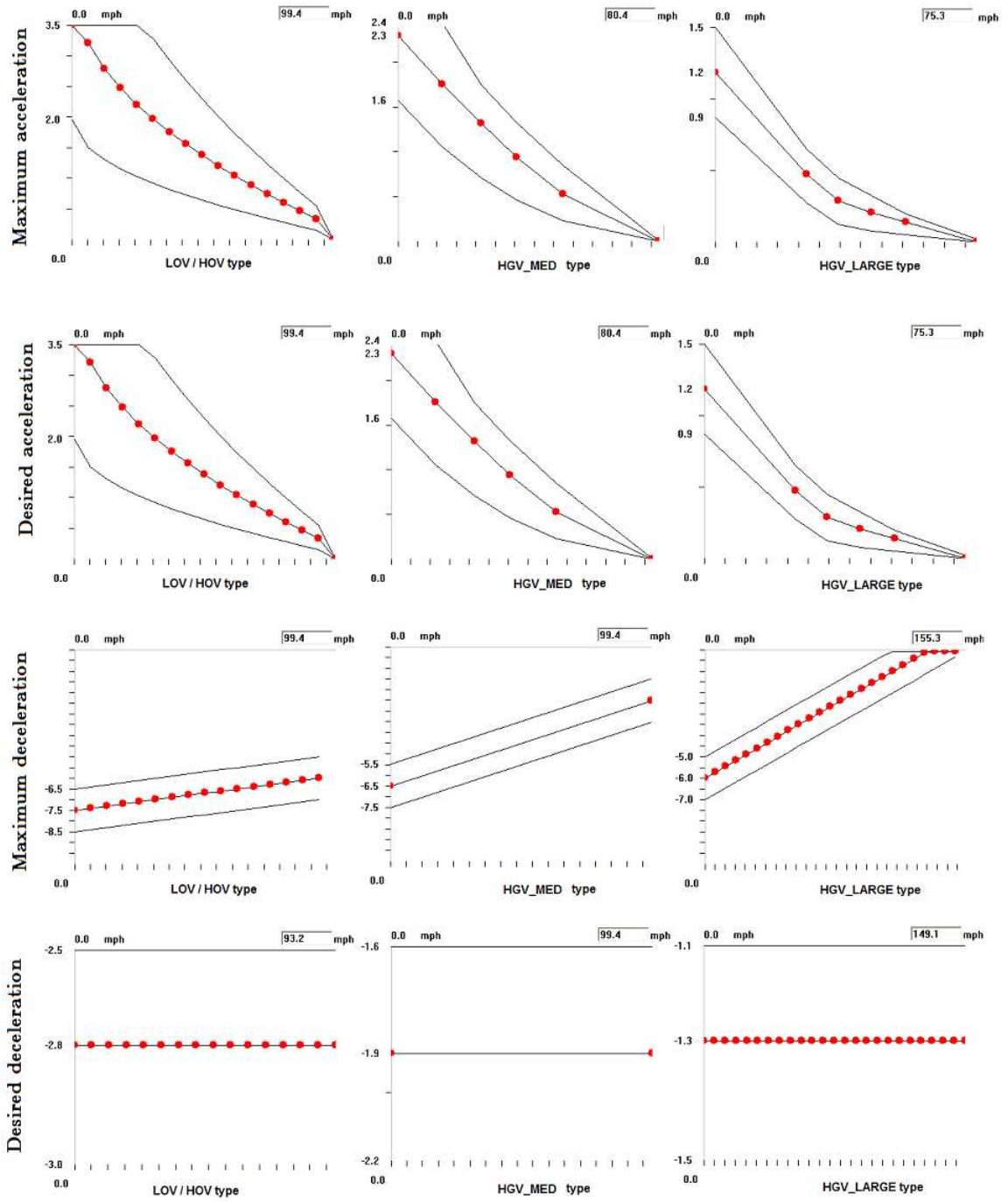Figure 9.12: Cars, small trucks, and large trucks - speed, weight, and power.

Figure 9.13: Cars, small trucks, and large trucks - acceleration and braking.
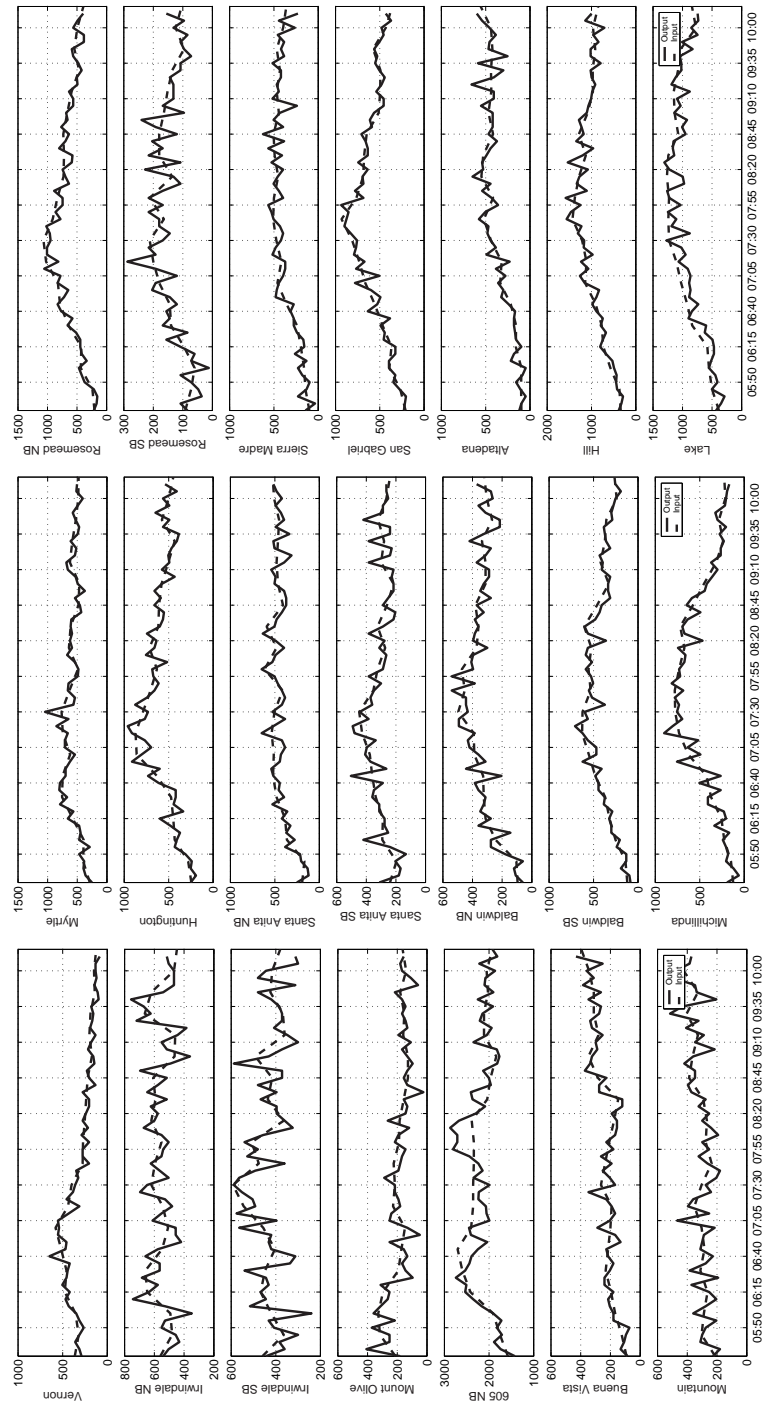
163

## C.4 VISSIM boundary flows



Figure 9.14: VISSIM input and output on-ramp flows.