Automatic Calibration of Freeway Models with Model-Based Sensor Fault Detection

by

Gunes Dervisoglu

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Roberto Horowitz, Chair
Professor J. Karl Hedrick
Professor Alexander Skabardonis

Fall 2012

The dissertation of Gunes Dervisoglu, titled Automatic Calibration of Freeway Models with Model-Based Sensor Fault Detection, is approved:

Chair _____ Date _____

_____ Date _____

_____ Date _____

University of California, Berkeley

Automatic Calibration of Freeway Models with Model-Based Sensor Fault Detection

Copyright 2012

by

Gunes Dervisoglu

Abstract

Automatic Calibration of Freeway Models with Model-Based Sensor Fault Detection

by

Gunes Dervisoglu

Doctor of Philosophy in Mechanical Engineering

University of California, Berkeley

Professor Roberto Horowitz, Chair

This dissertation presents system identification, fault detection and fault handling methodologies for automatically building calibrated models of freeway traffic flow. Using these methodologies, data driven algorithms were developed as part of a larger scheme of a suite of software tools designed to provide traffic engineers with a simulation platform where various traffic planning strategies can be analyzed. The algorithms that are presented work with loop detector data that are gathered from California freeways.

The system identification deploys a constrained linear regression analysis that estimates the so-called fundamental diagram relationship between flow and density at the location of a given sensor. A triangular fundamental diagram is assumed that establishes a bi-modally linear relationship between flow and density, the two modes being *free flow* and *conges-*

tion. An approximate quantile regression method is used for the estimation of the congested regime due to this mode's high susceptibility to various external factors.

The fault detection algorithm has been developed to facilitate the automatic model building procedure. The macroscopic cell transmission model, which is the model assumed in this study, requires consistent observations along the modeled freeway section for an accurate calibration to be possible. When detectors are down or missing, the model has to be modified to a less accurate representation to conform with a configuration where a sensor is assigned to each cell of the model. In addition, on most California freeways the ramp flows in and out of the mainline are not observed. Since the estimation of these unknown inputs to the system also hinge on healthy mainline data, the identification of faulty mainline sensors becomes crucial to the automatic model building process. The model-based fault detection algorithm presented herein analyzes the parity between simulated and measured state, along with estimated unknown input profiles. Subsequently, it makes use of a look-up table logic and a threshold scheme to flag erroneous detectors along the freeway mainline.

Finally, the fault handling algorithm that accompanies the fault detection aims to revert the model to its original configuration after the aforementioned modifications are made to the model due to missing or bad sensors. Using a relaxed model-constrained linear optimization, this algorithm seeks to fill in the gaps in the observations along the freeway that are a result of poor detection. This method provides a reconstruction of the unobserved state that conforms with the rest of the measurements and does not produce a state estimate in a control theoretical sense.

To my family,

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to thank the TOPL project members and our PI's, Professors Roberto Horowitz and Pravin Varaiya, for their guidance and support and for making this research possible. My advisor, Roberto Horowitz, has my utmost gratitude for giving me his unwavering support, through the good times and the worst, without which I would not have been able to complete my PhD study. My special thanks go to specific members of our group: Ajith Muralidharan, Rene Sanchez, Alex Kurzhanskiy and Gabriel Gomes, who had invaluable contributions to the realization of this work.

Next, I would like to thank my dissertation committee members, Professors Alex Skabardonis and Karl Hedrick for their patience and kindness. I am grateful to Professor Skabardonis for giving me the opportunity to work with him on sensor fault analysis, which translated into an essential part of this dissertation.

Finally, I would like to acknowledge the support of my friends and family, who had to endure all my mood swings over the last 6 years. I specifically thank my parents for their vision and for the way they raised me; my brother Derya for being the way he is, and doing the things I had to give up in the pursuit of this degree in my stead. And last but not least, I am grateful to have my wonderful wife Miray by my side, who is the person who had to make the most sacrifices for my dream to come true.

Chapter 1

INTRODUCTION

Traffic congestion in urban road networks is a major problem causing losses in time and energy and threatening public health and environmental safety. All these adverse effects can be quantified and regarded as monetary loss to states and municipal governing bodies. According to The 2011 Urban Mobility Report [47] generated by The Texas Transportation Institute, these costs accumulate to a nationwide total of over $100 Billion, which corresponds to approximately $750 yearly average cost per passenger. The report also points to the fact that traffic congestion has been a growing problem over the past decades. The yearly average delay per auto commuter over the 15 large urban areas in the United States that were included in the analysis increased from 19 hours in 1982 to 52 hours in 2010. The peak was observed in 2005 at 60 hours. Considering the ever increasing demand and the lagging nature of facility improvements, it can be conjectured that this trend is bound to continue.

As the problem grows, so does our arsenal of tools and strategies to tackle it. Obviously,

the most intuitive solution to the problem of excessive demand is to increase the supply, i.e. to enhance the present roadway infrastructure. This, however, is not a feasible solution for every given situation. In addition to physical facility expansion, a roadway structure can be made to operate more efficiently by the deployment of various strategies such as ramp metering control, speed limit control, demand management, etc. Software tools that model and simulate traffic flow become critical to the evaluation of such improvement strategies and analyzing their benefits and drawbacks before deploying them.

Traffic engineers and city planners rely on calibrated traffic corridor models and simulation software to devise mechanisms to improve road network performance. Two main approaches to traffic flow modeling can be classified as microscopic modeling and macroscopic modeling. Microscopic models of traffic model each individual vehicle as a separate entity and evaluate their interactions with other vehicles and the road. Macroscopic models treat the flow of traffic as a compressible fluid and evaluate the dynamic properties of the vehicles on a given road section as a coherent mass. In theory, microscopic models tend to be more accurate than their macroscopic counterparts, provided there exist measurements that facilitate their calibration. However, the data requirements for the calibration of microscopic models are often very demanding. Since the level of granularity for such models is the individual vehicle, a similar level of granularity and accuracy in the data is preferred. This can for instance be provided by video imaging of the road networks and the subsequent processing of these pictures. This, however, is not a feasible sensing mechanism to be implemented on a large scale. Therefore, most of the sensing on the roads in California are based

on inductive loop detectors embedded in the pavement, which are only capable of reporting aggregated point measurements of flow and occupancy. Moreover, even with adequate data present, the calibration process of microscopic models is a time and labor intensive task, which often includes manual tweaking of model parameters. In contrast, macroscopic models are very easy to calibrate and the data requirements are relatively more consistent with the readily available sensing infrastructure. Another major advantage of macroscopic models over the microscopic models is the simulation speed, which, due to the simplifying assumptions of the model, allows for a whole days' simulation to run within a few seconds. This leads to a more versatile deployment of macroscopic models in real-time applications such as live decision support for traffic operators.

The work presented in this dissertation has been carried out under TOPL (Tools for Operations Planning), a CalTrans and NSF funded project, launched in 2006 at PATH (Partners for Advanced Transportation Technology) and UC Berkeley. The goal of this project is to provide simulation tools to transportation engineers and researchers in order to assist them with operations planning. Traffic operations to alleviate congestion can be achieved through traffic control (e.g. ramp metering), through demand and incident management (e.g. incentivization and/or congestion pricing schemes), through traffic diversion to alternate routes, by providing information to travelers, etc. TOPL was launched to establish a software platform where computationally efficient calibrated models of traffic corridors could be used to qualitatively assess the benefits of such strategies.

The main contribution of this doctoral study to the TOPL project mentioned above is the

automation of the parameter estimation, fault detection and fault handling procedures that greatly facilitate the streamlined model building and calibration process. The questionable quality and, on occasion, the sparse availability of data has been a major challenge for the project and prior to the deployment of the presented work, the identification of faulty mainline detectors had to be carried out manually, which substantially increased the time and effort invested into building freeway models. With the addition of these algorithms, it is possible to automatically build and calibrate freeway models, provided the model geometry and the corresponding data are available to the software.

Thesis Outline

This thesis is constructed as follows: Chapter 2 explores the important concepts in the traffic sensing and modeling literature and introduces the main ideas in the established model-based fault detection framework. Preceding studies on fault detection in traffic sensing and modeling are also presented in this chapter. Chapter 3 first introduces the model building tools developed by the TOPL team. We then proceed to explaining the system identification and unknown input estimation algorithms along with the Link-Node Cell Transmission Model, the macroscopic model of choice for this study.

Chapter 4 presents the fault detection algorithm that is deployed by the TOPL suite of software tools to identify and eliminate faulty mainline detectors in an automated fashion. We first provide some motivating observations for the fault detection and then proceed to introducing the Asymmetric Cell Transmission Model on which the fault detection module is based. The next chapter demonstrates the application of the model building process on two

example freeways that have been modeled, namely: Interstate 80 in Northern California and Interstate 210 in Southern California. Finally, in chapter 6 we discuss the presented results and possible future work that can potentially extend upon the automated process presented herein.

Chapter 2

REVIEW OF RELATED LITERATURE

2.1 TRAFFIC MODELING

Traffic models in existing literature fall into two main categories: *macroscopic models* and *microscopic models*. In addition to strictly microscopic and macroscopic models, there also exist *mesoscopic models* that make use of the assumptions made by each paradigm to varying extents in an attempt to establish a specific trade-off level between model simplicity and accuracy. In essence, microscopic models of traffic flow are based on reproducing the behavior of individual vehicle units by replicating their interactions with other units and the road network. These are mostly built on car following models that adopt parameters such as headway, speed and position. Due to their detailed representation of the system, microscopic

models are not easy to calibrate and validate, especially with the available data collection infrastructure. Due to their increased complexity, simulators based on microscopic models are not suitable for large scale or real-time applications.

Macroscopic models, on the other hand, replicate the aggregate behavior of all vehicles on the road instead of focusing on individual units. In effect, macroscopic models treat the traffic flow as an incompressible fluid by specifying and evaluating quantities like speed, flow and density. Some important definitions regarding macroscopic traffic modeling are provided by the Highway Capacity Manual [7] as follows (x denotes the spatial and t denotes the temporal dependencies in the following):

- **Flow** $f(x,t)$**:** Flow is defined as the number of vehicles that pass a given cross-section on the road during a specified time interval divided by the length of that time interval. It is easily measured by point sensors on the road and is often given as an hourly rate.

- **Density** $\rho(x,t)$**:** Density is defined as the number of vehicles for a specific stretch of the road. Measurements of density are difficult to infer from point measurements and are usually approximated from flow and speed measurements as $\rho(x,t) = f(x,t)/v(x,t)$. Another approach to measuring density is to monitor the detector occupancy, which is defined as the portion of time during which the detection zone of the detector is occupied. This measurement is called occupancy and can be translated into the density measurement using equation (2.8).

- **Speed** $v(x,t)$**:** Speed is the average rate of motion of vehicles on the road. Depending

on the measurement, it can be expressed as a *space mean speed* and a *time mean speed*. Space mean speed is expressed as the average time spent by vehicles to traverse a certain length on the road. Time mean speed is the average speed of vehicles passing a given cross-section of the road, and as such, it is easier to capture using point measurements on the road.

- **Capacity:** Capacity is defined as the maximum rate of hourly vehicular flow through a certain section on the road under prevailing road conditions.

- **Bottleneck:** A bottleneck is a certain location on the road where the demand exceeds the prevailing capacity. Such locations usually trigger the onset of congestion that propagates upstream of the bottleneck location while the downstream of the bottleneck is typically in free flow. Bottlenecks can arise due to drops in number of lanes, merges of high volume onramps and certain topological features of the road such as grade and curvature.

- **Demand:** Demand is the number of vehicles that want to traverse a certain road section at a given time period.

Fundamental Diagram of Traffic Flow

Macroscopic models rely on constituting a relationship between the key parameters of flow, density and speed. These are usually captured in terms of the *Fundamental Diagrams*. This was first established by Greenshields in his seminal work on the Fundamental Diagram

relationships of traffic flow in 1934 [23]. He first approximated a linear relationship between speed and density based on his observations of captured images on a two-lane road. Based on this relationship, he expressed flow as a parabolic function of density using $\rho(x,t) = f(x,t)/v(x,t)$. A sketch of these relationships are shown in figure 2.1. In this figure, ρ^c denotes the critical density, where the maximum throughput is observed. Traffic flow is in congestion mode above this density value. Hence, all models of traffic flow are based on this bimodality of flow. ρ^J denotes the jam density for the road section, which is the maximum number of vehicles per mile that can be accommodated by that specific location. Note that both speed and flow are zero when this density is reached. The apex of this diagram corresponds to the section capacity and is attained at the critical density.

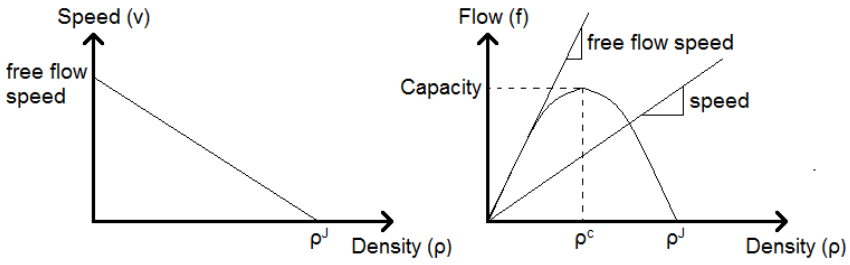

Figure 2.1: Greenshields' Fundamental Diagrams

In the existing literature, there are several different shapes assumed for the Fundamental Diagram, which essentially is an empirically determined constituent relationship between the key macroscopic elements. Thus, it can be established in several different ways, depending on what is chosen as the state of the model and on the chosen order of the governing differ-

ential/difference equations of the model. Models based on Speed-Flow [18], Speed-Density [30],[14] and Flow-Density [48],[20] have been established. The Flow-Density Fundamental Diagram was proposed as the model most suited for use in traffic control scenarios in [37], because it provides a direct relationship between flow, which would be the control input in a ramp metering exercise, and density, which would be the state to be controlled.

A challenging characteristic of Flow-Density diagram is the observed discontinuities near the capacity (maximum flow) of the road section, which gives rise to discussion on possible shapes of this fundamental diagram. This phenomenon was observed by Edie in [15] and in studies such as [20], [5]. Although there is widespread consensus that the Flow-Density curve is concave and the corresponding Speed-Density curve is monotone decreasing, different shapes for the theoretical curve are suggested: Triangular (or V-shaped), Reverse-λ shaped, Reverse parabolic and Trapezoidal Fundamental Diagrams have been suggested and evaluated in the literature. Figure 2.2 provides examples for each of these shapes. Works by Hall and Gunter [19] and Banks [4] support the V-shaped fundamental diagram. The argument made for the V-shaped fundamental diagram in [4] is that drivers obey the car following rule under dense and congested flow conditions but as their speed approaches the free flow speed, their sensitivity to vehicle spacing declines. Hence, as the density decreases, the speed of the vehicles increases at a higher rate which justifies the V-shaped fundamental diagram. On the other hand, microscopic models mostly agree on the Reverse-λ shaped Fundamental Diagram. [50], [33] and [17] are some examples of such studies. Contrary to the argument made for the V-shape, the following conjecture can be made for the Reverse-λ

claim: When the transition from congestion to free flow occurs, the vehicle speeds increase at a slower rate than the rate at which density decreases. Hence, the increase in flow lags the decrease in density because the downstream spacing supply is just recently available and it takes time for the vehicles queued upstream to speed up to the free flow speeds. This phenomenon was labeled *capacity drop* by Cassidy et al. in [9] and [31]. The transition from free flow to congestion occurs smoothly and the data points traverse through the apex of the fundamental diagram, whereas the transition from congested flow to free flow occurs more abruptly as a jump into the free flow line, undercutting the apex of the diagram. This phenomenon is also known as the hysteresis of traffic flow and is discussed in [27] and [51]. Current discussions on Macroscopic Fundamental Diagrams for Urban networks also consider this hysteresis [42].

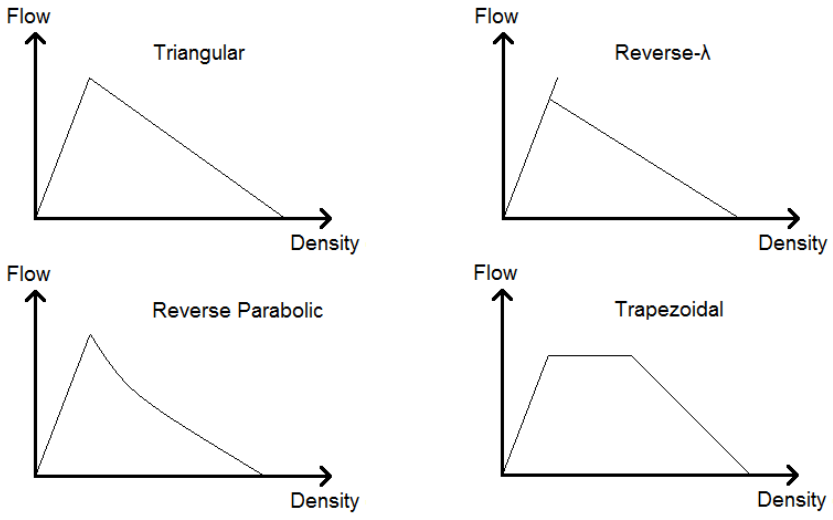

Figure 2.2: Various Flow-Density Fundamental Diagram Shapes

In our automated model calibration tool, we assume a V-shaped Fundamental Diagram due to its simplicity and ease of use with the Cell Transmission Model, which is the model adopted in this work. The V-shaped Fundamental Diagram is uniquely determined by 3 parameters: The free flow speed (the slope of the free flow line), the capacity (the maximum possible flow) and the congestion wave propagation speed (the slope of the congested regime line). The critical density and jam density parameters can be deduced from these three parameters.

The Cell Transmission Model

Macroscopic models are described by sets of partial differential equations that define the evolution of density, flow and speed over space and time. The Lighthill Whitham Richads (LWR) model [37], [46] constitutes the basis of the Cell Transmission Model (CTM), which is the basic model adopted in this study. All macroscopic models have at their core a vehicle conservation equation, similar to the mass conservation in fluid dynamics, and a fundamental diagram that relates density, flow and speed. In the case of the LWR model, these equations are given as:

$$\frac{\delta \rho(x,t)}{\delta t} + \frac{\delta f(x,t)}{\delta x} = 0 \tag{2.1}$$

$$\Phi(\rho(x,t)) = f(x,t) = \rho(x,t)v(x,t) \tag{2.2}$$

where equation (2.1) is the conservation equation and equation (2.2) specifies the fundamental diagram relationship. When this relationship is differentiable, the two equations can be combined as:

$$\frac{\delta\rho(x,t)}{\delta t} + \frac{d}{d\rho}\Phi(\rho(x,t))\frac{\delta\rho(x,t)}{\delta x} = 0 \tag{2.3}$$

First order models such as the LWR assume a static fundamental diagram for any given section of the road. Higher order macroscopic models complement the conservation equation with a so-called momentum equation, which are partial differential equations that describe the evolution of the space mean speed along the road. In effect, this aims to capture the complex dynamics of drivers' response to changes in the traffic state, rather than assuming those to be instantaneous as their first order counterparts do. For a more detailed treatise of higher order continuous time models, the reader is referred to the PhD thesis of Alex A. Kurzhanskiy [35], a fellow researcher in the TOPL project.

Since the main goal of the TOPL project is to create an automated and streamlined calibration and simulation tool that is applicable to large networks, a first order model is chosen, namely the Cell Transmission Model.

The Cell Transmission Model (CTM), introduced by Daganzo in 1994, is a discretization of the LWR model. The road network is represented as successive sections (cells) of uniform length L. Here, L is chosen to be such that $L = T_s v_f$ where T_s is the discretization time step and v_f is the free flow speed. With this formulation, the dynamics of density n_i of any given

cell i is described by the following difference equation:

$$n_i(k+1) = n_i(k) + f_{i-1}(k) - f_i(k) \tag{2.4}$$

where $f_{i-1}(k)$ is the flow entering cell i from the upstream cell i-1 and $f_i(k)$ is the flow leaving cell i and entering cell i+1 downstream at time k. These flows are determined by comparing the supply and demand of each cell dictated by their corresponding fundamental diagrams.

$$f_i(k) = min(Demand_i(k), Supply_{i+1}(k)) \tag{2.5}$$

$$Demand_i(k) = min(F_i, n_i(k)v_i) \tag{2.6}$$

$$Supply_{i+1} = min(F_{i+1}, w_{i+1}(n_i^J - n_i(k)) \tag{2.7}$$

Here, F_i, v_i and w_i are fundamental diagram parameters capacity, free flow speed and congestion wave propagation speed, respectively, for each section. A triangular fundamental diagram is assumed. Figure 2.3 provides a visual representation for this supply - demand relationship for determining exchanging flows between adjacent cells. In this temporal snapshot example, Cell 1 has a relatively low density and is therefore in free-flow. Cell 2 is congested but has enough room to accommodate the $Demand_1 = n_1(k)v_1$ from Cell 1. Hence $f_1(k) = Demand_1(k)$ in this case. Cell 2 wants to send $Demand_2 = F_2$ to Cell 3 since it is in congestion. Cell 3, however, is heavily congested and cannot accommodate $Demand_2$. As a result, $f_2(k) = Supply_3(k) = w_3(n_3^J - n_3(k))$ in this case. Looking at such a snapshot,

one would expect the density in Cell 2 to increase in the next time instance since flow into Cell 2 is greater than flow out of it at this time instance.

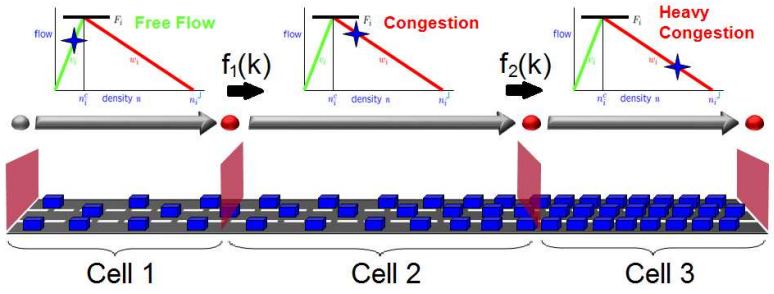

Figure 2.3: Cell Transmission Model send and receive relationship between successive cells

2.2 TRAFFIC SENSING

In this section, we explore the detection, reporting, analysis and archival tools available for gathering traffic data. These data can be used to calibrate simulation models, as in this study, or can be used in real-time by traffic operations planners for safety monitoring, incident detection and handling and providing information to travelers about travel times, accidents and other hazards.

Sensing Technologies There exist several sensing mechanisms deployed in California freeways. Here we review some of the most widely implemented technologies.

- *Inductive Loop Detectors:* These are inductive loops embedded into the pavement. Their operation principle is based on detecting induced eddy currents when a vehicle

passes over the loop. Loop detectors provide point measurements of flow and occupancy, and when installed adjacently, they also provide speed measurements. They are the most common sensing equipment implemented on California roadways. While they are used to measure flow, occupancy and speed on freeways, on arterials, they are mostly used to detect presence only, for queue responsive signalized intersections. The data they provide is easily archived and reported, which makes them very suitable for the automatic model building tools to be introduced in this dissertation. However, their installation and maintenance costs are relatively high and due to their low frequency of maintenance the data they provide are not very reliable. Nevertheless, when maintained appropriately, loop detectors provide the best accuracy in terms of counts, i.e. flows.

- *Magnetic Sensors:* Similar to the inductive loop detectors, magnetometers detect presence by evaluating the magnetic fields caused by passing vehicles. These also need to be embedded in the pavement but they provide the additional advantages of easier maintenance, longer life times and some models are equipped with wireless transmission capability.

- *Video Image Processors:* Video camera surveillance of roadway traffic is often used to transmit imagery to human operators to help them with real-time management. In addition, some stations are equipped with video image processors that digitize and analyze the imagery. Software tools are available for converting these into the desired

parameters like flow and density. Videos provide the highest quality and highest granularity data but come with a price tag of data complexity and size. Moreover, they are much more susceptible to changing weather and visibility conditions, especially at night time. Moreover, their maintenance requirements are higher due to the necessity of frequent cleaning of the camera lenses.

- *Microwave and Laser Radars:* Contrary to the passive detection technologies listed above, radars are active sensors that work on the principle of sending electro-magnetic waves to moving vehicles and and track their presence and movement through the reflection of these electro-magnetic waves. These have the advantage of not being affected by weather conditions over video surveillance. While they provide direct measurements of speed, additional analysis is required to infer flow and density from radar measurements.

Other sensor technologies include ultrasonic sensors, passive acoustic sensors and infrared sensors. Toll tag readers, license plate readers and probe vehicles can also be used for travel time measurements. For a comprehensive report on traffic sensing technologies the reader is referred to the Traffic Detector Handbook [36].

Measured Traffic Elements

The sensing technologies listed above are used to gather the following measurements related to traffic on road networks:

- *Occupancy:* Occupancy is the percentage of time the detection zone of a detector is

occupied by vehicles.

- *Flow:* Flow is the number of vehicles passing through the detection zone of a detector within a specified time interval.

- *Speed:* Depending on the detection method, space mean speeds and time mean speeds can be reported. Point sensors such as loop detectors and magnetic sensors usually estimate time mean speeds whereas detectors that have wider detection zones, such as video cameras, can be used to estimate space mean speeds. Use of multiple point sensors such as double loops provide accurate time mean speed measurements.

- *Travel Time:* Travel times can be estimated from re-identifying vehicles at two detection stations. Toll tag readers, license plate readers and video cameras have this capability. In addition, vehicle re-identification through magnetic signatures over shorter distances have been demonstrated to work using magnetic sensors [32]. These measurements can also be used to determine Origin-Destination data, which yields demand patterns necessary for demand management strategies.

Performance Measurement Systems (PeMS)

The automatic model calibration tools presented in this dissertation are based almost exclusively on data provided by the PeMS (Performance Measurement Systems) archival tool [2]. PeMS was launched in 1998 as a collaborative project of UC Berkeley, PATH and Caltrans and the software tools have been available online since 2003 for use by traffic engi-

neers and academics. PeMS collects data from over 25000 detectors on California freeways and arterials and has several tools to filter, process, aggregate and examine these data. Loop detectors are the main contributors to this extensive archive but PeMS also gathers data from sources like toll tag readers and CHP (California Highway Patrol) incident reports. It also provides reporting tools for a wide variety of statistical analyses from individual detector health to daily flow trends on a specific stretch of road. Based on loop detector data, PeMS reporting tools calculate performance measures such as VMT (Vehicle Miles Traveled), VHT (Vehicle Hours Traveled), Congestion Delay, Productivity Loss, etc. for a user-specified period of time and stretch of road.

For the purposes of this study, loop detector data from California freeways are of crucial importance. In fact, as will be explained in detail in section 3.1, the automatic model building tool relies on the detector locations on the freeway to determine the model geometry, i.e. the detector locations are used as a starting point when specifying the model. In their raw form, there are two components to loop detector data: 1) Flow 2) Occupancy. These are collected at 30s granularity and PeMS subsequently aggregates them into 5 minute intervals. PeMS also infers speed and density based on the occupancy and flow measurements following the formulae below:

$$Speed = \frac{Flow}{Occupancy}\text{g-factor} \qquad Density = \frac{Flow}{Speed} \tag{2.8}$$

where the g-factor is an adaptively estimated time-dependent effective vehicle length specific to the detector and the location [29].

In spite of this voluminous data collection and archival effort, detector and data health remains to be a major problem. Figure 2.4, taken from the PeMS website, shows the percentages of working vs non-working detectors over the 1-year period between August 2011 and August 2012. In the vertical axis are the Districts of The State of California and it can be seen that, on average, only about 70% of detectors were reported to be working correctly. It is also noteworthy that District 4, containing the larger Bay Area, has had an average detector health percentage of a dismal sub-50% value over the most recent annual period.

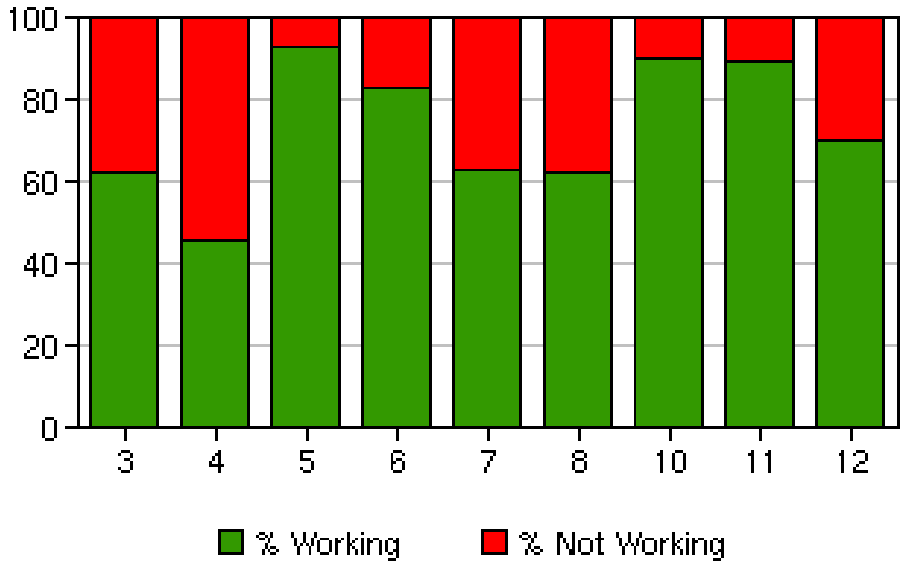

Figure 2.4: Detector Health throughout the Districts of California August 2011 - August 2012

PeMS implements some simple screening methods to determine detector health. It flags

a detector to be "not working" if the detector sends few or no samples. There are also set thresholds that are triggered when the observed values are too high to be realistic. In addition to these simple screens, the repository deploys other fault diagnosis methodologies which will be explored in the next section.

Another functionality PeMS provides is the imputation of missing measurements based on neighboring detectors and/or historical averages. This way, it is ensured that sparse missing chunks in the data do not affect the overall analysis. The imputation algorithm used by PeMS is based on a linear regression on the large historical database that is available. To predict missing data on freeway mainlines, this algorithm uses data from detectors on adjacent lanes and detectors that are immediately upstream and downstream of the erroneous detector [8].

2.3 FAULT DETECTION

Fault Diagnosis is a relatively new subfield of control theory that grew in the last several decades, parallel to the demands for higher system performance and better quality assurance along with increasing system complexity in industrial applications. In their 1997 survey of fault detection and isolation methods [44], Isermann and Balle provide the following definitions in order to establish a basic terminology for the field:

- *Fault:* An unpermitted deviation of at least one characteristic property or parameter of the system from the acceptable / usual / standard condition.

- *Fault Detection:* Determination of the faults present in a system and the time of detection.
- *Fault Isolation:* Determination of the kind, location and time of detection of a fault.
- *Fault Identification:* Determination of the size and time-variant behavior of a fault.
- *Fault Diagnosis:* Determination of the kind, size, location and time of detection of a fault.

Fault diagnosis has its historical roots in the implementation of redundant hardware components. If two or more system components are assigned the exact same task, any disparity between them is enough to conclude a fault event. Although still an important scheme, especially for safety critical applications, a more efficient and cost-effective approach has been to implement process models and quantitative analysis tools to tackle the issue. Current fault diagnosis methods can be classified into two major branches: 1) Signal processing-based approaches, 2) Model-based approaches.

Signal processing based methods of fault diagnosis are based on the assumption that some process signals contain information about faults. By analyzing the magnitudes or other statistical metrics of the signals in the time domain or by evaluating frequency domain functions like power spectral densities, one may be able to capture statistically aberrant measurements that are indicative of process or sensor faults. Especially in cases where multiple sensors are used to monitor the system, such as distributed sensor networks, for times long enough to

establish temporal and spatial correlations over time, time-series analysis tools become suitable along with hypothesis testing schemes that are designed to minimize the probability of false alarm.

Model-based fault diagnosis methods, on the other hand, make use of an analytical model of the underlying process to generate residuals between the model and the real system. Statistical methods can nonetheless be employed while evaluating these residuals. The general idea is depicted in figure 2.5 (figure adapted from [13]). Model-based methods of fault diagnosis are usually classified under three frameworks [13], [25]:

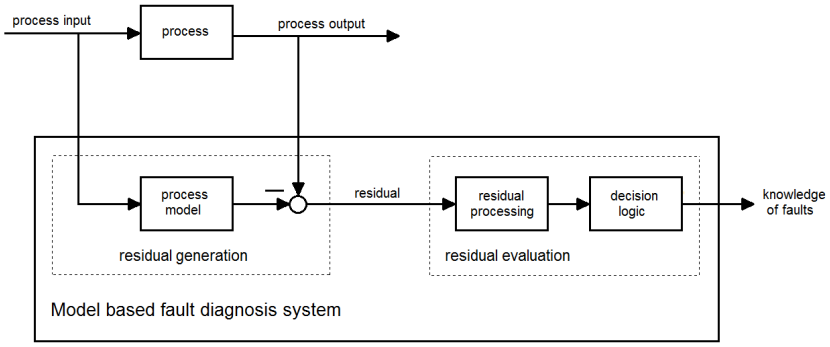

Figure 2.5: Model-based Fault Diagnosis Framework

1. **Observer-based Methods:** The main goal of this approach is to design an observer such that the resulting residuals have desirable features such as decoupling the faults from unknown inputs and disturbances along with fault detectability and isolability. These, of course, are mainly dependent on the system matrices but the observer design procedure may exploit known properties about the system components to produce the

most beneficial residuals for fault diagnosis. Figure 2.6 together with equations (2.9) - (2.11) gives the block diagram schematic of this approach (fault, noise and disturbance signals are not shown in the figure for simplicity - adapted from [25]).

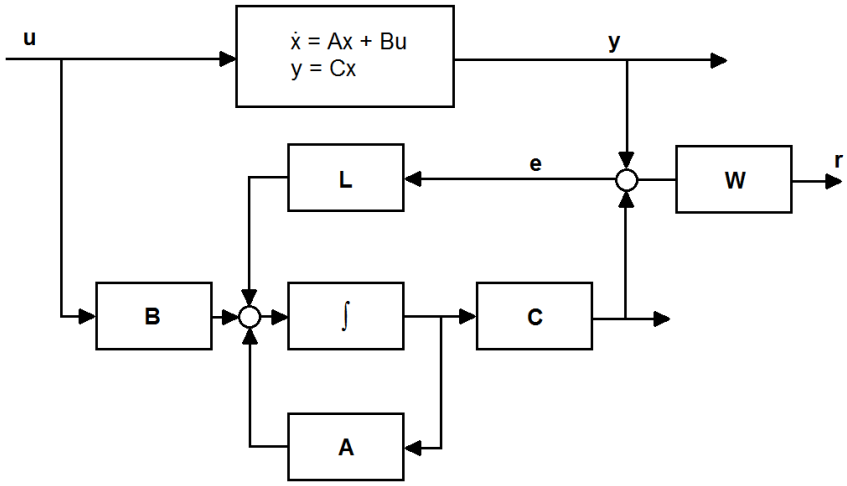

Figure 2.6: Observer-based Fault Diagnosis approach - Block Diagram

Plant equations:

$$\dot{x}(t) = Ax(t) + Bu(t) + E_d v(t) + F_s f_s(t)$$

$$y(t) = Cx(t) + E_n n(t) + F_o f_o(t) \tag{2.9}$$

Observer equations:

$$\dot{\hat{x}} = A\hat{x}(t) + Bu(t) + Le(t)$$

$$e(t) = y(t) - C\hat{x}(t) \tag{2.10}$$

Candidate Residuals:

$$\Delta x(t) = x(t) - x_0(t)\ , \qquad e(t)\ , \qquad r(t) = We(t) \tag{2.11}$$

2. **Parity Equations:** Also called the parity space approach, the use of parity equations for fault detection is a more straightforward approach where a fixed model is run in parallel with the process to generate residuals. These residuals are then evaluated, along with some threshold schemes to account for model uncertainties, to identify faults occurring in the process or the measurements. This approach is easier to demonstrate in discrete time.

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + E_d v(k) + F_s f_s(k) \\ y(k) &= Cx(k) + Du(k) + E_n n(k) + F_o f_o(k) \end{aligned} \tag{2.12}$$

Given a system in equation (2.12) with (C,A) observable and C full row-rank, we fix a time instance s time steps back in time and build the augmented matrices that map the input sequence to the output sequence from time $k-s$ to time k, in the fault- and disturbance-free case. In the system above, $x(k)$ denotes the state, $u(k)$ is the input signal, $v(k)$ is an unknown disturbance, $f_s(k)$ is the process fault, $n(k)$ is measurement

noise and $f_o(k)$ is the sensor fault. Defining:

$$y_s(k) = \begin{bmatrix} y(k-s) \\ y(k-s+1) \\ \vdots \\ y(k) \end{bmatrix} \quad , \qquad u_s(k) = \begin{bmatrix} u(k-s) \\ u(k-s+1) \\ \vdots \\ u(k) \end{bmatrix} \tag{2.13}$$

$$A_P = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^s \end{bmatrix} \quad , \qquad B_P = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-1}B & \cdots & CB & D \end{bmatrix} \tag{2.14}$$

we get the relationship:

$$y_s(k) = A_P x(k-s) + B_P u_s(k) \tag{2.15}$$

In equation (2.15), the only unknown is the state at time $k-s$, $x(k-s)$. If s is chosen such that $s >= n$, with the observability assumption, the matrix A_P becomes full column-rank with a non-empty left null-space. This left null-space constitutes the parity space as given in equation (2.16).

$$P_s = \{v_s | v_s A_p = 0\} \tag{2.16}$$

Using any row-vector from this left null-space, the unknown state $x(k-s)$ can be eliminated to produce the residual, which in the fault- and disturbance-free case is 0,

as expected.

$$r_s(k) = v_s(y_s(k) - B_P u_s(k)) = v_s A_P x(k-s) = 0 \tag{2.17}$$

When we also consider the faults and disturbances, building matrices E_{s_P}, F_{s_P}, E_{o_P} and F_{o_P} in a similar fashion to equation (2.14), we get the residual equation:

$$r_s(k) = v_s(E_{s_P} d_s(k) + E_{o_P} n_s(k) + F_{s_P} f_{s_s}(k) + F_{o_P} f_{o_s}(k) \tag{2.18}$$

The choice of the parity vector v has a significant effect on fault diagnosis performance and can be chosen to cancel one or more of the disturbances if any prior information is available on the nature of interactions between unknown signals. The parity approach is favorable due to its simplicity, especially for discrete time systems that do not have to be monitored in real time. For this reason, we adopt this approach in this study as will be shown in section 4.3. Figure 2.7 provides a simplified block diagram for the parity space approach (fault, noise and disturbance signals are not shown in the figure for simplicity - adapted from [25]).

3. **Parameter Estimation Methods:** In the parameter estimation-based methods, the fault diagnosis is performed online with the assumption that faults occurring in the process or the sensors affect the system parameters. Figure 2.8 (adapted from [13]) shows the main idea in block diagram form. This approach requires some nominal values for the unknown system parameters to distinguish deviant estimates from correct ones. Isolation and identification of faults is more difficult for this method than it is for the other two methods that have been introduced above.

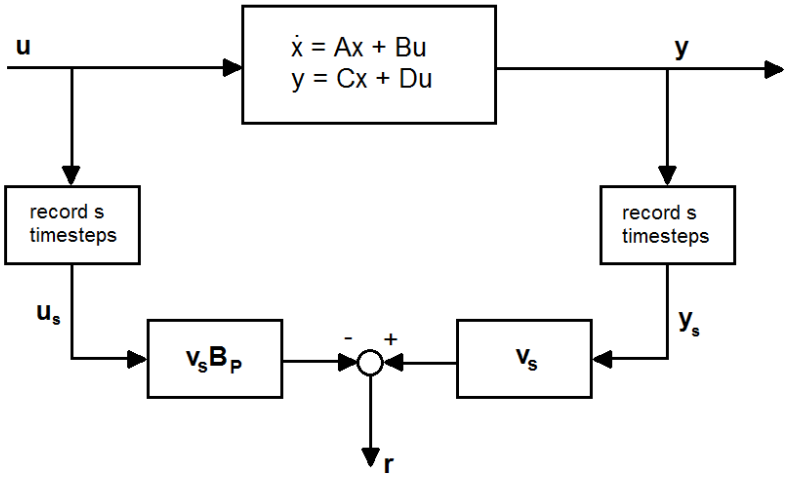

Figure 2.7: Parity-based Fault Diagnosis approach - Block Diagram

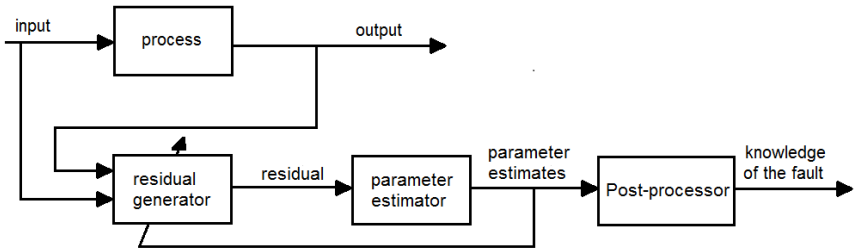

Figure 2.8: Fault Diagnosis with Parameter Estimation

Fault Detection in Traffic Systems

To the author's knowledge, fault diagnosis literature related to the sensing of vehicular traffic, especially using model-based approaches, is quite limited. Since the modeling and calibration effort presented in this dissertation is based on loop detector measurements, our

focus is on the detection of loop detector malfunction. Most of the previous work on the subject deploys statistical methods to detect faults.

The earliest works go back to early 70's, which coincides with the wide spread of loop detectors as the primary traffic sensing tool, and their main principle is to establish data reliability tests based on expected behavior and interrelation of occupancy and flow measurements. The theoretical foundations for these data reliability criteria were provided by Payne et. al back in 1976 [43], where they identified 5 types of detector errors and introduced upper and lower thresholds on flow, density and speed that can be reasonably assumed. Aside from heuristic "on the field" decisions made by operating agencies or detector installation and maintenance personnel to make sure the detector "works", the first signal processing based fault detection method that was deployed was the Washington Algorithm [28]. Developed by Jacobson et. al in 1990, this algorithm defines an acceptable Flow-Occupancy region and flags individual data points as unreliable if they fall outside this region. In 1991, Cleghorn et. al [11] introduced a tighter region of acceptability, with additional constraints for dual loops, increasing in effect the dimensions of the acceptable region to 3 by the addition of speed. In 2000, Turochy et. al [49] added new critical thresholds on top of the existing test on relationships between speed, flow and occupancy to better identify data points that were reporting unreasonably high values. PeMS uses the scheme devised by Chen et. al [8] in 2003, that introduces the first time-series methods to determine erroneous detectors. Instead of evaluating individual data points, they detect irregularities in the time series of a measurement taken during a day. They classify 4 such irregularities:

1. *Zero Occupancy and Flow for an extended period*

2. *Non-zero occupancy accompanied by zero flow*

3. *Very high occupancy*

4. *Constant occupancy and flow*

If a detector reports measurements that fall into one of the categories above, that detector is declared to be unreliable for that whole day and the data it provides are discarded. To replace this chunk of missing data, Chen et. al introduce an imputation algorithm that uses time-series data provided by neighboring detectors over the last 5 days (and reliable data provided by the same detector, if any, in the past 5 days) to replace these unreliable measurements. Improved statistical methods building upon this approach are provided by Kwon et. al [26] and by Rajagopal et. al [45].

To the author's knowledge, the only model-based approach to traffic sensor fault detection in the literature apart from the work presented herein is the recent work of Claudel and Bayen [10] where they model the freeway flow by Hamilton-Jacobi partial differential equations and test the consistency of data collected at certain locations along the freeway against the continuous model in the space-time domain. However, they ignore the effect of unknown ramp flow inputs into the system.

As far as loop detector data are concerned, the major advantage the model-based schemes have over statistical methods is the fact that some detectors are observed to carry consistent

offsets rather than or in addition to noise or other statistically discernable anomalies. Signal processing methods based on historical data may not be able to correctly pinpoint such anomalies if they are consistent throughout the data collection history. We will explore these phenomena in more depth in section 4.1

Chapter 3

AUTOMATIC FREEWAY MODEL CALIBRATION PROCEDURE

3.1 FREEWAY MODEL SPECIFICATION AND THE LINK-NODE CELL TRANSMISSION MODEL

The automated model building procedure for freeways can be summarized by the process flow diagram in figure 3.1. The first step is to specify the stretch of freeway section to be modeled. TOPL has a dedicated web-based software tool, named *Network Editor* [3], that implements *Google Maps*. This tool provides a user-friendly visual platform to build freeway or broader network geometries directly on the road network map. The user can also download

PeMS detector information and superpose it on the map. Figure 3.2 shows a snapshot of a specified geometry for a portion of northbound Interstate 15 near San Bernardino in Southern California. The network is specified by placing nodes at locations where traffic flow exchange occurs and where significant changes to mainline geometry exist, such as lane drops. However, it is customary to build the network such that all available detectors get associated with a separate link. Hence, the presence of detectors and their locations substantially dictate model geometry specification. The nodes are then connected by links and the number of lanes are specified for each link. Subsequently, sensors that are superposed on the map at their corresponding postmiles need to be associated with the according links by the user.

The Link-Node Cell Transmission Model

The macroscopic traffic flow model adopted by the TOPL project is the so-called *Link-Node Cell Transmission Model (LN-CTM)*, which is an extension of the Cell Transmission Model introduced in section 2.1. The LN-CTM was introduced in [35] which details the simulation tools developed by Alex Kurzhanskiy within the scope of the TOPL project.

LN-CTM represents the traffic network as a directed graph of links joined by nodes. Each link in the network is assumed to have uniform geometry and fundamental diagram parameters, the identification of which will be explained in the next section. Links can be classified as mainline links, sources and sinks. Source links are links that start with a source node and end at a mainline node. These sources specify the input boundaries to the network. Similarly, sink links start with a mainline node and end at a sink node, and they specify

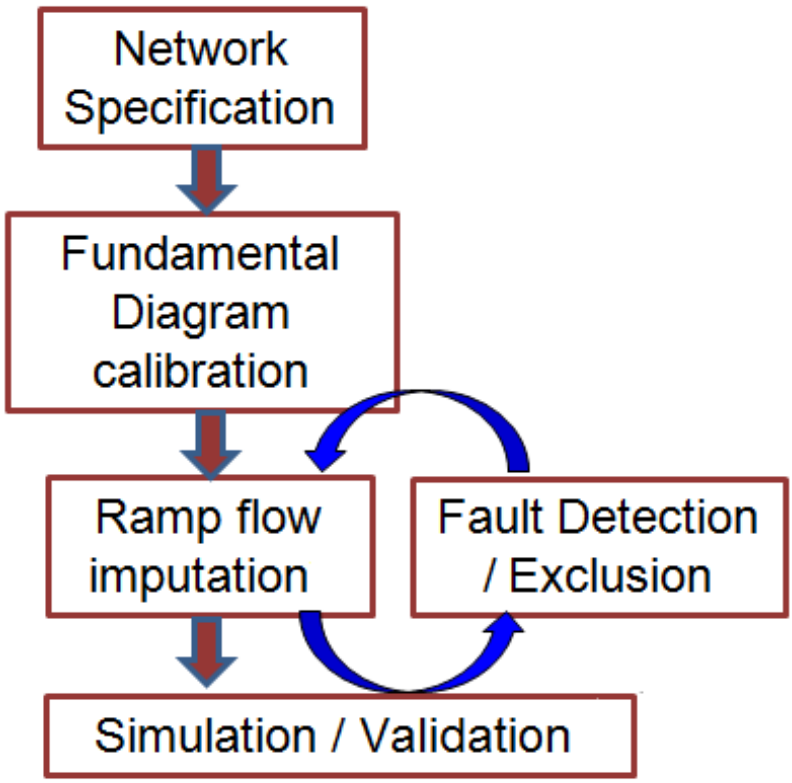

Figure 3.1: Model Building Process Flow Diagram

the output boundaries of the network. Traffic exchange at nodes are specified by means of a time-varying split ratio matrix which scales and assigns the incoming flows to outgoing flows, corresponding to the evaluation of the model. Nodes do not store any vehicles and flows are conserved, hence the entries of the split ratio matrix corresponding to each link add to 1. Freeway nodes are allowed to have at most one on-ramp and one off-ramp. Hence, it may be the case where two on-ramp merges are very close to each other, they may need to be bundled together and represented as a single on-ramp. Another restriction is on the link lengths. To ensure that the CFL conditions are satisfied by the discrete time simulation,

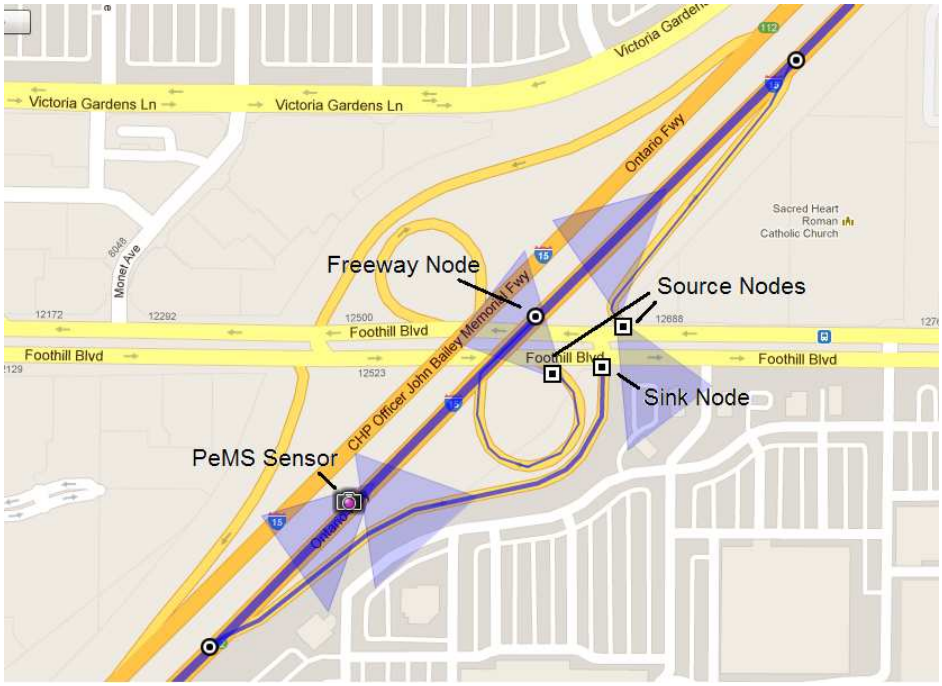

Figure 3.2: A Network Editor Snapshot

the link lengths should be specified such that they can not be traversed within a single time step of the simulation, i.e. $vT < L_i$ and $wT < L_i$ must hold for each link i where L_i is the length of link i, T is the simulation time step and v and w are the free flow and congestion wave speeds, respectively.

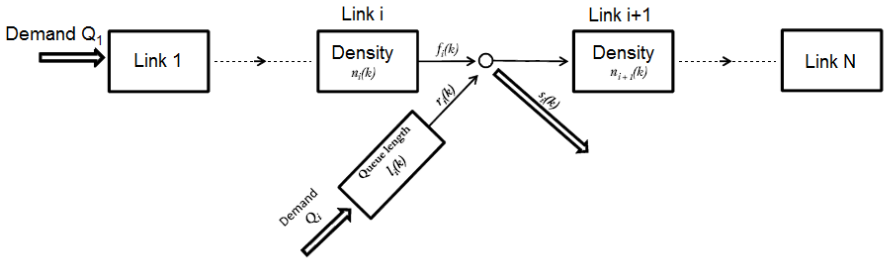

Figure 3.3: Graphical representation of the LN-CTM

The LN-CTM replicates the evolution of flow and density by a two-layered set of governing difference equations. A conservation equation governs the density updates using flows entering and leaving the links and the flow updates are governed by the supply-demand dependent send-and-receive relationship between adjacent links, in accordance with the Cell Transmission Modeling framework introduced in section 2.1. If at any exchange location the demand exceeds the supply, the flows are scaled down to match the minimum downstream supply and the excess demands contribute to density build-up in the upstream links. An additional conservation equation is introduced to evaluate the queues on on-ramps and the off-ramps are assumed to have infinite capacity, i.e. they are perfect sinks that do not get congested. This assumption may not always be accurate, in which case the model should be extended as suggested in [38]. For a network represented in Link - Node formulation as in figure 3.3 the flow and density update equations are as follows:

Density Update:

$$
\begin{aligned}
& n_0(k+1) = n_0(k) + Q_0(k) - f_0(k) \\
& n_i(k+1) = n_i(k) + f_{i-1}(k)(1 - \beta_{i-1}(k)) + r_{i-1}(k) - f_i(k) \qquad i = 1, \cdots, N \\
& l_i(k+1) = l_i(k) + Q_i(k) - r_i(k) \qquad i = 1, \cdots, N
\end{aligned} \tag{3.1}
$$

Flow Update:

$$
\begin{aligned}
& f_N(k) = D_n(k) \\
& f_i(k) = D_i(k) \times \frac{\min(R_i(k), S_{i+1}(k))}{R_i(k)} && i = 0, \cdots, N-1 \\
& r_i(k) = d_i(k) \times \frac{\min(R_i(k), S_{i+1}(k))}{R_i(k)} && i = 1, \cdots, N \\
& s_i(k) = f_i(k)(1 - \beta_i(k)) && i = 1, \cdots, N \\
& \textit{where} \\
& D_i(k) = \min(n_i(k)V_i, F_i), \\
& R_i(k) = D_i(k)(1 - \beta_i(k)) + d_i(k), \\
& S_{i+1}(k) = \min(W_{i+1}(n^J_{i+1} - n_{i+1}(k)), F_{i+1}) \\
& d_i(k) = \min\left(r^c_i(k), l_i(k), C_i\right)
\end{aligned} \tag{3.2}
$$

where Table 3.1 lists all the variables and parameters in the above equations.

When modeling a freeway with the LN-CTM, the convention adopted by the TOPL model building procedure is to collapse the stretch of the freeway between an off-ramp and the immediately following downstream on-ramp to a single node and specify the split ratio matrix such that flows from the incoming on-ramp to this node are not permitted into the outgoing off-ramp. Incidentally, most of the detectors in California freeways are located in this stretch. An example is shown in figure 3.4. The length of the upstream link (link i in the case of figure 3.4) is re-adjusted to include the length of the collapsed region. Hence, in effect, the off-ramp is being pushed downstream per assumption of the model geometry to

Symbol	Name	Unit
F_i	flow capacity of Link i	veh/period
v_i	free flow speed of Link i	section/period
w_i	congestion wave speed of Link i	section/period
n_i^J	jam density of Link i	veh/section
k	simulation time step (period)	dimensionless
$\beta_i(k)$	split ratio at node i	dimensionless
$f_i(k)$	flow out of Link i	veh/period
$n_i(k)$	number of vehicles (vehicle density) in Link i	veh/period
$s_i(k), r_i(k)$	off-ramp, on-ramp flow in node i	veh/period
$d_i(k)$	on-ramp i demand	veh/period
$l_i(k)$	queue length on on-ramp i	veh/period
C_i	flow capacity for on-ramp i	veh/period
L_i	queue capacity for on-ramp i	veh/period
$Q_i(k)$	input flow for on-ramp i	veh/period

Table 3.1: LN-CTM Variables and Parameters

be attached to the freeway at the same location as the on-ramp.

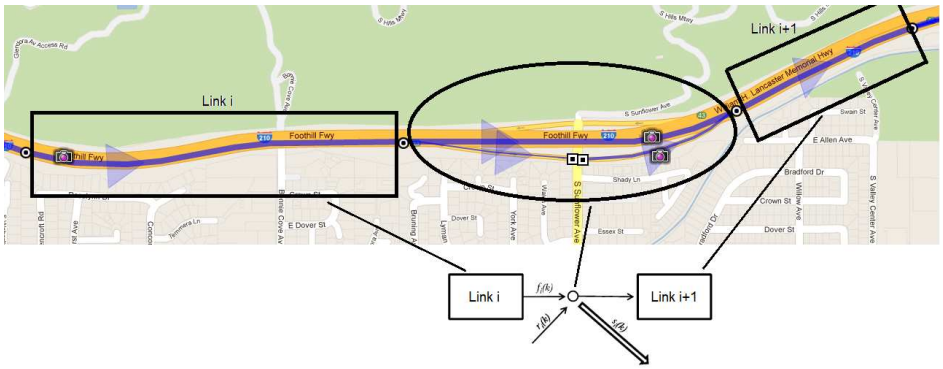

Figure 3.4: Link - Node representation of a freeway ramp merge location

3.2 FUNDAMENTAL DIAGRAM CALIBRATION

For both extensions of the Cell Transmission Model that are implemented by TOPL (LN-CTM introduced in section 3.1 and ACTM, to be introduced in section 4.2), and in fact, for

any macroscopic model of traffic flow, the identification of fundamental diagram parameters for each cell in the network needs to be performed to obtain an accurate model. Exploiting the abundance of Flow-Density data provided by PeMS, an automated fundamental diagram calibration methodology has been devised, which uses multiple days' data to calibrate a fundamental diagram for a section that includes a loop detector. The main goal of this algorithm is to fit the assumed triangular fundamental diagram relation to the Flow-Density scatter plot as shown in figure 3.5. In the scatter plot, every data point corresponds to a 5 minute aggregated Flow-Density pair.

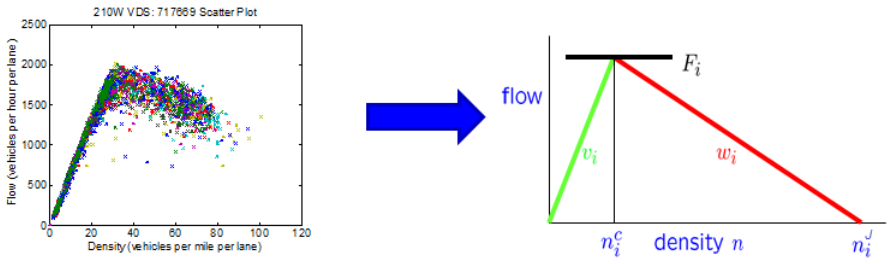

Figure 3.5: Fundamental Diagram fitting scheme

In this section, we demonstrate how the automated parameter identification process implemented by TOPL works. The first step is to gather data from PeMS for each loop detector station present in the modeled stretch of road network. Customarily, we download data from the PeMS Data Clearing House (DCH), which is a repository dedicated to provide users access to daily packets of data from each detector in the districts of California. Data collection from PeMS can also be automated by specifying database querying scripts.

To estimate the fundamental diagram parameters at a detector location, the observed 5

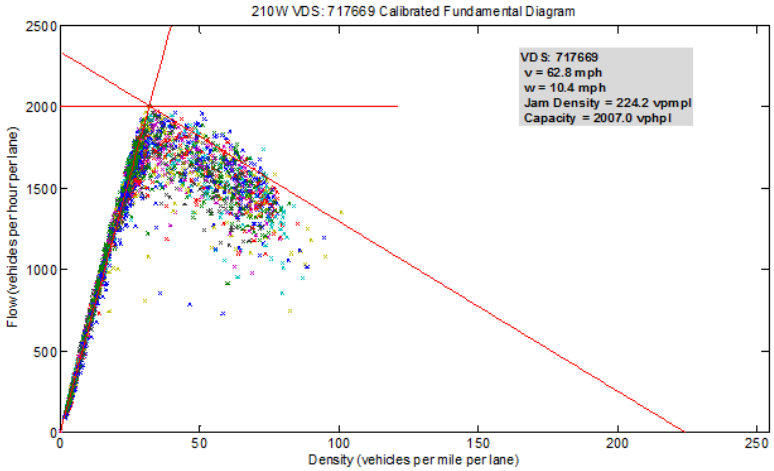

Figure 3.6: Example calibrated fundamental diagram on Interstate 210 Westbound

minute aggregated Flow - Density pairs are combined in a scatter plot as shown in figure 3.6. In this example plot, there are 108 days' data combined in one plot, every color corresponding to a different day and this yields a total of 108x288 = 31104 individual data points. Three parameters need to be identified to uniquely determine the triangular fundamental diagram:

1. **Free Flow Speed (v):** The free flow speed corresponds to the slope of the left-hand side portion of the triangle, where the density is below critical. This slope is determined by performing a linear least squares fit on the set of Flow-Density pairs that correspond to measured speeds that are higher than 55 mph. 55 mph is the PeMS criteria for a threshold speed between free flowing and dense traffic conditions. Here we are making use of the fact that PeMS provide estimates of speeds in addition to flows and densities.

2. **Capacity:** The capacity of a freeway section is defined as the apex of the fundamental

diagram in the macroscopic modeling context. However, care must be taken when attributing observed maximum flows to the actual capacity of the section. First of all, the section must get congested during the analyzed day for the capacity to be observed. Second, the causes of congestion also determine whether the observed maximum flow in fact corresponds to the actual section capacity. If the section gets congested due to a lane drop as a result of an incident or because it is hit by a back-propagated congestion wave from downstream, the observed maximum flows may be below what can ideally be achieved. Furthermore, it is demonstrated in author's own work [12] that capacity, when defined as maximum flow observed during a day on which the section got congested, shows significant variability. Figure 3.7 demonstrates this variability in capacity (or in daily observed maximums to be more accurate) over a stretch of southbound Interstate 880 in East Bay approximately from Oakland to Fremont. The figure shows box plots of daily maximum flows for each detector station listed in the horizontal axis. The vertical axis shows the maximum flow values and listed along the upper horizontal edge of the figure are the number of days that particular section got congested, i.e. the number of data points for each particular box plot. The red lines inside the box plots correspond to the median of the observed capacities among days. The lower and upper box boundaries represent 1st and 3rd quartiles, i.e. 25th and 75th percentiles, respectively. The whiskers span from either end of the box to the smallest and largest data points that are non-outliers, where "outliers" are defined as any points that are more than 1.5 interquartile range away from box boundaries.

The interquartile range is defined as the difference between the 25th percentile and the 75th percentile of the data. Individual outliers are marked as points. It is noteworthy that in the extreme case of detector 400352, the capacity varies between 2186 and 1324 vehicles per hour per lane. Note that both these values reflect the highest observed flow on a single day of operation across this detector. Another important observation is that most of the outliers fall below the lower whisker, indicating that external random factors such as weather, incidents, events, etc., which were not controlled for in this study, affect the capacity adversely. As for the outliers above the higher whisker, they may be attributed to the expected offset and noise in the detectors, which will also be explored in section 4.1. Under the light of these observations, the capacity of a section is assigned as the *maximum non-outlier* value, defined as

$$\text{Capacity} = \max_{f_i}(f_i | f_i < Q_3 + 1.5IQR) \tag{3.3}$$

where f_i are the individual maximum flows observed on each congestion-bearing day, Q3 is the 75th percentile of all the daily observed maximums, and IQR = 75th percentile - 25th percentile. For sections that do not get congested for any of the observed days, a nominal value (usually 2000 vehicles per hour per lane) is assigned as the capacity. Of course, this nominal value should be higher than the observed flows in free flow.

Capacity was demonstrated to be the parameter to which the adopted models are most sensitive by Muralidharan et. al in [40]. Hence, the correct estimation of capacity is of utmost importance and is an ever evolving debate in the traffic literature.

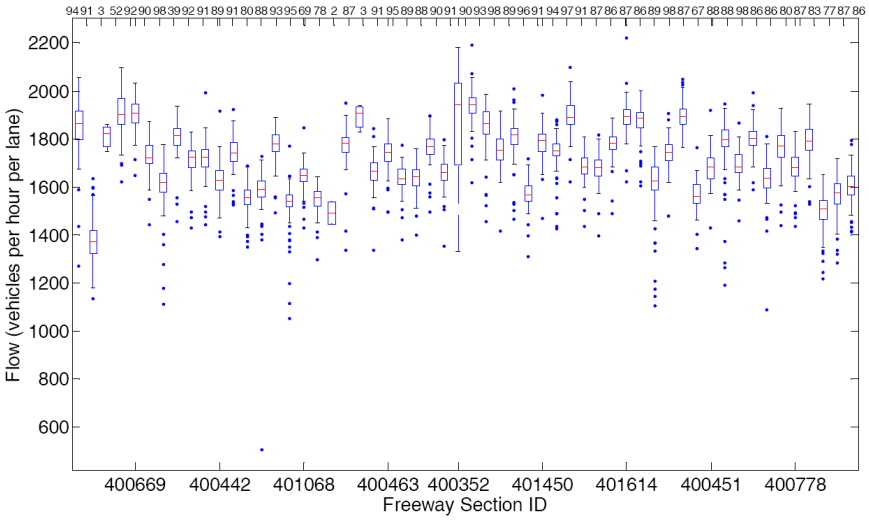

Figure 3.7: Box Plots of Capacity along I-880 South

3. **Congestion Wave Speed (w):** This parameter determines the speed at which the congestion propagates upstream of the section. Like capacity, this parameter also shows significant variability [6]. Following the reasoning made for the variability in capacity, we want to exclude artificially lower flow data points since those points most likely correspond to "worse-than-ideal" operating conditions for the freeway. Since we want to capture the "ideal" values of the parameters and use our simulation tools to replicate "worse-than-ideal" conditions, we adopt an approximate quantile regression scheme [34] to fit the right-hand-side portion of the triangular fundamental diagram. Once the capacity flow is determined, the corresponding density value is deemed the *critical density*. The data points that fall to the right of the vertical line drawn from the apex to the critical density constitute the congested regime data. To estimate the congestion wave speed, these data points are partitioned along the density axis into

non-overlapping bins of 10 data points each. From each of these bins, one representative point is chosen as the "maximum non-outlier" within the bin, and then a constrained least squares fit is performed to fit a line through these representative points and the apex of the triangle. Figure 3.8 shows a detailed view of this procedure, where the green rectangles in the zoomed view show the representative points for each bin. The estimation of the w-line also establishes the *jam density* for the section which is the density value where the w-line intersects the horizontal axis. When the section to be calibrated does not get congested for any of the available days, a nominal value of 10 mph is assigned as the congestion wave speed. This value is an average value over all the available estimations that have been carried through as part of the TOPL project.

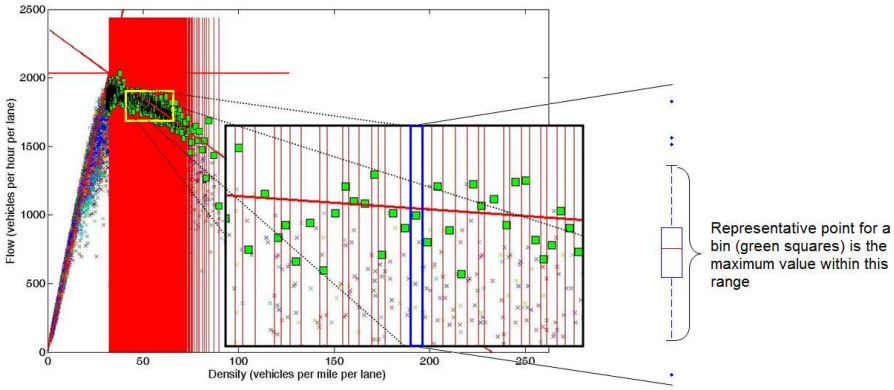

Figure 3.8: Estimation of the congestion wave speed (w)

Once the fundamental diagram parameters are identified at each detector location, they need to be assigned to their corresponding links. In all freeways analyzed and modeled by the TOPL project (I-680, I-210, SR-101, I-15, I-80, I-880), the detector placement is such that

measurements are taken just upstream of on-ramps and/or just downstream of off-ramps. The LN-CTM, introduced in the previous section, assumes a link to start just at the on-ramp merge. Hence we assign the fundamental diagram estimated at a sensor location to the link just downstream of the detection. Here, the choice is between assigning the upstream detector versus assigning the downstream detector to the cell. The choice of the upstream assignment is justified by the model assumptions. Of the two modes a cell can be in, during free flow we do not expect any significant difference in the normalized density (normalized by lane and space) between the measurement location and inside of the cell and when the cell is congested, the congestion spills back in the upstream direction, giving rise to a cell density better captured by the upstream detector location rather than the downstream.

Due to malfunctioning or, more frequently, a lack of adequate detector placement, there are several links with no detection. In such cases, the parameters estimated at the nearest upstream detector station are assigned to these links (see the double assignment of the fundamental diagram on the right, in the top part of figure 3.9). Another important exception to the rule is the case of major on-ramp merges, such as freeway to freeway connectors where the downstream detector gives a more accurate flow-density relationship for the link (see the double assignment of the fundamental diagram on the left, in the bottom part of figure 3.9).

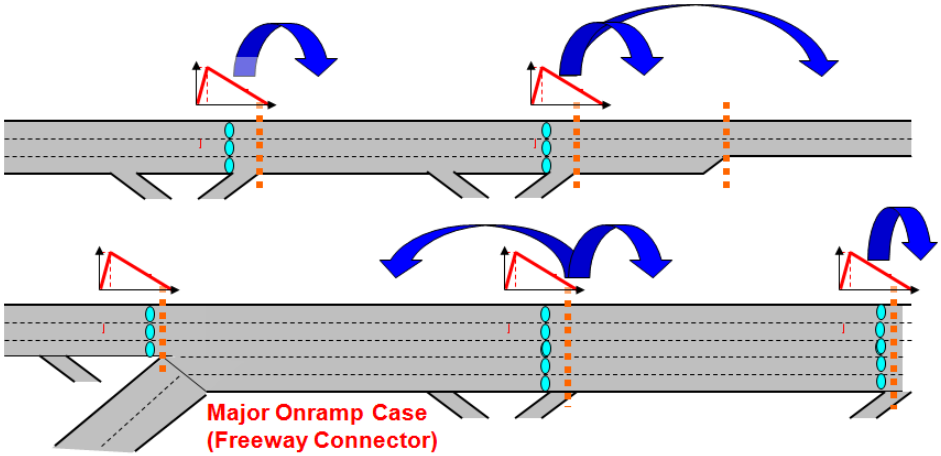

Figure 3.9: Assignment of Fundamental Diagrams to freeway links

3.3 IMPUTATION OF UNKNOWN RAMP FLOWS

After the fundamental diagram parameters are calibrated for each link, the next step to completing the model calibration is the estimation of unobserved ramp flows. These unknown ramp flows are estimated using an imputation algorithm devised by Ajith Muralidharan, another member of the TOPL team. In this section, we present a concise overview of this algorithm to complete the picture of the automated model building procedure. For a more detailed account and convergence analysis of the algorithm, the reader is referred to [38].

The imputation algorithm that is deployed within TOPL is a model-based algorithm built upon the LN-CTM introduced in section 3.1. The goal of the algorithm is to construct artificial ramp flows at specified locations using an iterative learning algorithm to match model densities and flows with the measurements taken from the mainline.

For the purposes of the imputation algorithm, the following additional terms need to be

introduced,

- *Capacity adjusted free-flow speed ($\bar{v}(k)$):* $\bar{v}_i(k) = \bar{v}_i(n_i(k)) = \min\left(v_i, \frac{F_i}{n_i(k)}\right)$
- *Capacity adjusted congestion wave speed ($\bar{w}(k)$):* $\bar{w}_i(k) = \bar{w}_i(n_i(k)) = \min\left(w_i, \frac{F_i}{(n_i^J - n_i(k))}\right)$
- *Total/effective demand ($c(k)$):* $c_i(k) = n_i(k)\bar{v}_i(k)(1 - \beta_i(k)) + d_i(k)$

Under these definitions, the demand, and supply functions are represented as

$$\begin{aligned} D_i(k) &= \min(n_i(k)v_i, F_i) = n_i(k)\bar{v}_i(k) \\ S_i(k) &= \bar{w}_i(k)(n_i^J - n_i(k)) \end{aligned} \tag{3.4}$$

and the model equations become

Density Update:

$$n_i(k+1) = n_i(k) + f_i^{in}(k) - f_i^{out}(k) \tag{3.5}$$

Flow Update:

$$\begin{aligned} f_i^{in}(k) &= \min\left(\bar{w}_i(k)(n_i^J - n_i(k)), c_{i-1}(k)\right) \\ f_i^{out}(k) &= n_i(k)\bar{v}_i(k) min\left(1, \frac{\bar{w}_{i+1}(k)(n_{i+1}^J - n_{i+1}(k))}{c_i(k)}\right) \\ c_i(k) &= n_i(k)\bar{v}_i(k)(1 - \beta_i(k)) + d_i(k) \\ d_i(k+1) &= d_i(k) + Q_i(k) - r_i(k) \\ s_i(k) &= \beta_i(k) f_i^{out}(k) \\ r_i(k) &= \frac{min(c_i(k), \bar{w}_{i+1}(k)(n_{i+1}^J - n_{i+1}(k)))}{c_i(k)} d_i(k) \end{aligned} \tag{3.6}$$

Due to the dual nature of both the incoming and the outgoing flows for each link, LN-CTM is a piecewise nonlinear model with four modes. Depending on flow conditions upstream and downstream of a link, i.e. whether these are in free flow or in congested operating conditions, the four modes and their corresponding density update equations can be written as:

$$
\begin{aligned}
n_i(k+1) &= n_i(k) + c_{i-1}(k) - n_i(k)\bar{v}_i(k) && \textbf{(FF mode)} \\
n_i(k+1) &= n_i(k) + c_{i-1}(k) - n_i(k)\bar{v}_i(k)\frac{\bar{w}_{i+1}(k)(n^J_{i+1} - n_{i+1}(k))}{c_i(k)} && \textbf{(FC mode)} \\
n_i(k+1) &= n_i(k) + \bar{w}_i(k)(n^J_i - n_i(k)) - n_i(k)\bar{v}_i(k)\frac{\bar{w}_{i+1}(k)(n^J_{i+1} - n_{i+1}(k))}{c_i(k)} && \textbf{(CC mode)} \\
n_i(k+1) &= n_i(k) + \bar{w}_i(k)(n^J_i - n_i(k)) - n_i(k)\bar{v}_i(k) && \textbf{(CF mode)}
\end{aligned}
\tag{3.7}
$$

where **F** stands for free flow and **C** stands for congestion in the mode labels following the equations. The first letter is for the flow conditions upstream and the second is for downstream. With this formulation, the imputation algorithm solves a two stage problem:

1. Estimate the effective demand vectors $c_i(k)$ for each link on the freeway using an adaptive learning scheme to match the density measurements on the mainline.

2. Using the estimated effective demands, evaluate incoming and outgoing flows to and from each link. Using these flows and the measured flows (usually measured at the

location depicted by figure 3.10), solve a linear program to produce on-ramp flows $\hat{r}_i(k)$ and off-ramp split ratios $\hat{\beta}_i(k)$, where the "hats" denote estimated variables.

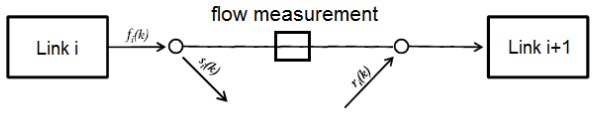

Figure 3.10: Location of measurement with respect to the LN-CTM

3.4 MODEL VALIDATION

The identification of fundamental diagram parameters and the estimation of unknown ramp flows (assuming there are no faulty detectors) produces a complete model. The next step is the open-loop simulation of this model and assessing its performance. The performance criteria adopted by TOPL are as follows:

- **Density Error:** The mean absolute error between the model calculated and measured densities along the freeway.

$$\text{Density error } = \frac{\sum_i \sum_k |n_i(k) - n_i^{meas}(k)|}{\sum_i \sum_k |n_i^{meas}(k)|} \tag{3.8}$$

- **Flow Error:** The mean absolute error between the model calculated and measured flows along the freeway.

$$\text{Flow error } = \frac{\sum_i \sum_k |f_i(k) - f_i^{meas}(k)|}{\sum_i \sum_k |f_i^{meas}(k)|} \tag{3.9}$$

- **VMT Error:** The relative error in hourly Vehicle Miles Traveled

$$\text{VMT error } = \frac{\sum_\kappa |VMT(\kappa) - VMT^{meas}(\kappa)|}{\sum_\kappa VMT(\kappa)} \tag{3.10}$$

- **VHT Error:** The relative error in hourly Vehicle Hours Traveled

$$\text{VHT error } = \frac{\sum_\kappa |VHT(\kappa) - VHT^{meas}(\kappa)|}{\sum_\kappa VHT(\kappa)} \tag{3.11}$$

- **Delay Error:** The relative error in hourly Delays

$$\text{Delay error } = \frac{\sum_\kappa |VCD(\kappa) - VCD^{meas}(\kappa)|}{\sum_\kappa VCD(\kappa)} \tag{3.12}$$

where VMT, VHT and Delay are defined as:

$$\begin{aligned}
VHT(\kappa) &= \sum_i \sum_{k=\kappa*3600/T_s}^{(\kappa+1)*3600/T_s} n_i(k) \\
VMT(\kappa) &= \sum_i \sum_{k=\kappa*3600/T_s}^{(\kappa+1)*3600/T_s} f_i(k) \\
Delay(\kappa) &= \sum_i \sum_{k=\kappa*3600/T_s}^{(\kappa+1)*3600/T_s} (n_i(k) - f_i(k)/V_i)\, \mathbb{I}(V_s(k) < 55mph)
\end{aligned}$$

In the delay formulation, the 55 mph speed is a user-specified reference speed set at this value for freeways. In addition to the calculated measures above, the contour plots of speeds,

flows and measurements along the freeway provide an important visual inspection tool to assess if the model is able to replicate the congestion patterns and important bottleneck locations on the freeway. An example set of these plots are provided in figures 3.11 and 3.12. In these contour plots, the horizontal axis is the spatial coordinate whereas the vertical coordinate axis corresponds to the time of day in hours. The plots show a single day, i.e. a 24 hour period. The traffic flows from right to left in these particular plots. The color palettes next to the plots define the color vs value matchings for the contours. This example shows a calibrated 23 mile stretch of eastbound I-80 on August 15h, 2008 for a 24 hour time range.

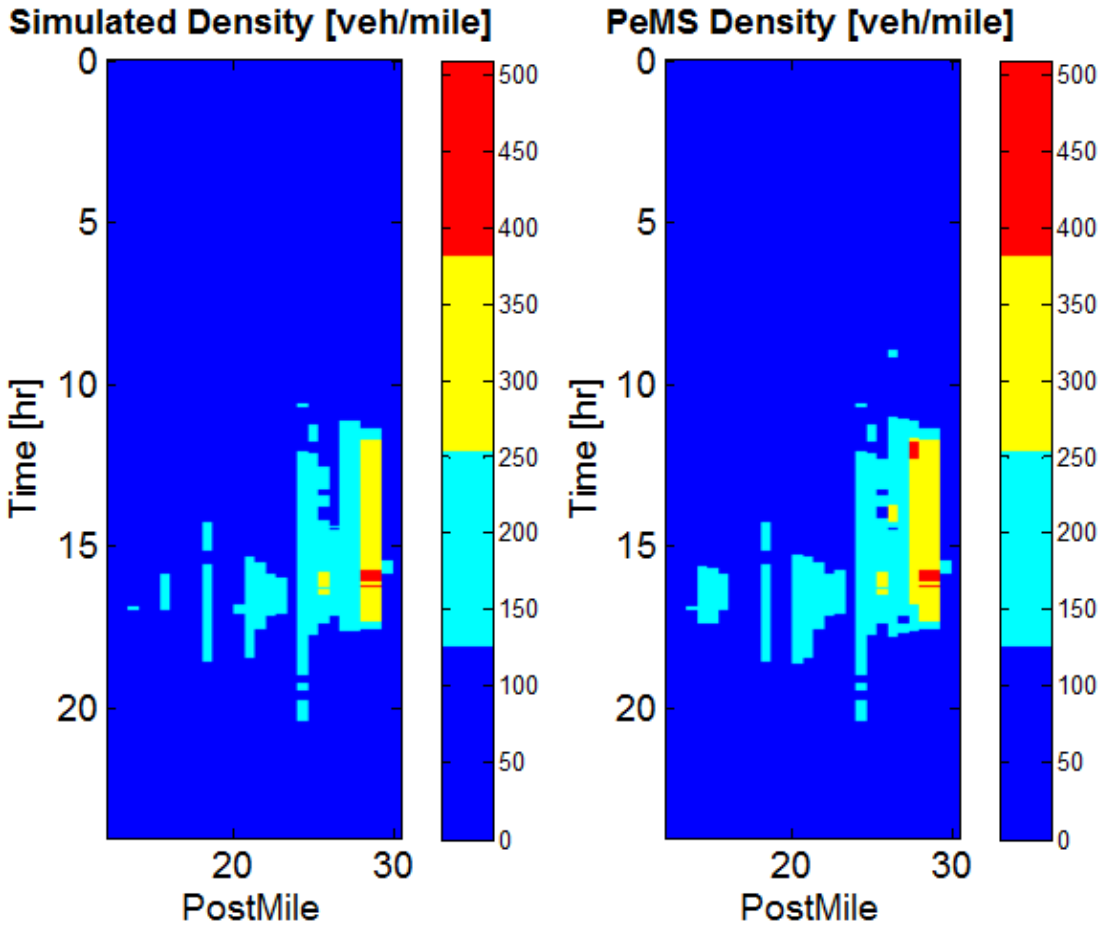

Figure 3.11: Example Density Contour Plots for a model of I-80 Eastbound

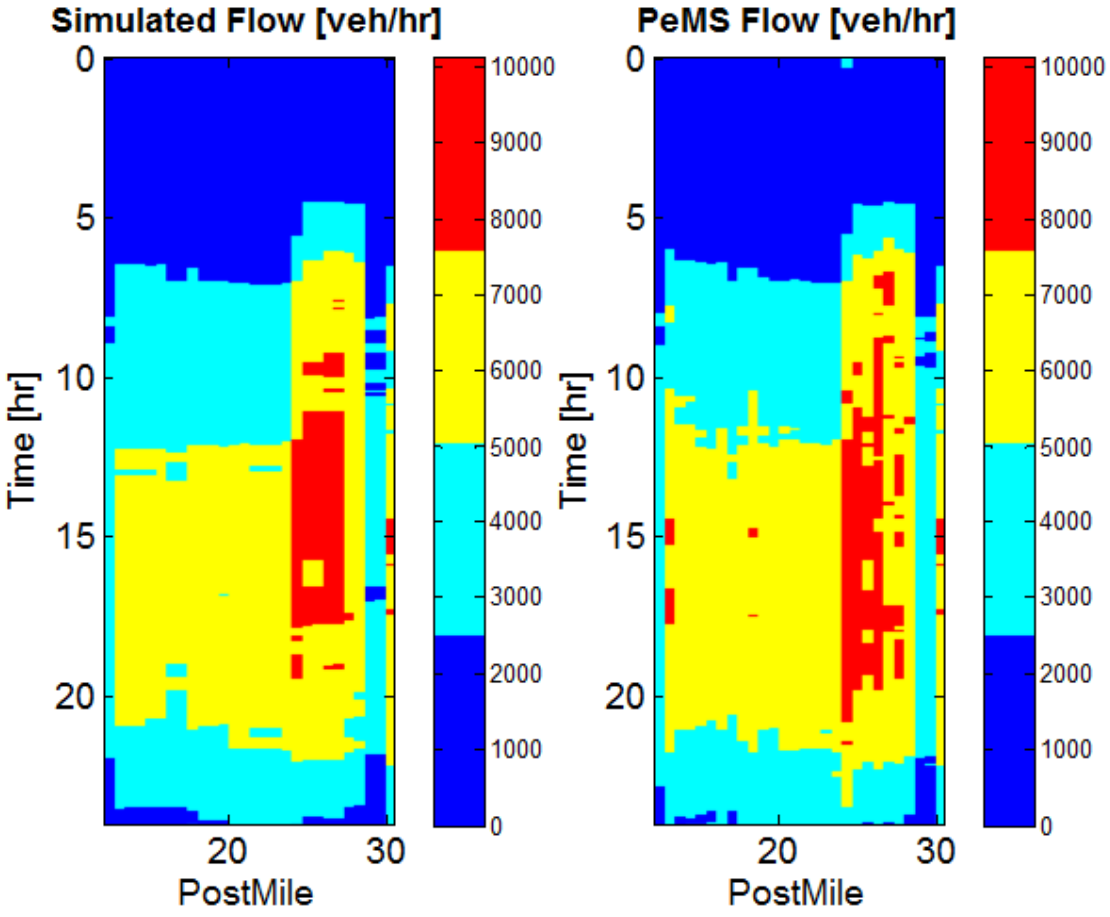

Figure 3.12: Example Flow Contour Plots for a model of I-80 Eastbound

Chapter 4

FAULT DETECTION AND HANDLING

4.1 MOTIVATION FOR MODEL-BASED FAULT DETECTION

Before introducing the model-based fault detection scheme, we first provide a background on common measurement errors that cannot be identified by the statistical methods applied on the raw data by PeMS. Since the model building tools developed by TOPL are based on data already processed and filtered by PeMS, the aim of the model-based approach is to detect further errors in the data that have slipped through the PeMS filters. These errors will usually arise in the form of consistent biases and off-sets rather than statistically aberrant

measurement noise. Therefore, the decision logic to be introduced in section 4.3 classifies potential errors as bias in flow and density measurements. It is clear that such errors can not be detected by comparative statistical methods if they occur in a consistent manner over days of operation.

Measurement bias in flow counts was evidenced as a result of an ancillary study that investigated the discrepancy between adjacent loop detectors that report to two separate authorities but should nevertheless, in theory, report the same counts. There are many federal Census detectors on several locations along California Freeways and a significant portion of these are located adjacent to PeMS detectors which are under Caltrans authority, insofar as to share the same physical inductive loop in a few cases. It was observed that even though these Census and PeMS detectors are right next to each other, they sometimes show consistent bias towards one of the reported total daily counts. Figure 4.1 (left) shows a scatter plot of 143 days of total daily flows at such a location. Looking at the plot, it is obvious that the counts are highly correlated but the Census counts are, on average, 20% higher than the PeMS counts, as the right hand side plot of the figure exhibits. Two such sites have been further examined in an effort to explain this discrepancy. To this goal, hand counts were established from video feeds at these two sites to compare to the filtered PeMS data. The details and complete results of this study are beyond the scope of this work; however, the findings about PeMS counts are highly pertinent to the problem of finding consistently biased measurements. For the two analyzed sites, it was found that 5 minute aggregated PeMS data contains 4.3% positive bias relative to the hand counts. Among the

two sites, a total of 6 hours of data were analyzed which amounts to a sample size of 72 data points. It was also observed that the PeMS counts were always higher than the hand counts, i.e. the bias was always positive. The error in hand counts was estimated to be around 1%. Figure 4.2 shows the 15 minute aggregated relative errors in PeMS counts with respect to the hand counts for a four hour period. Another observation was that the relative bias with respect to the hand counts shows no significant dependence on flow volume or traffic conditions.

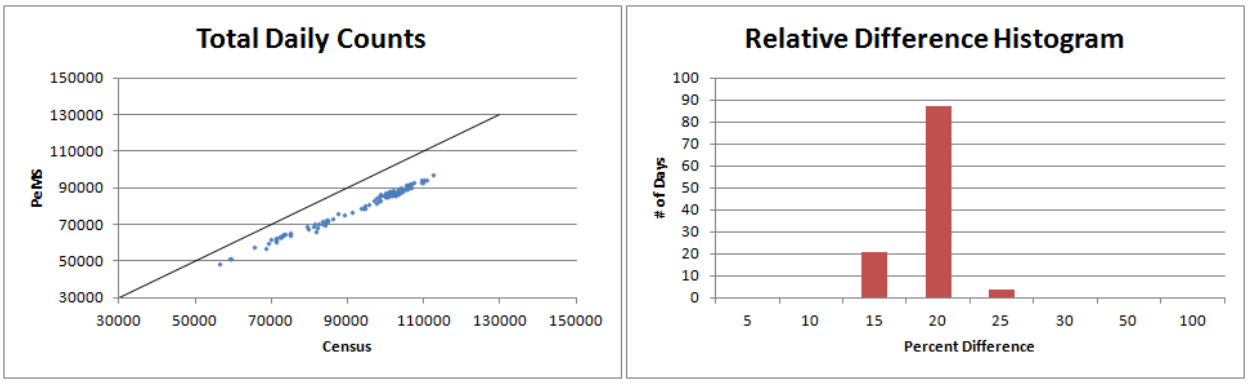

Figure 4.1: Census ve PeMS daily total flows

For density measurements, the cause of measurement bias is a predominantly structural problem related to sensor placement and model assumptions. On California freeways the loop detectors are customarily placed immediately upstream of on-ramps and downstream of off-ramps. When specifying the model geometry, every link is associated with one of these detectors. Figure 3.9, shows this association in the LN-CTM sense. ACTM has a similar detector-link association scheme depicted in figure 4.3. In this figure, the links are enumerated as 1 and 2, the ramp flows are indicated by the arrows below as r and s and the

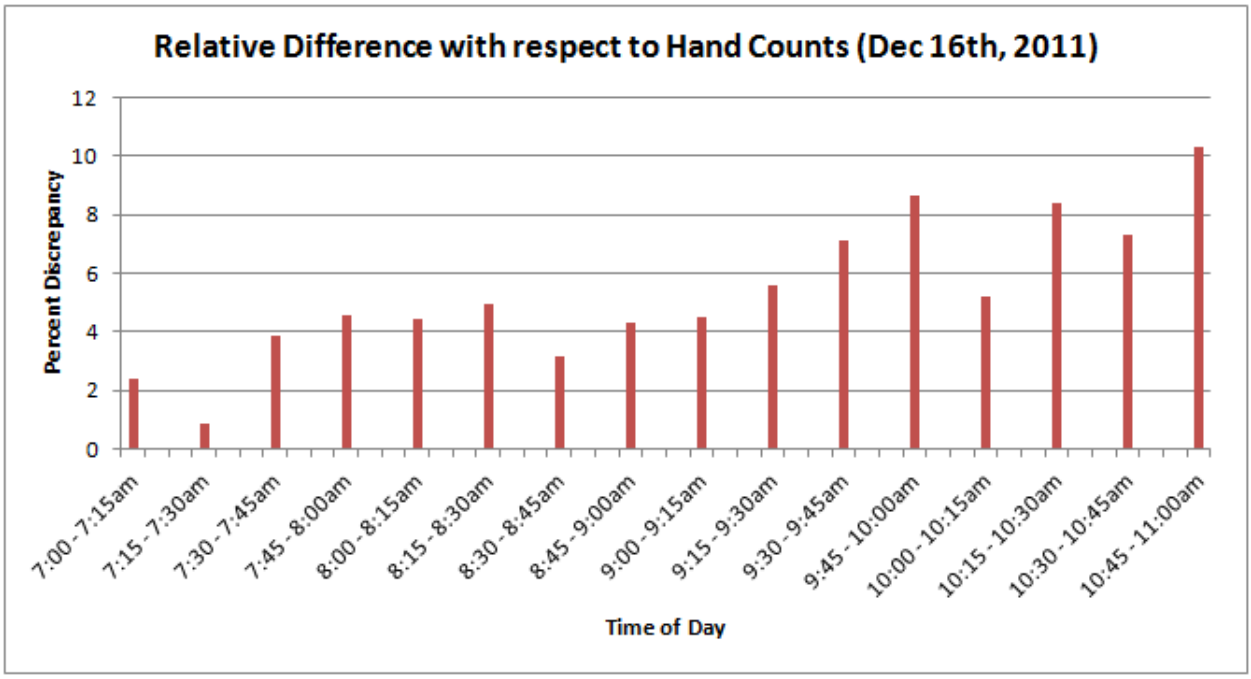

Figure 4.2: Relative bias in PeMS flows with respect to hand counts

measurements f (flow) and n (density) are taken at the locations indicated by the circles. Every detector is associated with the link immediately downstream of the detector location. Since the ACTM defines a link with an on-ramp at the beginning and an off-ramp at the end, in this formulation the detector actually is within the geometrical confines of the link but is nonetheless upstream of the on-ramp and thus may fail to capture the density of the whole link accurately. Due to this placement of the detectors which is incompatible with the model assumptions, it is expected that all density measurements carry an unavoidable bias. Since density is the essential state of the model on which both the imputation and the simulations are based, this bias can have a significant effect on the accuracy of the model. This phenomenon constitutes the basic reason for a model-based fault detection algorithm.

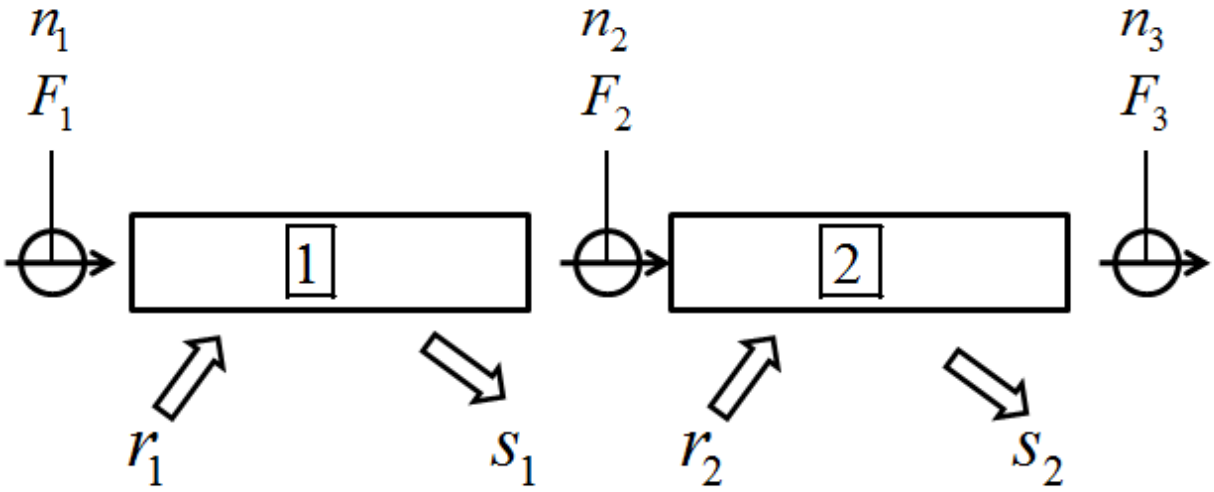

Figure 4.3: Basic ACTM representation and Location of Detectors

4.2 THE ASYMMETRIC CELL TRANSMISSION MODEL

The Asymmetric Cell Transmission Model (ACTM) is another extension of the Cell Transmission Model adopted by earlier deployments of the TOPL freeway modeling and calibration process [22], [21]. Compared to LN-CTM, introduced in section 3.1, ACTM is a simpler model that is particularly suited to modeling unidirectional freeways rather than more complex networks. The fault detection algorithm to be introduced in the next section exploits some favorable properties of ACTM which the LN-CTM does not provide. For this reason, the fault detection algorithm is based on the ACTM. In this section we summarize the ACTM and the ACTM-based imputation algorithm as an essential background to the

discussion of fault detection in the next section.

The ACTM specifies the freeway as a sequence of cells, each with at most one on-ramp near the beginning of the section and at most one off-ramp near the end of the section (see figure 4.4). Every cell is assumed to have a detector station associated with it, and when specifying the model geometry, it is aimed to build the cells in a such a way that every cell contains one vehicle detector station. This is depicted in figure 4.5 and should be compared to figure 3.4 for a visual assessment of the discrepancy between the ACTM and the LN-CTM.

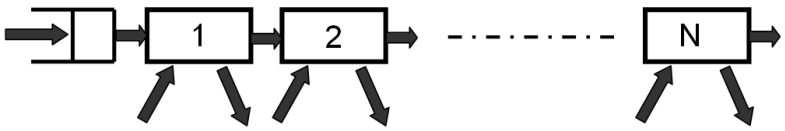

Figure 4.4: ACTM schematic representation

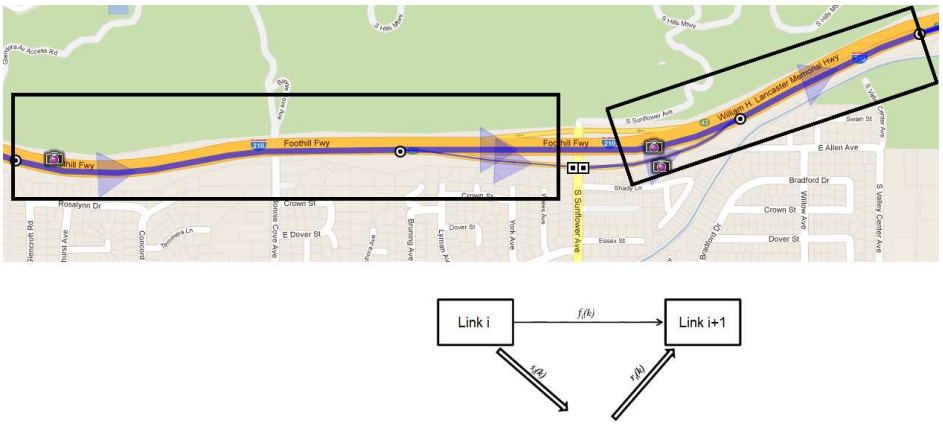

Figure 4.5: ACTM representation of a freeway portion

The equations that govern the evolution of density and flow for a given link i in the

ACTM are as follows:

Density:

$$n_i(k+1) = n_i(k) + f_{i-1}(k) - f_i(k) + r_i(k) - s_i(k), \quad 1 \leq i \leq N \quad (4.1)$$

Flow:

$$f_i(k) = \min(v_i n_i(k) - s_i(k), w_{i+1}[n_{i+1}^J - n_{i+1}(k)], F_i) \quad 1 \leq i < N \quad (4.2)$$

Boundary Flows:

$$f_N(k) = \min(v_N n_N(k) - s_N(k), F_N) \quad (4.3)$$

$$f_0(k) = \min(w_1[n_1^J - n_1(k)], Q(k) + f_{in}(k)) \quad (4.4)$$

$$Q(k+1) = Q(k) + f_{in}(k) - f_0(k) \quad (4.5)$$

where i is the cell index and k is the discrete time index. The model variables and parameters are as listed in table 4.1.

It is customary to specify the downstream boundary conditions as free flow and at the upstream boundary, flow is assumed to be fed to the freeway through a queue. A cell i is defined to be in free flow if its corresponding discharge flow $f_i(k)$ satisfies

$$v_i n_i(k) - s_i(k) < \min[w_{i+1}[n_{i+1}^J - n_{i+1}(k)], F_i] \quad (4.6)$$

and it is in congestion if the inequality above is violated. Hence, the system essentially is a *piecewise affine switching mode system* with congested and free flow modes for each cell.

Symbol	Name	Units
F_i	maximum flow (capacity) of section i	veh/period
v_i	free flow speed of section i	section/period
w_i	congestion wave speed of section i	section/period
n_i^c	critical density of section i	veh/section
n_i^J	jam density of section i	veh/section
$f_i(k)$	flow from section i to $i+1$ at period k	veh/period
$s_i(k), r_i(k)$	off-ramp, on-ramp flow in section i at period k	veh/period
$n_i(k)$	number of vehicles in section i at period k	veh/section
$Q(k)$	number of vehicles in the input queue to section 1 at period k	veh
$f_{in}(k)$	input flow at upstream queue at period k	veh/period

Table 4.1: Asymmetric Cell Transmission Model (ACTM) Variables and Parameters

In contrast, the LN-CTM is a *piecewise non-linear switching mode system* which complicates the building of model-based schemes like imputation and fault detection upon it.

ACTM IMPUTATION

The imputation algorithm, briefly introduced in section 3.3, was first developed for the ACTM and it is essential to the fault detection scheme to be introduced in the following sections. The ACTM-based imputation algorithm is also due to fellow TOPL researcher Ajith Muralidharan [41], [38].

The ACTM-based imputation algorithm uses an *iterative learning control (ILC)* approach rather than the adaptive control approach described in section 3.3 for the LN-CTM-based imputation. The goal again is to match the mainline flows and densities by fitting artificial

ramp flows at specified locations. In the following, the model and learning dynamics are represented in continuous time for ease of representation and analysis, but in practice they are implemented in discrete time with a small time step and small gains, so that the imputation procedure as well as the model are stable. The upstream section is specified with the subscript *up* and the downstream section is specified by *dn*. Figure 4.6 shows the parameters and measurements used for imputation of ramp flows in section i. The upstream boundary conditions include the upstream density, fundamental diagram parameters and the off-ramp flow $s_{i,up}$. Since all the parameters and variables carry the subscript i, for clarity, we drop it in the following equations.

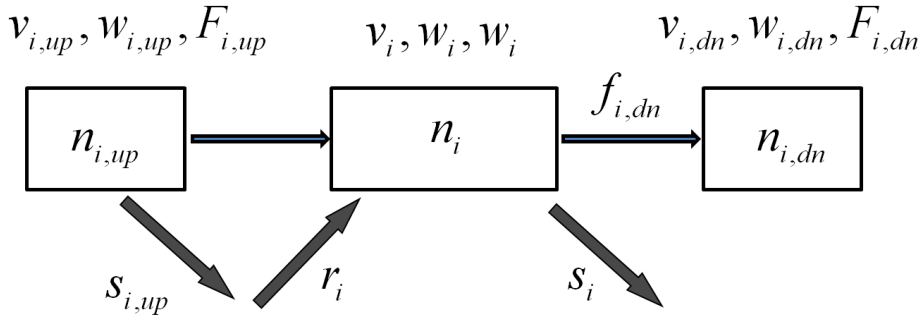

Figure 4.6: Imputation Parameters and Cell Definitions

It is assumed that the density and ramp flow profiles are 24 hour periodic, and the on-ramp and off-ramp flows are represented as a convolution of a kernel with a constant periodic ramp flow parameter vector:

$$r(t) = \int_0^T K_r(\tau, t) c_r(\tau) d\tau, \quad s(t) = \int_0^T K_s(\tau, t) c_s(\tau) d\tau \tag{4.7}$$

In equation (4.7), $K_r(\tau, k)$ and $K_s(\tau, k)$ represent periodic, time dependent kernel functions with a period of 24 hours. A typical kernel function would be an impulse or a gaussian window centered at time k.

For each section, the imputation procedure assumes initial estimates for the constant ramp parameter vectors $\hat{c_r}$ and $\hat{c_s}$. These estimates are then adapted so that the model calculated densities match the density profile measured by the vehicle detector station. The model variables and the estimates are denoted by a hat accent ($\hat{}$) and the residuals by a tilde accent ($\tilde{}$). The true values are represented without any accent.

The four modes considered by the imputation algorithm are:

1. **P-F:** Plant in Free Flow
2. **P-C:** Plant in congestion
3. **M-F:** Model in Free Flow
4. **M-C:** Model in Congestion

These modes for both the plant and the model are determined by equation (4.6). The mode dependent adaptation laws for the parameters at each step are given by:

- **P-F , M-F** (plant and model are both in free-flow):

$$\dot{\hat{c}}_r(\tau, t) = G_1 K_r(\tau, t)\tilde{n}(t)$$
$$\dot{\hat{c}}_s(\tau, t) = -G_2 K_s(\tau, t)\tilde{f}_{dn}(t) \tag{4.8}$$

- **P-C , M-C** (plant and model are both in congestion):

$$\dot{\hat{c}}_r(\tau,t) = G_1 K_r(\tau,t)\tilde{n}(t)$$

$$\dot{\hat{c}}_s(\tau,t) = -G_1 K_s(\tau,t)\tilde{n}(t) \tag{4.9}$$

- **P-C , M-F** (plant is in congestion and model is in free flow):

$$Case(i) \qquad \tilde{n}(k) > 0$$

$$\dot{\hat{c}}_r(\tau,t) = G_1 K_r(\tau,t)\tilde{n}(t)$$

$$\dot{\hat{c}}_s(\tau,t) = -G_1 K_s(\tau,t)\tilde{n}(t) - G_2 K_s(\tau,t)\frac{\tilde{f}_{dn}(t) + |\tilde{f}_{dn}(t)|}{2}$$

$$Case(ii) \qquad \tilde{n}(t) \leq 0$$

$$\dot{\hat{c}}_r(\tau,t) = G_1 K_r(\tau,t)\tilde{n}(t)$$

$$\dot{\hat{c}}_s(\tau,t) = -G_2 K_s(\tau,t)\tilde{f}_{dn}(t) \tag{4.10}$$

- **P-F , M-C** (plant is in free flow and model is in congestion):

$$Case(i) \qquad \tilde{n}(k) < 0$$

$$\dot{\tilde{c}}_r(\tau,t) = G_1 K_r(\tau,t)\tilde{n}(t)$$

$$\dot{\tilde{c}}_s(\tau,t) = -G_1 K_s(\tau,t)\tilde{n}(t) - G_2 K_s(\tau,t)\tilde{f}_{dn}(t)$$

$$Case(ii) \qquad \tilde{n}(k) \geq 0$$

$$\dot{\tilde{c}}_r(\tau,t) = G_1 K_r(\tau,t)\tilde{n}(t)$$

$$\dot{\tilde{c}}_s(\tau,t) = -G_2 K_s(\tau,t)\tilde{f}_{dn}(t) \tag{4.11}$$

where G's are user defined positive gains. The model density is updated at each time step using the most recent estimates of the unknown ramp flow inputs by the following set of equations:

$$\tilde{n}(t) = n(t) - \hat{n}(t) \tag{4.12}$$

$$\dot{\hat{n}}(t) = \hat{f}_u - \hat{f}_d + \hat{r}(t) - \hat{s}(t) - a\tilde{n}(t) \tag{4.13}$$

$$\hat{f}_u = min(n_{up}(t)v_{up} - s_{up}(t), F_{up}, w(n^J - \hat{n}(t))) \tag{4.14}$$

$$\hat{f}_d = min(\hat{n}(t)v - \hat{s}(t), \bar{w}_{dn}(t)(n^J_{dn} - n_{dn}(t))) \tag{4.15}$$

$$\tilde{f}_{dn}(t) = f_{dn}(t) - (n(t)v - \hat{s}(t)) \tag{4.16}$$

where $\bar{w}_{dn}(t) = min\left(\frac{F}{n^J_{dn} - n_{dn}(t)}, w\right)$ is introduced in order to absorb the capacity flow into the congestion mode and the parameter a in (4.13) is chosen so as to make the error

equations asymptotically stable. The adaptation is carried out for the entire density profile multiple times, so as to reduce the 24-hour 'residuals' $\sum_k |n(k) - \hat{n}(k)|$ and $\sum_k |f_{dn}(k) - \hat{f}_{dn}(k)|$. This procedure is repeated until either both the residuals become small (eg. $< 0.5\% \quad of \quad \sum_k n(k)$ and $< 0.5\% \quad of \quad \sum_k f_{dn}(k)$) or the residuals do not improve (eg. change of errors of $< 0.5\%$ of $\sum_k n(k)$ and $< 0.5\%$ of $\sum_k f_{dn}(k)$ across iterations)

Due to the piecewise affine formulation of the ACTM, the imputation can be carried out sequentially for each section, starting from the most upstream section. Hence, for the estimation of ramps in section i, only the immediate upstream section $i-1$ and the immediate downstream section $i+1$ are considered. This property is instrumental to the isolation of faults for the model-based fault detection scheme to be introduced momentarily.

4.3 FAULT DETECTION ALGORITHM

In the modeling process flow outlined in figure 3.1 the identification of faulty mainline detectors, when carried out manually, is the most time-consuming step of the procedure. Therefore, an automated procedure based on the evaluation of residuals between the measured and simulated densities and flows has been devised. Figure 4.7 shows the block diagram of this framework.

With regard to the treatise in section 2.3, this framework can be regarded as a *combination of the parity and parameter estimation methods*. The relatively simple parity approach is applicable in this case because we have a discrete time model of the system and the data

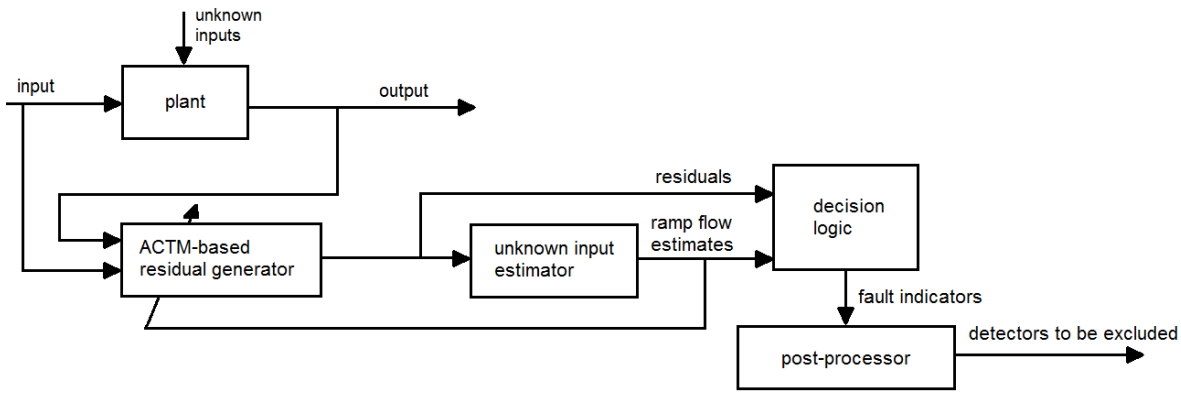

Figure 4.7: Fault Detection Framework

become available as daily packets in the PeMS repository. This allows for the iterative learning of the unknown ramp flows, as explained in section 4.2, and facilitates the residual generation for the period of one complete day. The caveat of this structure is that the method can not be implemented in real-time due to its dependence on a complete set of daily measurement profiles.

To analyze the effects of faults on the freeway model, we start with an artificial freeway of only four links. Simulated data are used as measurements and fault signals are superposed upon these measurements at a known isolated location. When the unknown ramp flows are imputed based on these artificially polluted mainline data, the fault signals manifest themselves in the estimated variables in very specific patterns depending on the freeway geometry and the nature of the fault signal. A convenient consequence of using the ACTM is that the fault in the measurements of a link only affects the link itself, and the two links adjacent to it upstream and downstream, but not the other links.

To demonstrate an example, we consider 3 successive cells, all of which have both an on-ramp and an off-ramp and none of the ramp data are available (i.e. they are all to be imputed). This configuration is depicted in figure 4.8. We inject bias terms, ϕ and ψ, respectively, in the flow and density measurements just upstream of cell 2, which the model assigns as the flow entering cell 2 and density of cell 2, respectively. To conform with the model definition in chapter 4.2, we denote the flow measured at this location f_1 whereas the density measurement is named n_2 but in reality these two measurements belong to the same vehicle detector station. We make no additional assumptions about the additive artificial error terms. Hence they can, in theory, be any function of time.

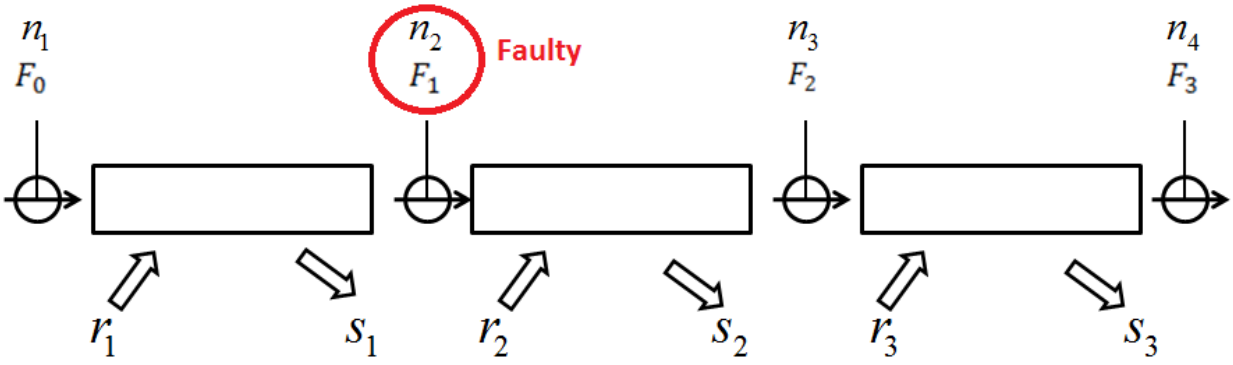

Figure 4.8: Example Ramp Configuration

We have:

$$f_1^m = f_1^T + \phi_1$$

$$n_2^m = n_2^T + \psi_2$$

where superscript m stands for measured and T stands for truth, i.e. the simulated artificial data. In the presence of the artificial faults, the dynamics of the plant, model and residuals in free flow evolve according to the following equations:

Plant:

$$\dot{n}_1 = -v_1 n_1 + f_0 + r_1$$

$$\dot{n}_2 = -v_2 n_2 + v_1 n_1 - s_1 + r_2$$

$$\dot{n}_3 = -v_3 n_3 + v_2 n_2 - s_2 + r_3$$

Model:

$$\dot{\hat{n}}_1 = -v_1 \hat{n}_1 + f_0^m + \hat{r}_1$$

$$\dot{\hat{n}}_2 = -v_2 \hat{n}_2 + v_1 \hat{n}_1 - \hat{s}_1 + \hat{r}_2$$

$$\dot{\hat{n}}_3 = -v_3 \hat{n}_3 + v_2 \hat{n}_2 - \hat{s}_2 + \hat{r}_3$$

Residuals:

$$\dot{\tilde{n}}_1 = -v_1 \tilde{n}_1 + \tilde{r}_1$$

$$\dot{\tilde{n}}_2 = -v_2 \hat{n}_2 + v_2 \psi_2 + \dot{\psi}_2 + \tilde{r}_2 - \phi_1$$

$$\dot{\tilde{n}}_3 = -v_3 \tilde{n}_3 + \tilde{r}_3$$

where the hat accents are reserved for the model and the tilde accents are reserved for residuals. In the residual equations, we have used the fact that under free flow conditions

$$v_1\hat{n}_1 - \hat{s}_1 = f_1^m$$

$$v_2\hat{n}_2 - \hat{s}_2 = f_2$$

On the other hand, during congestion, the dynamics evolve according to the equations below:

Plant:

$$\dot{n}_3 = w_3(n_3^J - n_3) - f_3 + r_3 - s_3 \text{ (not affected)}$$

$$\dot{n}_2 = w_2(n_2^J - n_2) - w_3(n_3^J - n_3) + r_2 - s_2$$

$$\dot{n}_1 = f_0^m - w_2(n_2^J - n_2) + r_1 - s_1$$

Model:

$$\dot{\hat{n}}_2 = w_2(n_2^J - \hat{n}_2) - f_2 + \hat{r}_2 - \hat{s}_2$$

$$\hat{n}_1 = f_0 - w_2(n_2^J - \hat{n}_2) + \hat{r}_1 - \hat{s}_1$$

Residuals:

$$\tilde{n}_2 = -\tilde{n}_2 w_2 + w_2 \psi_2 + \dot{\psi}_2 + \tilde{d}_2$$

$$\tilde{n}_1 = \tilde{n}_2 w_2 - w_2 \psi_2 + \tilde{d}_1$$

where $d_i = r_i - s_i$ is the total demand.

To simplify the analysis, only the two cases where the middle cell is either in free flow or in congestion are considered and it is assumed that the adjacent cell affected by the middle cell (which is the upstream cell when the middle cell is in congestion and the downstream cell when it is in free flow) is in the same mode as the middle cell. This does not compromise the algorithm's ability to detect faults since the instances when two successive cells are on different flow modes are transient and do not prevail for extended periods of time. After the imputation algorithm is applied, the model profiles converge to the following values during free flow and congestion respectively:

Free Flow:

$$\hat{n}_1 \rightarrow n_1$$

$$\hat{n}_2 \rightarrow n_2^m (= n_2^T + \psi_2)$$

$$\hat{f}_1 \rightarrow f_1^m (= f_1^T + \phi_1)$$

$$\hat{d}_1 \rightarrow d_1^T + \phi_1$$

$$\hat{d}_2 \rightarrow d_2^T - \phi_1 + \dot{\psi}_2$$

Congestion:

$$\hat{n}_1 \rightarrow n_1$$

$$\hat{n}_2 \rightarrow n_2^m (= n_2^T + \psi_2)$$

$$\hat{f}_1 \rightarrow f_1^m - \phi_1 - w_2\psi_2$$

$$\hat{d}_1 \rightarrow d_1^T - w_2\psi_2$$

$$\hat{d}_2 \rightarrow d_2^T + w_2\psi_2 + \dot{\psi}_2$$

It is seen that the faults manifest themselves in two distinct ways for this particular configuration of unknown ramp flows:

1. The measured and simulated flows discharged by cell 1 do not match during congestion

2. The estimated total demands for cells 1 and 2 show a significant instant change when the mode switches from free flow to congestion.

Figure 4.9 depicts the artificially generated flow, density and demand profiles in comparison to the simulated model based on these artificial data and the injected fault terms. The fault signatures that were derived above are clearly visible in the figures. In the demand (total ramp flows) plots, the red lines that show the "true" artificial ramp flows are not present in real applications but only the blue "imputed" ramp flows are available. Hence it is only the sudden switch in the profiles that provides the fault signature.

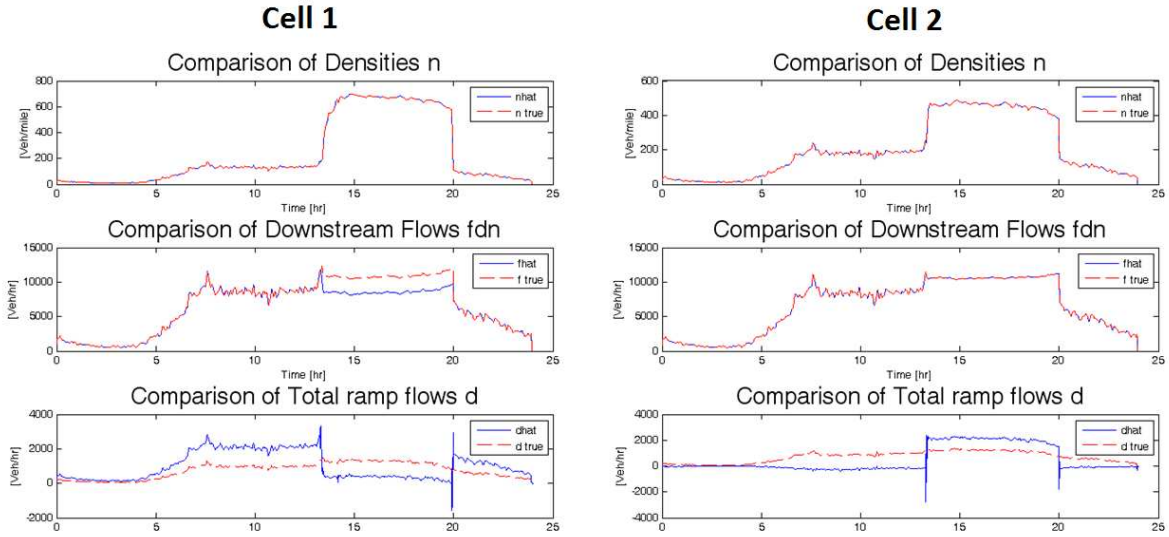

Figure 4.9: Fault Signatures due to Positive Bias in Density and Flow in an "All Ramps Present" configuration (Artificial Data)

There are 5 possible fault signatures that show up as residuals between the model and the plant:

1. Density mismatch during free flow

2. Density mismatch during congestion

3. Flow mismatch during free flow

4. Flow mismatch during congestion

5. "Jump" in estimated demand when there is a switch in flow conditions

Since the fault signatures are dependent on which ramps are to be estimated in a cell triplet, they were evaluated for all 64 possibilities of ramp configurations, similar to the example given above. The given example happens to be case 64, whereas case 1 is a triplet of cells with no ramps at all. The signatures for all ramp configuration cases are listed in the table in Appendix A.

4 modes of systematic fault were considered:

1. Positive Bias in Density

2. Negative Bias in Density

3. Positive Bias in Flow

4. Negative Bias in Flow

If the middle cell in a three cell cascade is subject to any of the above failure modes, a unique combination of the 5 signatures occur in all the three cells. In the given example this

combination is "Signature 4 in Cell 1 and Signature 5 in Cells 1 and 2", which are due to "Positive Bias in Density" and "Positive Bias in Flow".

The automated algorithm is essentially a look-up-table logic that takes the cells on the freeway in sets of 3, determines the specific ramp configuration and checks whether the density and flow residuals show characteristics pertaining to any of the failure modes listed above. The next chapter presents the results of this algorithm on real freeway sections with different ramp configurations and different fault signatures than the ones presented in this section.

Before we present the results, there are several important characteristics of the algorithm that need to be explained.

- **Thresholds:** As suggested by the discussion in section 4.1, most measurements are subject to varying degrees of bias. For this reason the algorithm requires thresholds to be set to accurately flag density and flow mismatches as actual fault signatures. The models calibrated by the TOPL project customarily aim for and consistently achieve mean absolute errors of around 3-5% and 5-10% for density and flow respectively (see equations (3.4) and (3.8)). Such calibration efforts can be reviewed in previous publications produced by the TOPL researchers, such as [24], [16] and [39]. In light of the extensive previous calibration experience (roughly around 50 calibrated models) on several different freeways the 3-5% and 5-10% errors for density and flow have been established as achievable and acceptable error figures. These error percentages are

averaged values over all sensors available on the modeled stretch of freeway. Therefore, the fault detection algorithm uses these figures as a starting point for determining the degree of mismatch for a particular sensor. Subsequently, **hard thresholds have been set at 3% for density and 10% for flows** at individual sensor locations.

- **Unobservable faults:** As shown in the example above, detection of fault is conditioned on the system switching its mode during the day. If this does not occur for a specific sensor location, i.e. the section does not get congested, the fault in the measurement is unobservable. In most cases, the model will be able to match the faulty measurements by feeding ramp flows that will match the measurement perfectly and the estimated ramp flow profiles will not exhibit any anomalies that can be picked up by the algorithm. On the other hand, since we are basing the accuracy of our model on how it compares to the measurements, this constitutes a case of 'ignorance being bliss" and gets ignored.

- **Known ramp flows:** Although the majority of ramp flows are not monitored and/or archived on California freeways, this is not always the case. In the rare cases where the ramp flow measurements are available, the algorithm assumes those measurements to be fault-free and classifies the geometry as a case where the given ramp does not exist, since it is not available to the imputation algorithm to be estimated. The decision logic treats the case accordingly when evaluating the mainline data for faults.

- **Faults in adjacent detectors:** It has been empirically established (similar to the

example above), that when two neighboring detectors both have a measurement bias, they nevertheless produce the corresponding fault signatures in the residuals and estimated demand profiles. These signatures do superpose but it is virtually impossible for them to cancel each other because this would require the fault signals to perfectly complement each other. The superposition of the fault signatures does not constitute a problem for the decision logic because it looks for an exact subset of signatures for each fault mode and does not get thrown off by the fact that there are additional signatures. In terms of classifying the fault however, it can report the section to exhibit both positive and negative bias in the same measurement. To summarize, adjacent faults do not deteriorate *fault detection* performance but they may hinder *fault identification*.

- **Algorithmic characterization of the "Jump" phenomenon:** In the automated algorithm, there is a need to algorithmically evaluate and categorize the jump phenomenon. Mathematically, this is done by taking the mean value of the demand profile during free flow and congestion and calculating the difference. If this difference is above the threshold of 1000 vehicles per hour, the demand profile is deemed to exhibit the jump phenomenon. A more exact approach would be to check if the profile produces a jump at the time instance when the flow conditions transition between free flow and congestion. However, at these exact time instances, we also expect some spikes in the profiles due to the discretization which can corrupt the identification of an authentic "jump". Figure 4.10 shows the comparison of these two approaches. It can be seen that

the former approach is less sensitive and thus more selective in terms of identifying the "jump". Here, the threshold of 1000 vehicles per hour is a rather conservative value. This figure is also based on previous model calibration and fault identification efforts. Mainline flows have their maximum values at approximately 8000 vehicles per hour when averaged over multiple days and compared to this average value, a difference of 1000 vehicles per hour is a conservative yet evidently effective threshold for detecting faults.

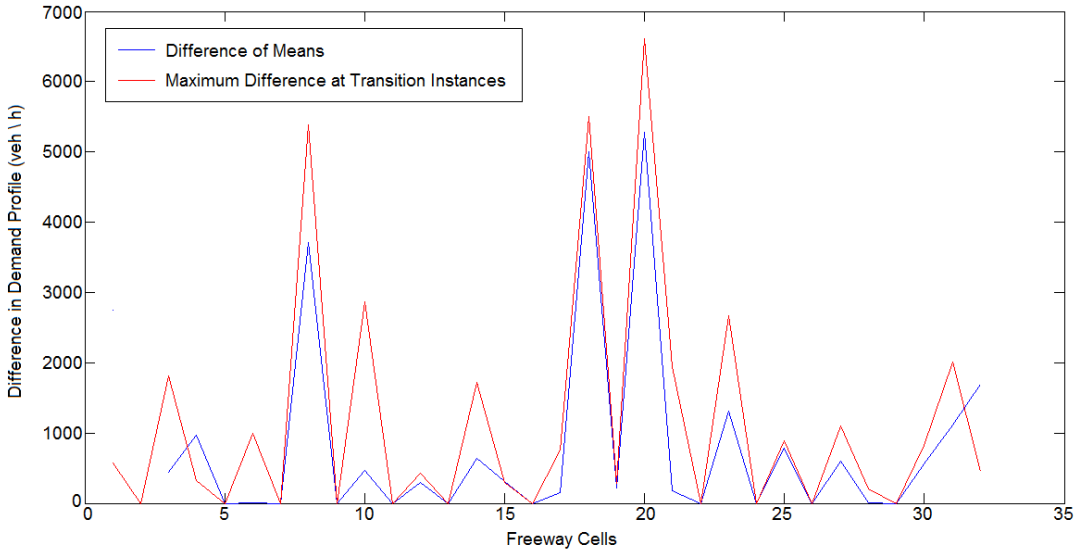

Figure 4.10: Severity of the jump phenomenon in the estimated demand profiles on an example freeway section

- **Arbitrary Fault Profiles:** When the fault signals that contaminate the measurements are not biases but rather arbitrary profiles superposed on the underlying correct measurements, they also produce fault signatures of the types enumerated above. An

example of such a fault would be encountered when a detector is reporting measurements from the opposite direction of the freeway and is highly inconsistent with its neighboring detectors. Such faulty measurements produce similar residuals, and in fact these residuals are much more severe and are easily identifiable with most threshold settings. These often trigger multiple fault modes that contradict each other such as "Negative Bias in Density" and "Positive Bias in Density" simultaneously. This is not surprising since an arbitrarily contaminated measurement can be both higher or lower than the true value at different time instants throughout the day.

The next section introduces the default post-processing and fault handling procedures that complete the picture for the automated fault diagnosis scheme.

4.4 FAULT HANDLING

Once the faulty detectors are accurately flagged, the next step is to analyze and try to figure out how to handle the fault. For the sake of complete automation, a post-processing unit deploys a default scheme of fault handling, which is to be introduced shortly. Here, it is important to note that the performance of this automated scheme will be inherently poor when compared to user-directed fault handling. The automated procedure discards the flagged bad detectors completely whereas in reality further analysis may reveal that the faulty detector was not necessarily reporting inaccurate data but instead it was placed at the incorrect location during the specification of the freeway geometry, detailed in section

3.1.

Due to the unknown inputs, a complete estimation of the fault signatures is not possible. Instead, faults can only be detected, as shown in the previous section. The automated fault handling approach therefore is to discard the bad detector from the analysis completely. This method is labeled as *fault exclusion* in the established fault diagnosis literature.

An important caveat of discarding bad detectors is the fact that data, which already are in short supply to begin with (about an average of 1.5 healthy detectors per mile on the modeled freeways), are being completely disregarded once flagged as faulty. Similar to the setting of thresholds, mentioned in the previous section, a cost vs. benefit analysis is required in an effort to be responsible with scarce resources. Every excluded detector reduces the model complexity and accuracy. So, although we want to exclude faulty detectors, we don't want to reduce our model too much by doing so. The thresholds mentioned in the previous section are set based on previous calibration experience and if for a certain freeway section and a given day they tend to be too strict, too many detectors (up to 50% in extreme cases) may be deemed as faulty by the decision logic. In such cases it behooves the user to either adjust the thresholds manually or, as it is more often the case, take a closer look at sensor placement on the freeway geometry. When too many detectors are flagged, and if they also tend to be succeeding each other in a certain portion of the freeway, it is usually a sign of a structural problem with the archived data. Two such cases were encountered in the calibrated freeways so far. One instance was in the modeling of the Eastbound Interstate 80, in the portion between Emeryville and El Cerrito, encompassing Berkeley. Eight mainline detectors in this

stretch were found to be reporting data from the loops installed on the opposite direction of the freeway, i.e. the westbound Interstate 80. A similar problem was encountered during the modeling of northbound Interstate 680. On a portion of the freeway to the immediate south of Walnut Creek, 4 successive detectors on the mainline were identified to be reporting data from the opposite direction loops.

To address the issue of too many flagged detectors without compromising the complete automation of the model building procedure, an heuristic post-processor has been devised. It is not uncommon for some flagged detectors to not have a very adverse contribution to the mean absolute error throughout the whole freeway. Following this argument, the flagged detectors can be ranked in terms of how badly they affect the mean absolute error of the complete model. Within this ranking, the default automatic threshold has been set at an arbitrary 0.5% contribution level, i.e. if the mean absolute error of the complete model decreases by only less than 0.5% in both density and flow errors, those detectors are not discarded. In effect, this serves as an additional threshold scheme acting on a post-processing level. However, attention must still be paid to the estimated unknown ramp flows in the vicinity of such detectors. Even though the mainline profiles may match more or less within acceptable accuracy, the estimated ramp flows may look unrealistic for detectors that are flagged but not discarded. The software is structured so that the user can specify and override which detectors to be excluded after the automated process determines the detectors to be eliminated from the analysis.

The exclusion of a detector is equivalent to the very common case where a detector is

already reported to be bad by PeMS or a detector is not installed at all at an important location on the freeway. This physical or artificial lack of mainline detectors coupled with the unobserved ramp flows gives rise to the creation of the so-called *Mega-Cells* within the model geometry. In effect, these Mega-Cells are ordinary links in both the LN-CTM and ACTM senses. However, both models assume the individual junction nodes (or the individual links for ACTM) to have at most one on-ramp and at most one off-ramp. When a detector is missing or gets discarded, it may be the case that more than one of each ramp falls between two adjacent working detectors. In such cases, the imputation algorithm can only estimate an effective total on-ramp flow and an effective total off-ramp flow no matter how many on-ramps and off-ramps are between the two working mainline detectors in reality. In other words, the model assumptions restrict the algorithm to represent an unobserved stretch of the freeway in a simplified manner, effectively deteriorating the geometric accuracy of the model. This theoretical formation of the so-called Mega-Cell is schematically demonstrated in figure 4.11.

This formulation of mega-cells, although essential for the estimation of the unknown ramp flows, poses the problem of geometric incompatibility of the model with the real world. TOPL strives for producing calibrated models for traffic researchers to run reliable simulations and test various planning and control strategies. To this goal, a model that represents multiple ramps that join the freeway at different locations as a single effective ramp connecting to the freeway at a single location is not an acceptable product. Therefore, further post-processing of the estimated ramp flows is required to revert the geometry back to its original form.

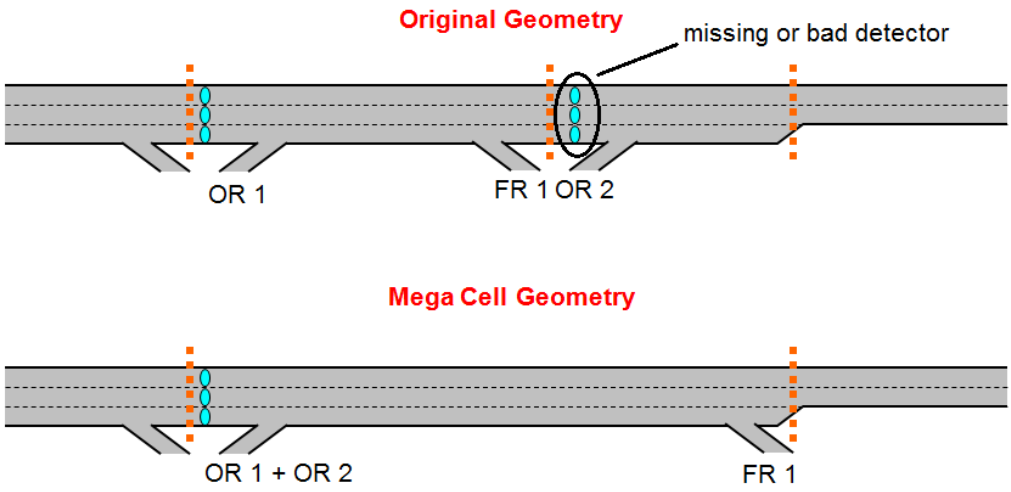

Figure 4.11: Forming of a Mega-Cell

This is achieved by converting the mega-cell to its original shape and splitting the estimated effective flows in such a way that the new geometrically accurate, no-mega-cells model agrees with the geometrically inaccurate model that was used in the imputation of effective ramp flows. Obviously, this agreement is only sought at the measurement locations which are part of the mega-cell model, i.e. the detectors that were used for the imputation.

This splitting of flows poses a challenge for the model construction process because the new ramp flows assigned to the separated ramps affect the state of all links in the rest of the model. This corresponds to going back to the top configuration in figure 4.11 from the bottom configuration. It is easy to see that an arbitrary splitting of the total effective on-ramp $OR1 + OR2$ may lead to extra congestion, if for instance the on-ramp $OR1$ is assigned zero flow as an extreme case and all of the estimated effective total on-ramp demand is

assigned to the on-ramp $OR2$. The link to which $OR2$ feeds to has become shorter and does not have an off-ramp anymore and hence can not accommodate the total demand without getting unrealistically congested. This extra congestion would also spill upstream and disrupt the rest of the model. Therefore, it is imperative to devise a separation scheme for such ramps that are bundled together due to forming of mega-cells.

A graphical representation of the mega-cell splitting problem is depicted in figure 4.12. Based on this figure, the problem can be stated as follows: For a given mega-cell $i+1$ with simulated profiles as given in the top half of the figure, reconstruct the profiles encircled in blue in the bottom half of the figure.

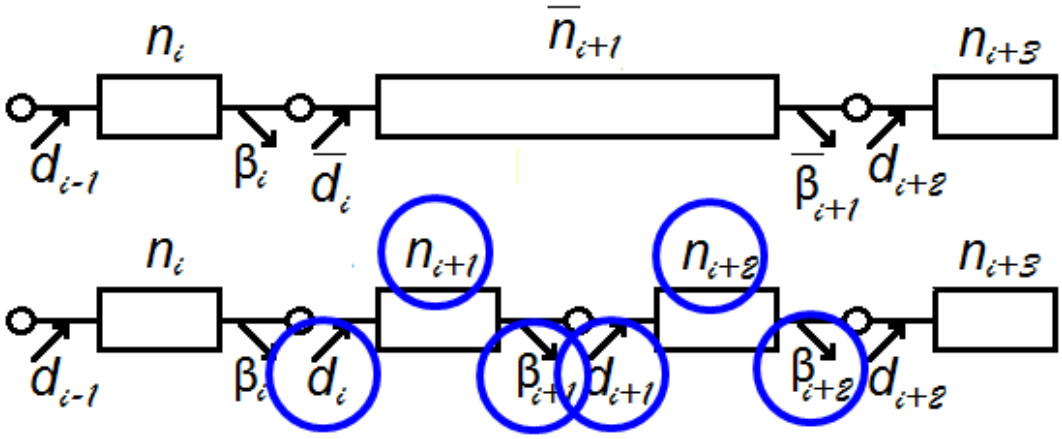

Figure 4.12: Mega-cell splitting problem

In the earlier stages of development, ad-hoc separation schemes were evaluated. These would start by separating the imputed total demands equally and adjusting the scaling according to the effect the new demands have on the rest of the model. If for example

an on-ramp demand is too high for the new geometry to accommodate without excessive congestion, the algorithm would decrease the portion that gets assigned to that particular on-ramp. Proceeding in this fashion the algorithm would iteratively come up with the best splitting proportion for the total demands while keeping the mainline density and flow profiles at the boundaries of the mega-cell the same. However, this approach did not prove successful in practice. Especially for the geometries where the original geometry has no ramp at the very beginning or the end of the mega-cell portion, as in the case with $FR1$ in figure 4.11, this algorithm could not provide a satisfactory solution.

Therefore, a re-imputation based approach has been adopted. Recall that the sole reason for skewing the original geometry is for imputing the unknown ramp flows and if the mainline measurements were appropriately available, no mega-cell would have been created to begin with. Following this reasoning, the alternative approach is based on finding an artificial *pseudo-measurement* profile for the mainline that would agree with the model and the measurements taken at the boundaries of the mega-cell. Given such a pseudo-measurement at the missing or bad detector location in figure 4.11, there would be no need to bunch the unknown ramps together and the imputation algorithm would estimate the ramp demands separately.

To find these so-called pseudo-measurements, we start with the calibrated model with mega-cells present. The terminology here bears significance. Due to the unknown inputs, the real state of the system is inherently *unobservable* when a healthy mainline detection is not present. Therefore, we can not proceed with a more rigorous state estimation approach.

Instead we formulate the problem as a linear program and search for a mainline density profile that agrees with the model assumptions and the model-generated flows and densities in the links neighboring the mega-cell. This linear optimization problem is posed as follows:

$$\min_{r_1,s_1,r_2,s_2,n_1^p,n_2^p,f_{mid}} \quad |n_1^p - n_1| + |n_2^p - n_2| \tag{4.17}$$

$$s.t. \quad An_1^p = f_0^{out} + r_1 - (f_{mid} + s_1) \qquad An_2^p = f_{mid} + r_2 - (f_2^{out} + s_2) \tag{4.18}$$

$$f_0^{out} + r_1 \leq w_1(\rho_1^J - n_1^p) \qquad f_{mid} + r_2 \leq w_2(\rho_2^J - n_2^p) \tag{4.19}$$

$$f_{mid} + s_1 \leq F_1 \qquad f_2^{out} + s_2 \leq F_2 \tag{4.20}$$

$$r_1 \leq f_0^{out}\frac{F_{r_1}}{F_1} \qquad r_2 \leq f_{mid}\frac{F_{r_2}}{F_2} \tag{4.21}$$

$$f_{mid} \geq 0 \qquad \rho_1^J \geq n_1^p \geq 0 \qquad \rho_2^J \geq n_2^p \geq 0 \tag{4.22}$$

$$r_1 \geq 0 \text{ (or } r_1 = 0) \;\; s_1 \geq 0 \text{ (or } s_1 = 0) \;\; r_2 \geq 0 \text{ (or } r_2 = 0) \;\; s_2 \geq 0 \text{ (or } s_2 = 0) \tag{4.23}$$

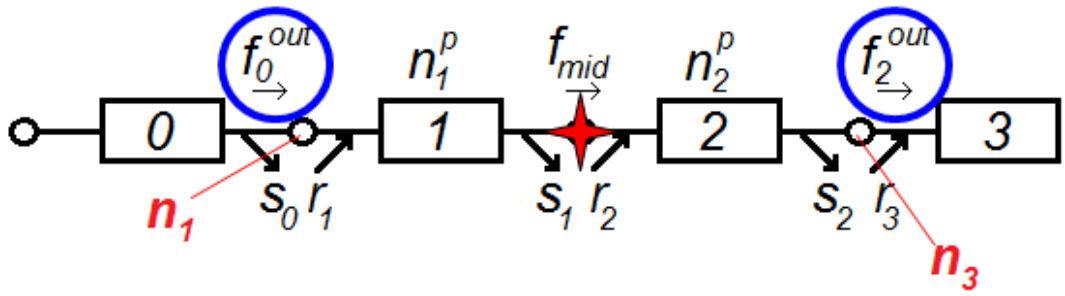

Figure 4.13: Mega-cell splitting problem - LP formulation

Figure 4.13 provides a visual clarification to the above problem, where it can be seen that n_1 and n_2 correspond to the measured density profiles and the flows on the boundaries of the

mega-cell, encircled in blue, are generated as the result of simulation using imputed effective ramp flows. In the equations above, n_1^p and n_2^p denote the pseudo-measurements of density in the two parts of the mega-cell that is being split, whereas n_1 and n_2 are the measured densities taken upstream and downstream of the mega-cell, respectively. The fundamental diagram parameters that are assigned to the two halves are the ones calibrated based on the data received at the measurement locations of n_1 and n_2 as well. The $|.|$ operator in the objective function 4.17 denotes l1-norm. All variables in the problem are vectors that span values for the given day of calibration, midnight to midnight. The set of constraints in the problem can be interpreted as a relaxed LN-CTM where we have relaxed the flow equations to bounding inequalities due to the unknown ramp flows. The constraints 4.18 specify the dynamics for the evolution of density in the two cells where the matrix A is the difference matrix specified as:

$$A = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & 0 & -1 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & -1 & 1 \end{bmatrix}$$

The constraints in 4.19 ensure that flows entering the cells do not exceed the supply during congestion, and similarly, discharge flows are bounded above by section capacities via the constraints in 4.20. The constraints 4.21 are imposed to ensure that ramp flows are

comparably smaller than mainline flows and so that the solution set does not include cases where flow into the downstream half is supplied predominantly by the ramp and the mainline flow between the cells hits zero. The rest of the constraints specify the trivial bounds within which the profiles should stay. Depending on the ramp geometry, constraints in 4.23 are set to either equal or greater than zero for each individual ramp inside the boundaries of the mega-cell.

The results of the above linear program are only the two pseudo-measurement profiles for the densities of the two new halves of the former mega-cell. The ramp flows that correspond to the solution of the linear program are not accurate estimates for the unknown ramp flows because of the relaxed nature of the model specification in the optimization. However, the main goal of the mega-cell de-construction is to obtain on-ramp demands and split ratio profiles for the individual ramps in the original geometry. This is accomplished by re-imputation using the pseudo-measurements of density along with the actual measurements where available. To summarize, the iterative algorithm that produces the separated ramp flow profiles is structured as follows:

Mega-cell Splitting Algorithm:

1. Use the linear program 4.17 - 4.23 to find pseudo-measured densities for the new cells

2. Feed pseudo-measurements into the imputation algorithm along with the actual measurements

3. Impute for the unknown ramp flows of the new configuration

4. Simulate to get new boundary flows at the boundaries of the mega-cell

5.
   - If after the splitting, the second cell is still a mega-cell (i.e. still has ramps that are bundled together), assign the density and the downstream flow of the first cell as the new upstream boundary conditions and go back to step 1.
   - Otherwise, proceed to the next mega-cell in the geometry.

Before proceeding to the results chapter, there is a need to clarify the choice of target density profile n_2 in the objective function 4.17 for the downstream half of the mega-cell. This target profile actually is the measurement that is assigned to the link downstream of the mega-cell. However, due to the relaxed formulation of the problem, the pseudo-measured density profile n_2^p needs to be associated with a realistic density profile. Otherwise, there is too much freedom in the free-flow mode, which gives rise to unrealistically high densities and flows at times during the night, where we expect both to be low. Therefore, a realistic density profile is required for the estimated n_2^p to hinge on.

Another practical modification that was implemented was to internalize the equality constraints 4.18 in the optimization by adding them to the objective function as follows:

$$\min_{r_1,s_1,r_2,s_2,n_1^p,n_2^p,f_{mid}} \quad |n_1^p - n_1| + |n_2^p - n_2| + |An_1^p - f_0^{out} - r_1 + (f_{mid} + s_1)| \qquad (4.24)$$
$$+ |An_2^p - f_{mid} - r_2 + (f_2^{out} + s_2)|$$

$s.t.$ 4.19 - 4.23

This relaxation is required because of the excessive number of equality constraints that these constraints impose on the problem, especially with small simulation time-steps such as 5 seconds which makes the A matrix, although sparse, a 17280-by-17279 matrix. The range of values for densities are often in the range of 50-300 vehicles per mile, whereas the common range for flows are around 1000-8000 vehicles per hour. Hence, the flows are penalized more severely by a factor of approximately 20 in the cost formulation (4.24).

The next chapter provides case studies on the modeling of two freeways that make use of all of the tools introduced so far.

Chapter 5

EXAMPLE CASE STUDIES AND RESULTS

5.1 INTERSTATE 210 IN SOUTHERN CALIFORNIA

In this section we present the results of the automated modeling and calibration procedure applied to a certain portion of the eastbound Interstate 210 in Southern California. The modeled section of the freeway spans the portion of the freeway starting near Pasadena and extending to near San Dimas, California. The first step of the model building procedure is to construct the geometry in Network Editor as shown in 5.1.

The modeled section is 27.3 miles long with 44 mainline detectors placed along the

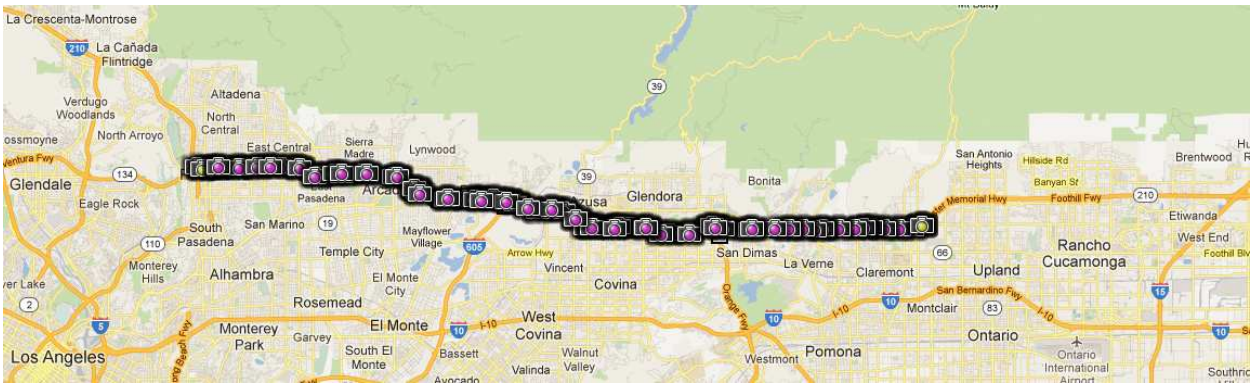

Figure 5.1: Modeled portion of the eastbound Interstate 210

mainline. There are 31 on-ramps and 27 off-ramps in this stretch. 26 of the on-ramps have working detectors whereas the remaining 5 on-ramp detectors were off-line for the day of the calibration. None of the off-ramp flows are observed. Hence, the 5 on-ramp and all of the off-ramp flows need to be imputed. There is a 24 hour active *HOV lane* (High Occupancy Vehicle Lane, also commonly known as *car pool lane*). This HOV lane has detectors separate from the mainline vehicle detector stations with individual IDs. Therefore, the model treats this lane as a separate road parallel to the mainline and specifies vehicle-type specific split ratio profiles for the mixing of flows. The key parameter for the model here is the *PPV (passenger per vehicle)* which specifies what portion of the vehicles are classified as HOV vehicles. This is customarily set to a nominal 1.2 which translates into 20% of the vehicles being HOV eligible. The TOPL simulation tool has the capability of keeping track of differen vehicle types such as single occupancy vehicles, high occupancy vehicles and trucks.

The freeway is represented by 66 links. These links need to be joined to form cells such that each cell has a working mainline detector associated with it as required by the

imputation algorithm. Since the imputation of unknown ramp flows hinges on the mainline data, the 66 link freeway representation needs to be condensed into a cell representation where a new cell is initiated at each detector location and the succeeding links are attached end to end until a new detector is reached along the freeway. Due to missing or not working detectors, this particular freeway needs to be reduced to a 30 cell representation. On the day of February 15^{th}, 2011, for which the following model calibration results are presented, 10 mainline detectors were reported to be down by PeMS and 4 detectors were discarded due to redundancy. When two detectors are in close proximity with no ramps in between, they do not provide additional information and one must define very short links to make use of both detectors. These short links can violate the CFL conditions during simulation and are therefore avoided. Figure 5.2 shows a spatio-temporal detector health contour plot for the year 2011, spanning the months between February and October. The horizontal axis is the space vertex listing the post miles of the detectors and the vertical axis is the temporal vertex showing the dates. Hence, every row in the plot corresponds to a day and every column to a detector. Detector health is represented as a percentage value which corresponds to the percentage of data received by PeMS from the detector for the given day. We can see from the plot that the month of February shows relatively better overall detector health.

The next step is to calibrate the fundamental diagrams for each of the detectors and feed the calibrated model into the imputation and fault detection algorithms. We first present the simulation results with imputed ramp flows without performing any fault detection on the available mainline data. We visualize the simulation results in form of the speed, density

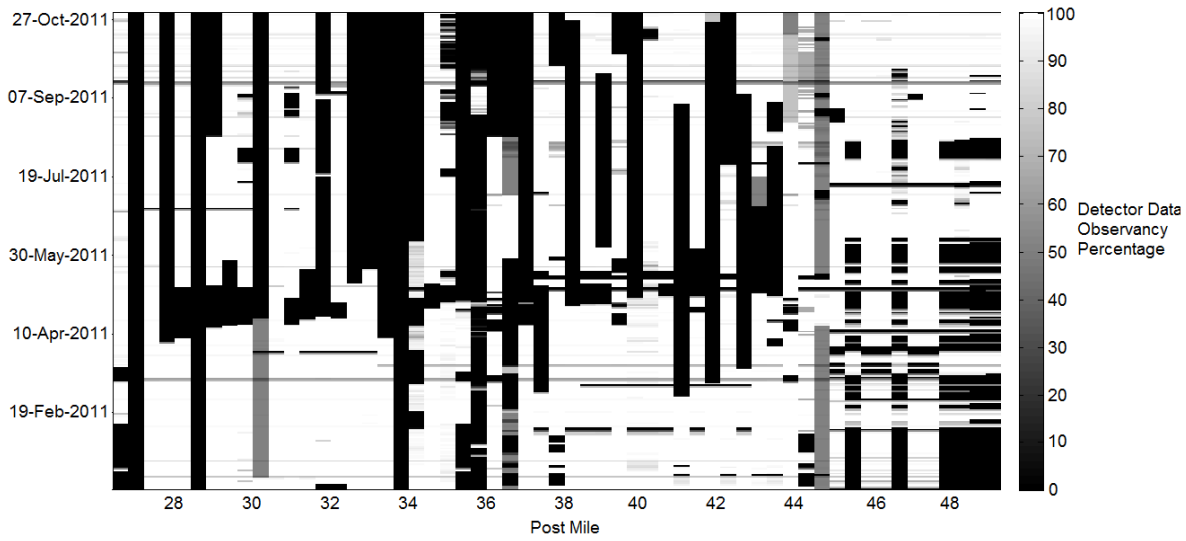

Figure 5.2: Detector Health Space Time Diagram for eastbound Interstate 210

and flow contour plots shown in figures 5.19, 5.20 and 5.21. These plots are similar to figure 5.2, where each column corresponds to a detector and each row to a 5 minute time instance during the day. Therefore, although the horizontal axis is labeled according to the post-miles of the detectors, the column widths are uniform and hence do not reflect actual detector spacing along the freeway. In the speed contours, the color bar is not included since there are only 3 shades of color associated with these plots: Black corresponds to speeds lower than 40 mph (congested flow), Grey corresponds to flows between 40 mph and 55 mph (dense flow) and White corresponds to speeds higher than 55 mph (free flow).

Before the faulty detectors are excluded, the mean absolute errors for density and flow are 5.63% and 10.53% respectively. Out of the 30 available detectors, 5 were flagged to be faulty and excluded. After removal of these bad detectors, the density and flow errors

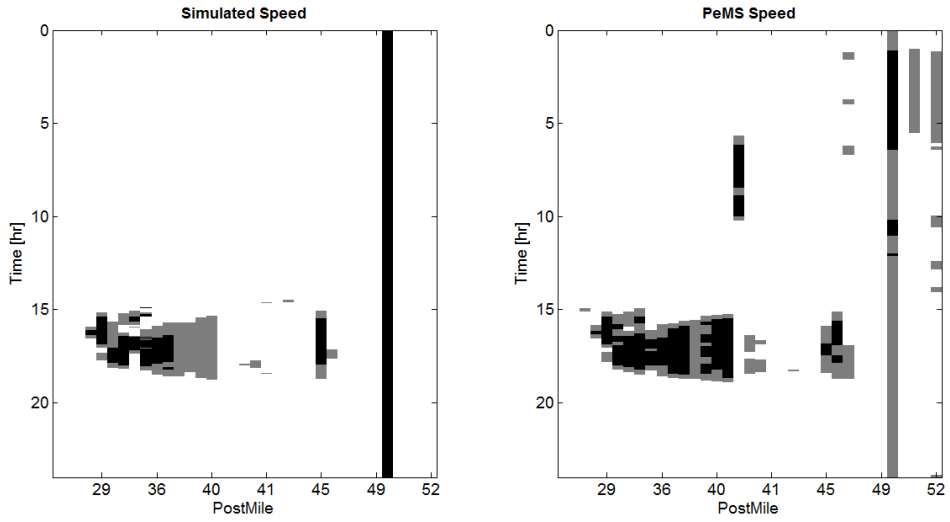

Figure 5.3: I-210E Speed Contours before Fault Detection / Exclusion

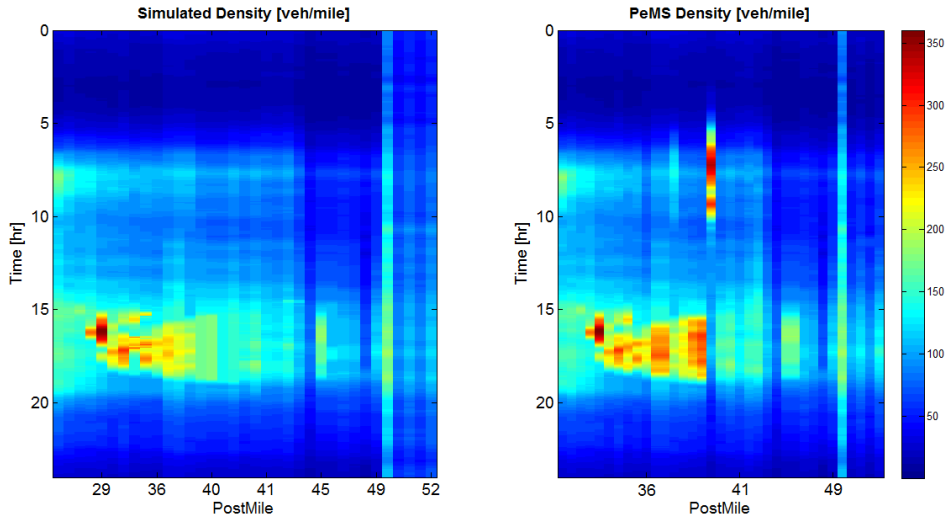

Figure 5.4: I-210E Density Contours before Fault Detection / Exclusion

decrease to 1.96% and 6.33% respectively. The speed, density and flow contours after fault detection and exclusion are given below.

In the latest configuration of the freeway, there exist 11 mega-cells to be de-constructed

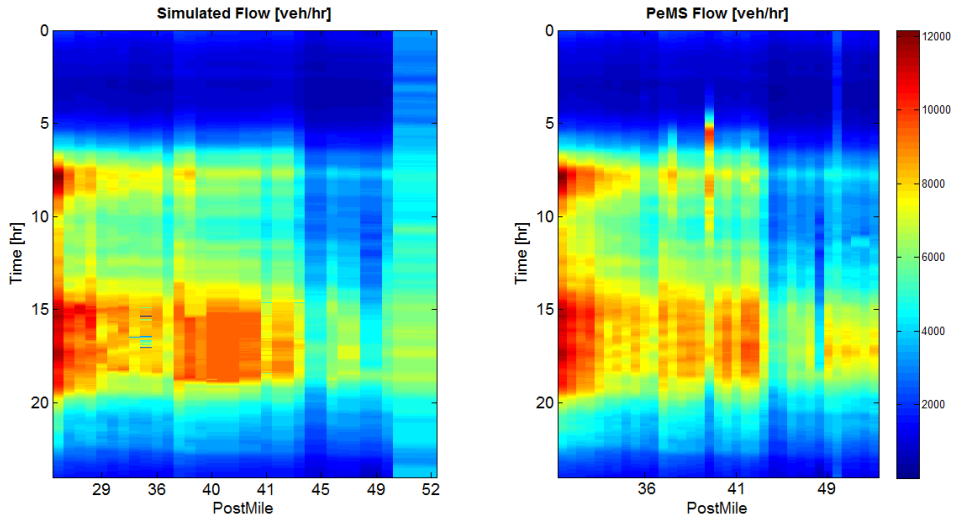

Figure 5.5: I-210E Flow Contours before Fault Detection / Exclusion

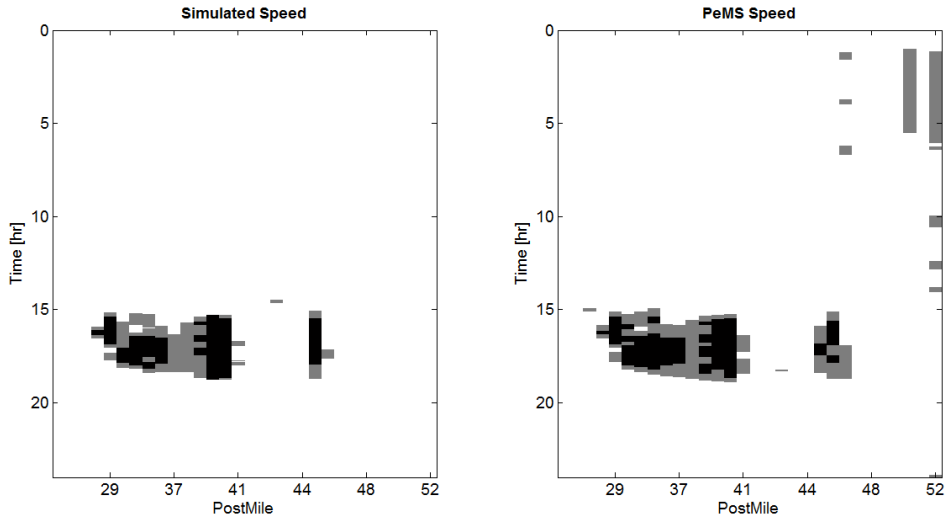

Figure 5.6: I-210E Speed Contours after Fault Detection / Exclusion (11 mega-cells present)

into their corresponding original geometries. The first step is estimating pseudo-measurements for each conjoined link inside the mega-cells. Recall that these are daily density profiles that are feasible in the LN-CTM sense and are as close as possible to the nearest available mea-

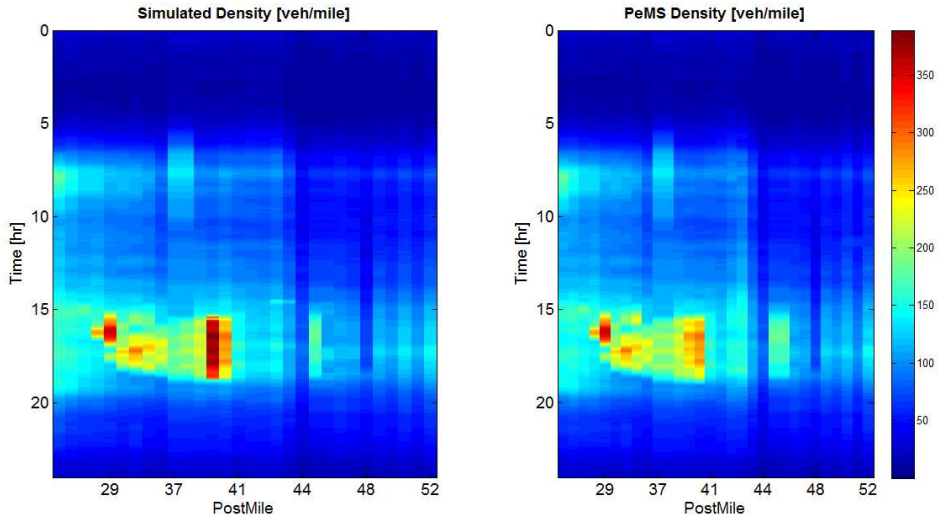

Figure 5.7: I-210E Density Contours after Fault Detection / Exclusion (11 mega-cells present)

surements. The objective function of the linear program (4.24) is also posed to preserve the flow profiles at the boundaries of the mega-cell. Flow and density profiles for an example section with the original configuration (i.e. with mega-cells still intact), and after re-imputation are shown in figures 5.9 to 5.12. That is to say, in the following plots, the green dotted lines are the actual measurements, the blue lines are the simulated profiles with mega-cells intact and the red lines are simulated profiles after the re-imputation using the pseudo-measurements obtained from the optimization problem.

To assess the resulting calibrated model after the de-construction of the mega-cells, we present the speed, density and flow contours with mega-cells intact and mega-cells de-constructed juxtaposed against the available data along the modeled freeway section. These plots are similar to the above plots except for the uniform column width for each cell. The

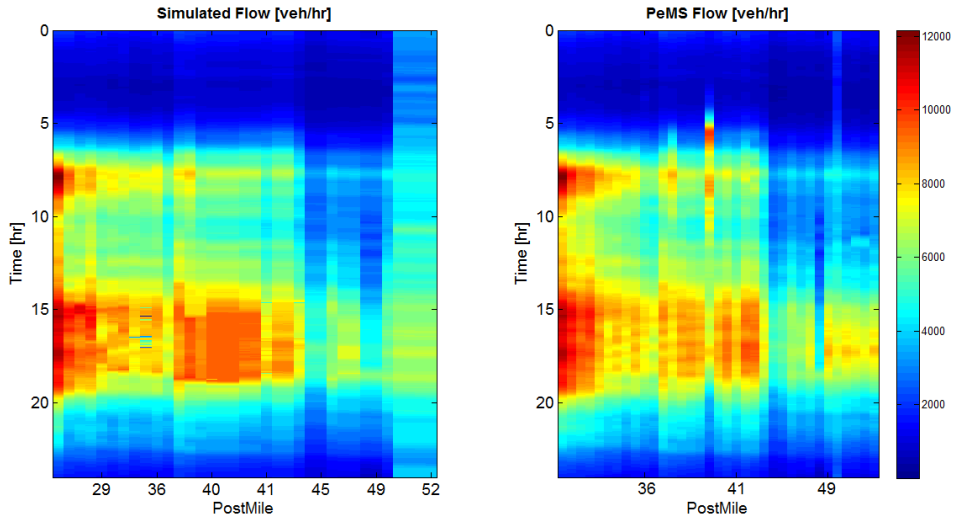

Figure 5.8: I-210E Flow Contours after Fault Detection / Exclusion (11 mega-cells present)

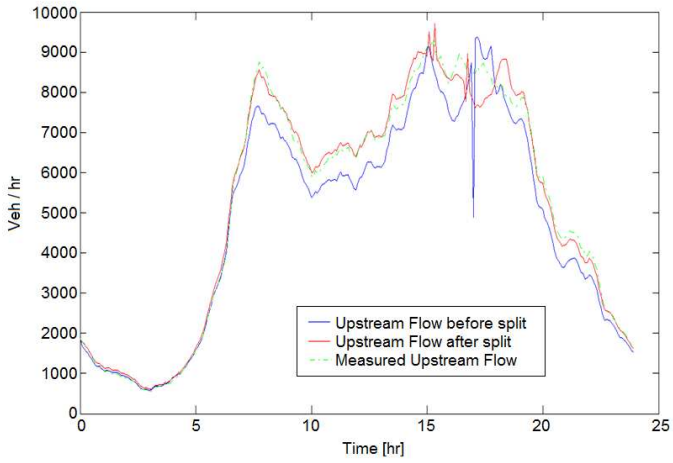

Figure 5.9: Flows at the upstream boundary of the example mega-cell

columns widths in these plots are directly proportional to cell lengths. That is to say, each cell is assigned a number of successive columns proportional to its size. This way, it is easy to see how big the mega-cells are and for how long a portion of the modeled freeway the lack of data extends. The top plots in all three of the figures reflect the taken measurements and

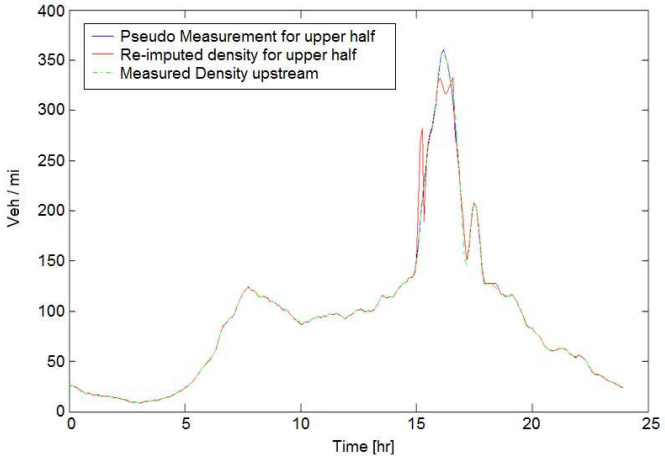

Figure 5.10: Densities of the upstream half of the example mega-cell

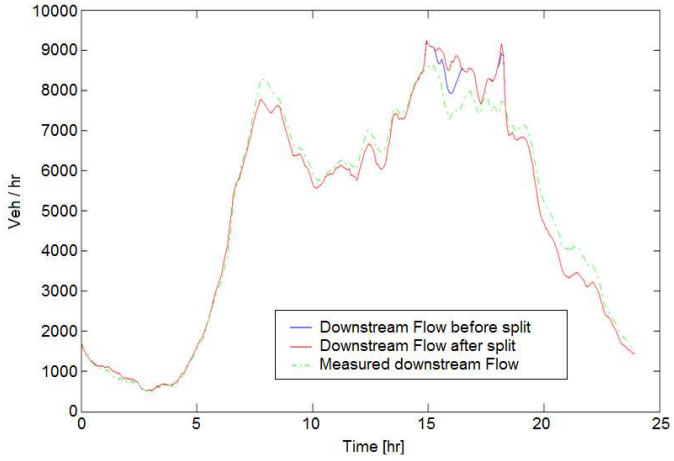

Figure 5.11: Flows at the downstream boundary of the example mega-cell

the blacked-out bands correspond to the portions where the data is bad or non-existent. The set of columns corresponding to mega-cells are bounded by boxes superposed on the plots and the numbers on top of these plots specify the number of original links that are bundled together to construct that specific mega-cell.

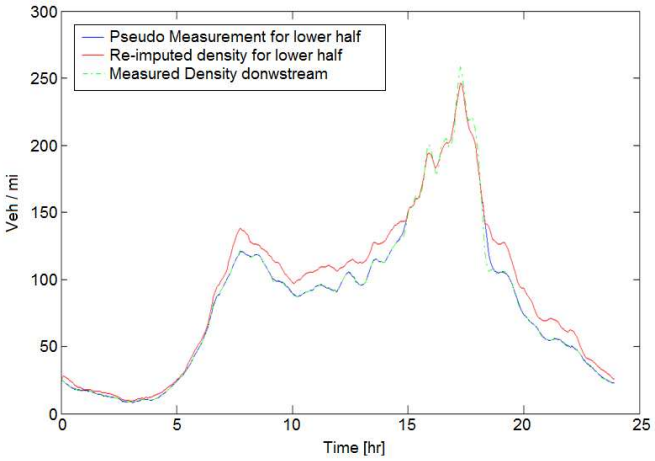

Figure 5.12: Densities of the downstream half of the example mega-cell

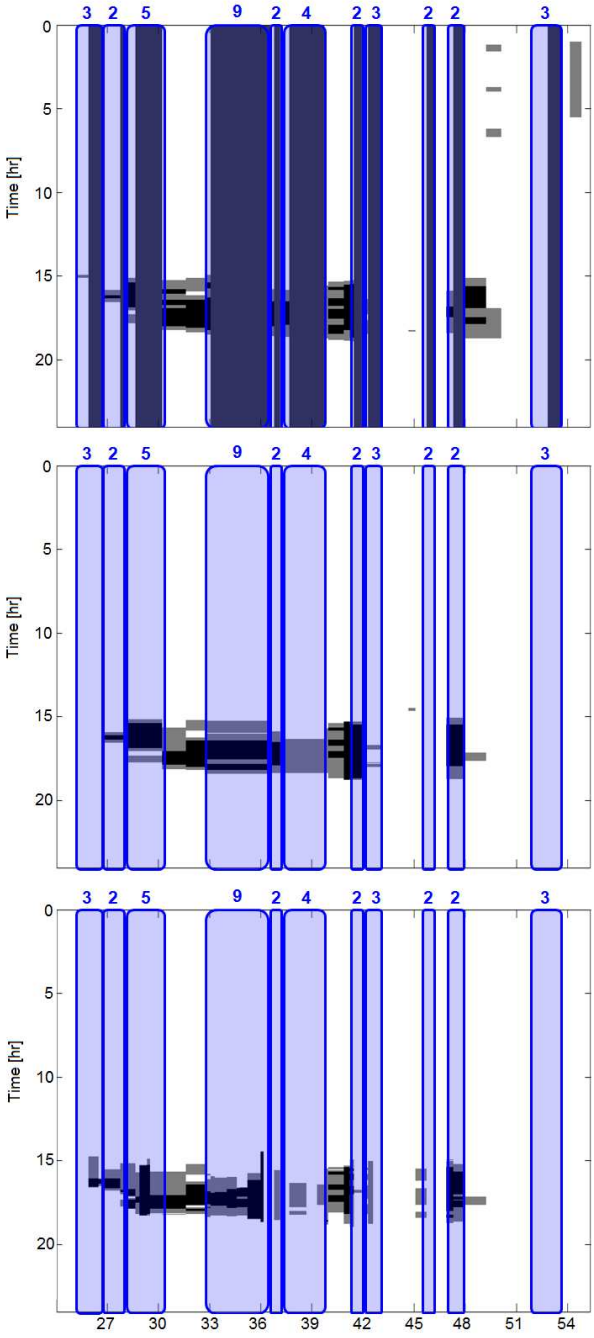

Figure 5.13: I-210E Speed Contours: Measurements (top), Mega-cells intact (middle), Mega-cells split (bottom)

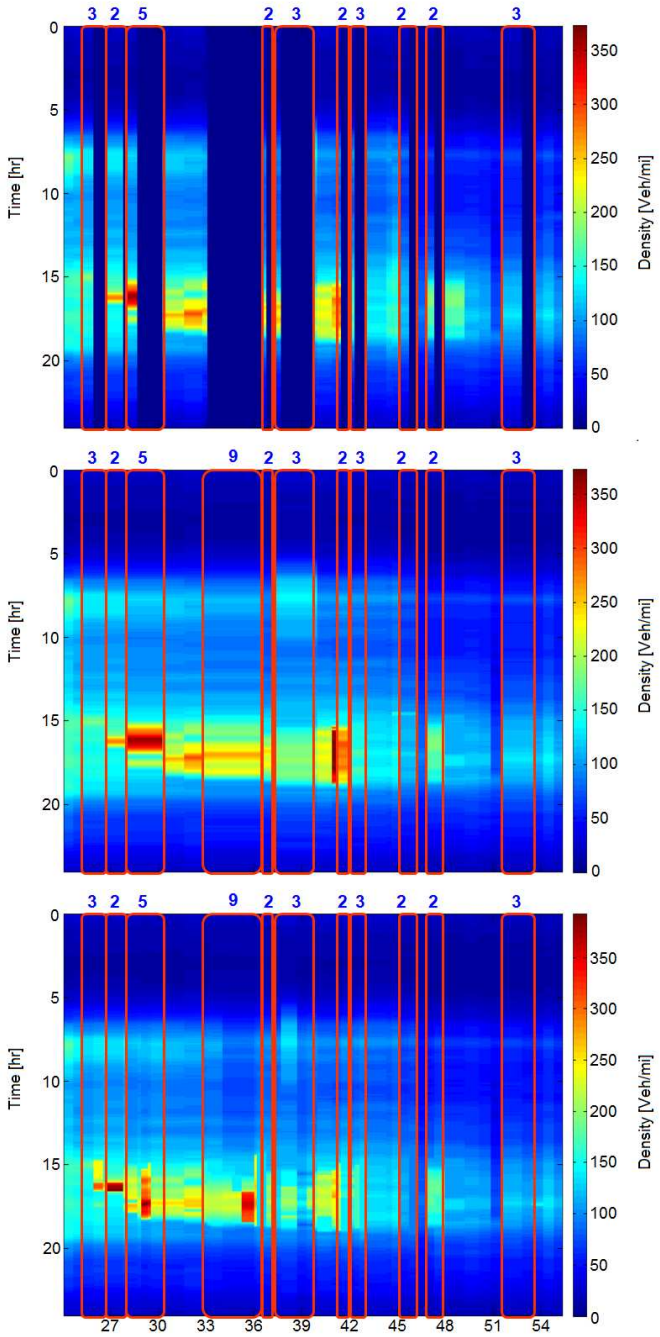

Figure 5.14: I-210E Density Contours: Measurements (top), Mega-cells intact (middle), Mega-cells split (bottom)

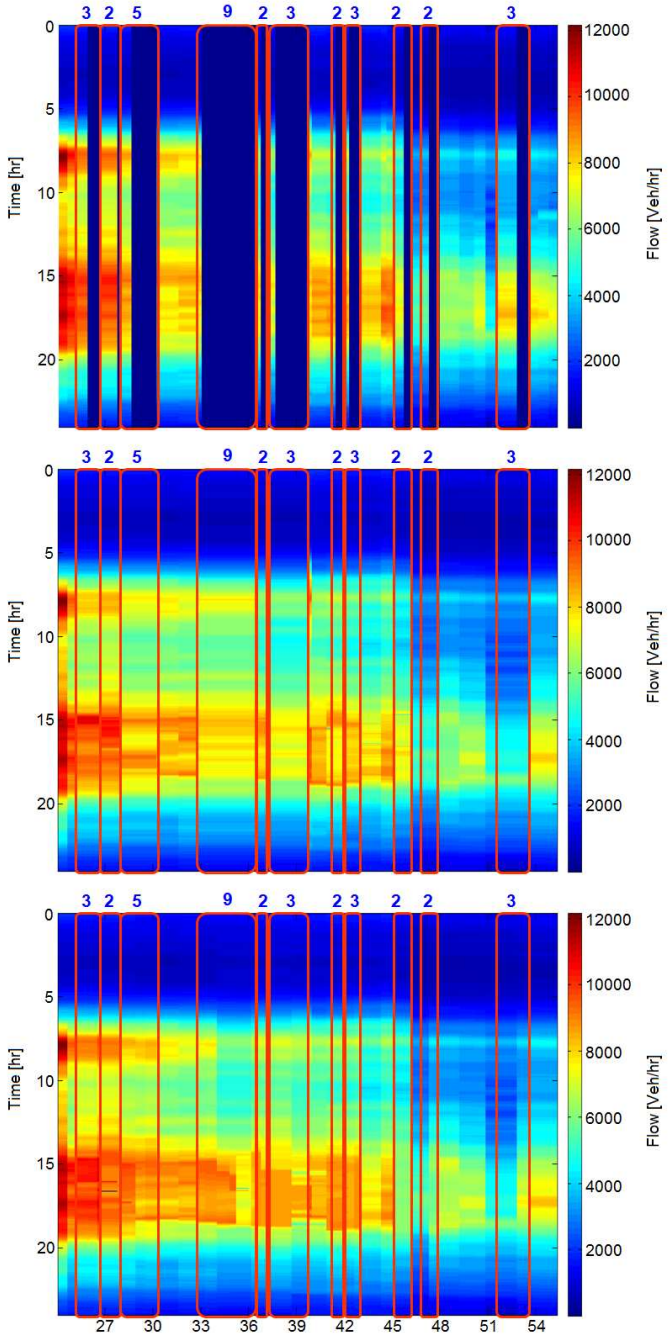

Figure 5.15: I-210E Flow Contours: Measurements (top), Mega-cells intact (middle), Mega-cells split (bottom)

Finally, figure 5.16 shows an example of the ramp flow profiles before and after splitting the mega-cell. The three plots on the left hand side show the total on-ramp, total off-ramp and net (on-ramp − off-ramp) ramp flows before and after the decomposition of the mega-cell from a single cell to 5 individual cells. The right hand side plots show the on-ramp and off-ramp flows that are assigned to each newly created individual cell in an alternating fashion. I.e., the top two profiles are the on-ramp and off-ramp of the first (most upstream) cell, the third and fourth plots belong to the second cell and so on. Due to the fact that this mega-cell consists of 5 individual cells, we get 10 such plots on the right hand side.

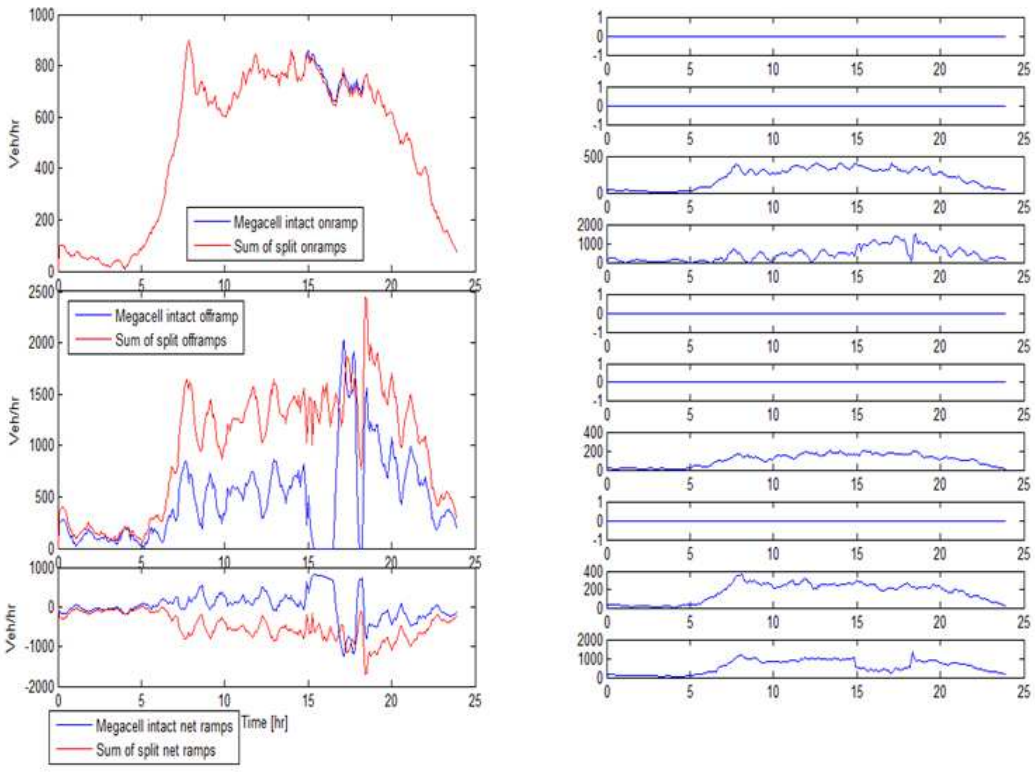

Figure 5.16: I-210E Ramp Flow and Demand Profiles for Mega-cell 3

5.2 INTERSTATE 80 IN NORTHERN CALIFORNIA

In this section we present the results of the automated modeling and calibration procedure applied to a certain portion of the eastbound Interstate 80 in the San Francisco Bay Area. The modeled section of the freeway spans the portion of the freeway starting near the end of the Bay Bridge in Oakland / Emeryville and extending to the beginning of the Carquinez Bridge in Crockett California. The portion of the modeled freeway section is first constructed in the Network Editor to produce the geometry shown in 5.17, including the mainline detectors placed along the freeway.

This modeled section is 21.6 miles long with 51 mainline detectors placed along the mainline. Contrary to the I-210 example, this freeway has no observation of ramp flows and for this reason all of the ramps need to be estimated using the imputation algorithm. Within the modeled portion there exist 25 on-ramps and 23 off-ramps. There is a time-of-day-activated HOV lane, active during the morning and evening peak hours of 5:00am to 10:00am and 3:00pm to 7:00pm, respectively. Unlike the I-210 case study, the detection on the HOV lane is not separate. Therefore the data reported by the single loops on the left-most lanes are used to calibrate the fundamental diagrams for the HOV lanes. In simulation, the activation and de-activation of the HOV lane is achieved by specifying correspondingly time-varying split ratios at the nodes where mainline and HOV lane flow interactions occur.

The freeway is represented by 79 links, most of which need to be joined into mega-cells

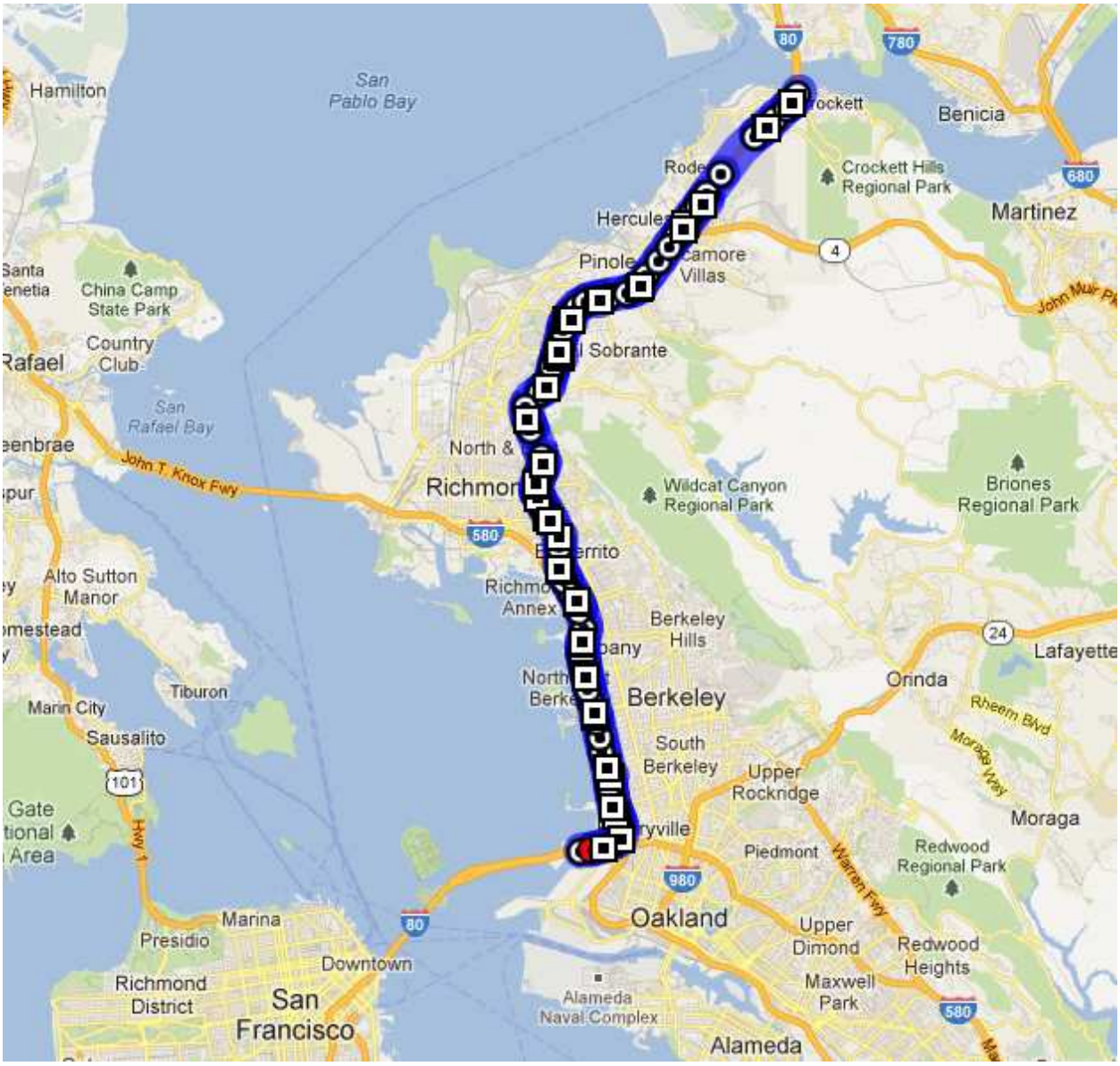

Figure 5.17: Modeled portion of the eastbound Interstate 80

for the imputation algorithm to estimate the unknown on-ramp demands and off-ramp split ratios. This is again due to either the poor health or a lack of mainline detection. Moreover, on the 2.6 mile stretch between the Powell street and the University Avenue exits, the PeMS

archives (a total of 8 detectors) were found to be reporting data from the opposite freeway direction. With no PeMS data available for this stretch, an alternate source of data was used. These 8 sensors happen to fall into the exact test bed of the Berkeley Highway Laboratory Project [1] and the archives from the year 2008 were available to us for use. Hence, the calibration and imputation were carried through based on 2008 data. In particular, the unknown ramp flows are imputed for the single day of August 21^{st}, 2008 in the following analysis. Looking at a 3 month period in 2008, the spatio-temporal detector status contours provide an overall picture of data health along the modeled freeway (5.18). The portion where the BHL data is used corresponds to the post mile range between 9.31 and 11.91. It can be seen in the plot that PeMS reports relatively healthy detection for this stretch.

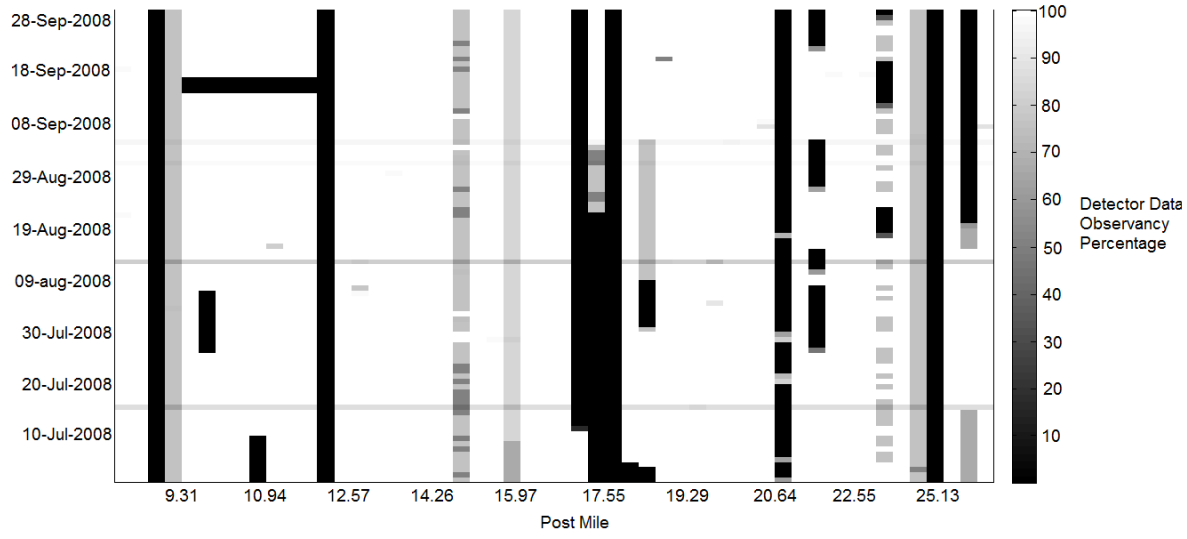

Figure 5.18: Detector Health Space Time Diagram for eastbound Interstate 80

On the particular day of August 21^{st} 2008, the PeMS analysis reports 14 of the 51

mainline detectors to be of poor health. Again, for the imputation algorithm, the 79 link freeway representation needs to be condensed into the cell representation. For this particular case, the 79 link representation is transformed into a 33 cell representation due to the 14 bad detectors reported by PeMS and a few other additional merges as a result of redundant detectors. It is sometimes the case that two detectors are placed within close proximity and do not provide additional information. In such cases, one of the two is discarded, especially if the distance between the two is shorter than the required minimum cell length for the CFL condition to be met by the simulation. Hence, in this particular example we have a 33 cell representation of the freeway although we have 37 healthy detectors available.

The next step is to calibrate the fundamental diagrams for each of the detectors and feed the calibrated model into the imputation and fault detection algorithms. We first present the simulation results with imputed ramp flows without performing any fault detection on the available mainline data. Similar to the 210 example, the speed, density and flow contour plots are shown in figures 5.19, 5.20 and 5.21.

Before the faulty detectors are excluded, the mean absolute errors for density and flow are 6.22% and 8.39% respectively. Out of the 33 available detectors, 4 were flagged to be faulty and excluded. After removal of these bad detectors, the density and flow errors decrease to 3.08% and 6.8% respectively. The speed, density and flow contours after fault detection and exclusion are given in figures 5.22 through 5.24 below.

In this latest configuration of the freeway, there exist 7 mega-cells to be de-constructed into their corresponding original geometries. We again present the speed, density and flow

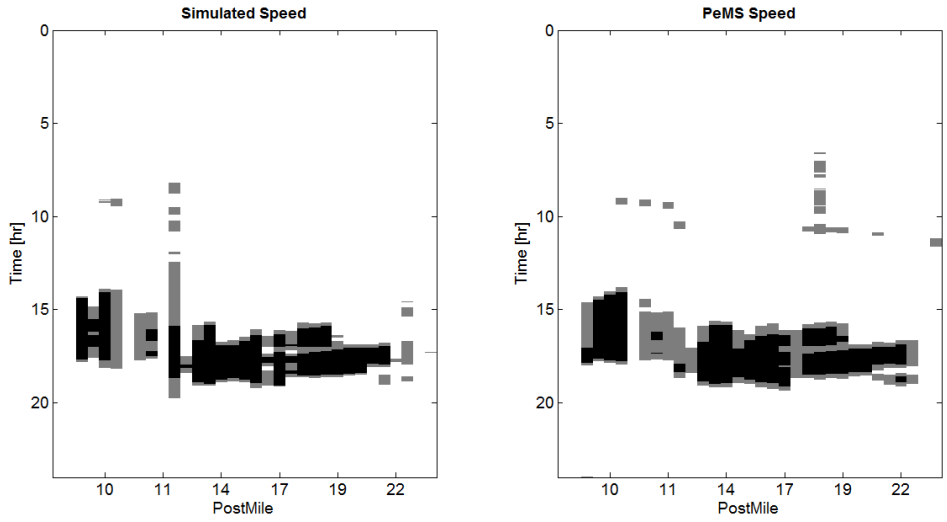

Figure 5.19: I-80E Speed Contours before Fault Detection / Exclusion

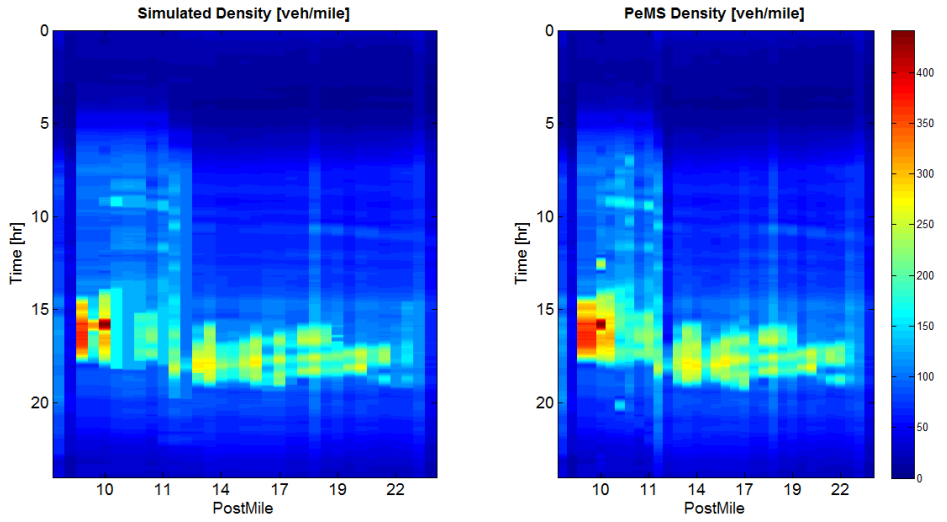

Figure 5.20: I-80E Density Contours before Fault Detection / Exclusion

contours with mega-cells intact and mega-cells de-constructed juxtaposed against the available data along the modeled freeway section. Contrary to uniform column width for each cell, the columns widths in these plots are directly proportional to cell lengths. The top

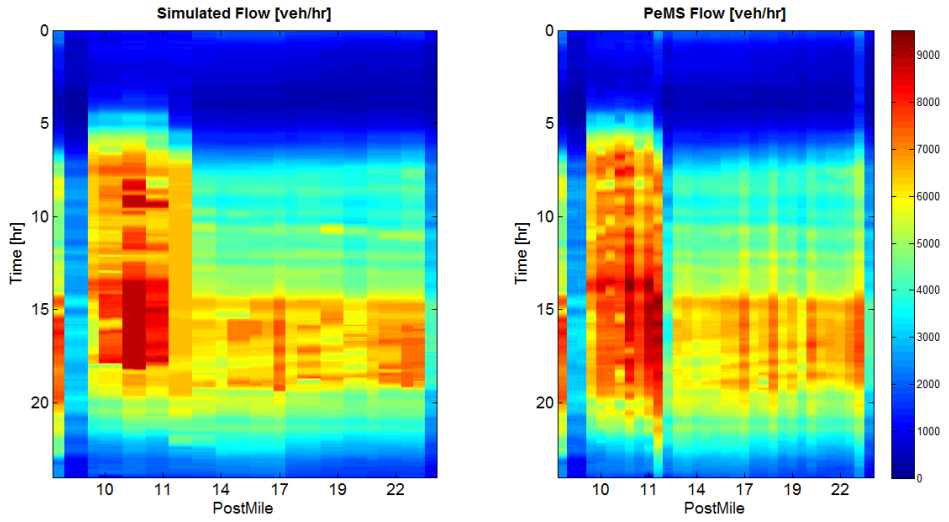

Figure 5.21: I-80E Flow Contours before Fault Detection / Exclusion

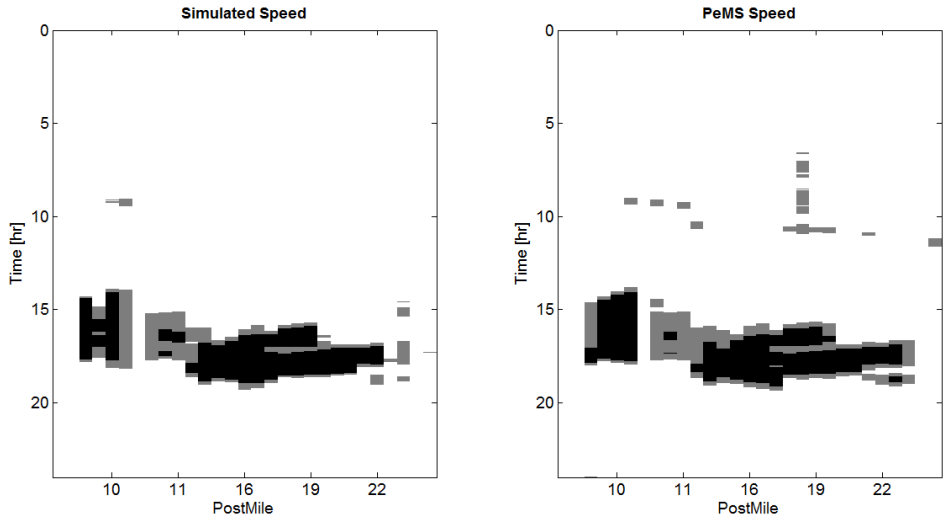

Figure 5.22: I-80E Speed Contours after Fault Detection / Exclusion (7 mega-cells present)

plots in all three of the figures reflect the taken measurements and the blacked-out bands correspond to the portions where the data is bad or non-existent. Finally, figure 5.28 shows an example of the ramp flow profiles before and after splitting the mega-cell.

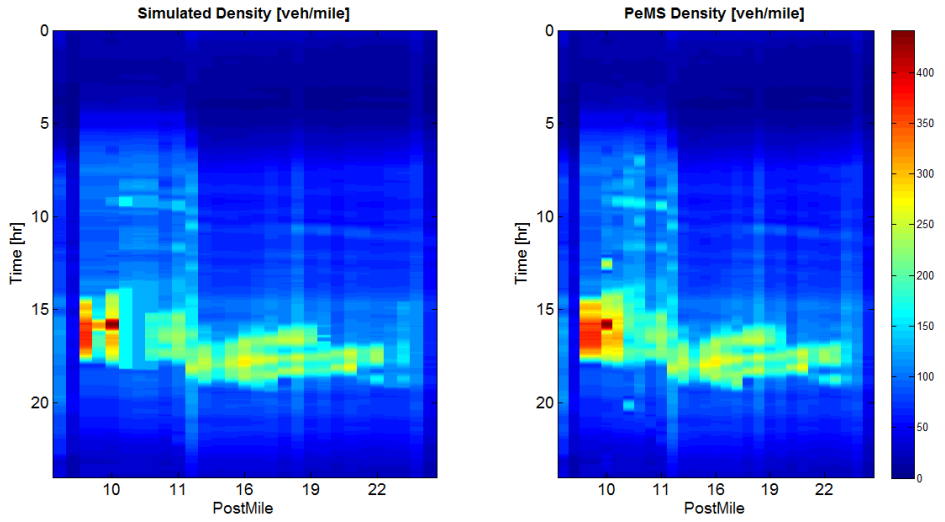

Figure 5.23: I-80E Density Contours after Fault Detection / Exclusion (7 mega-cells present)

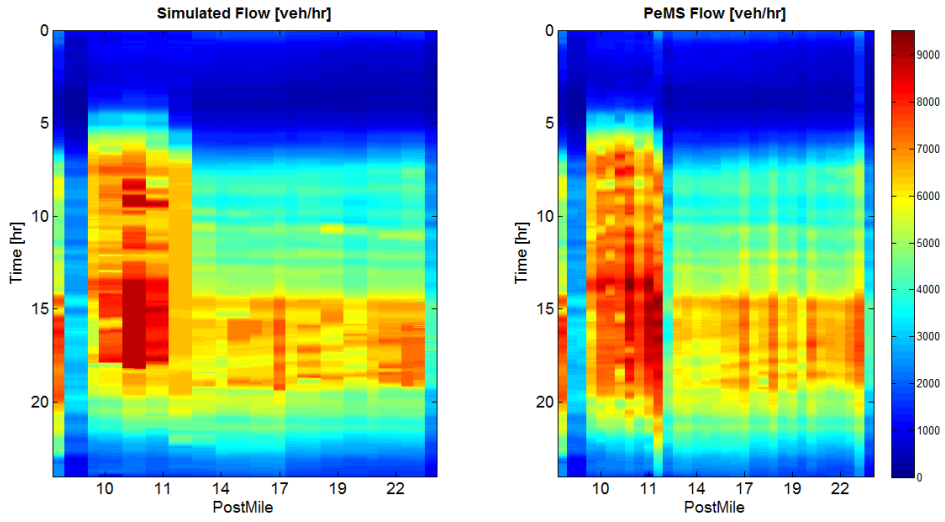

Figure 5.24: I-80E Flow Contours after Fault Detection / Exclusion (7 mega-cells present)

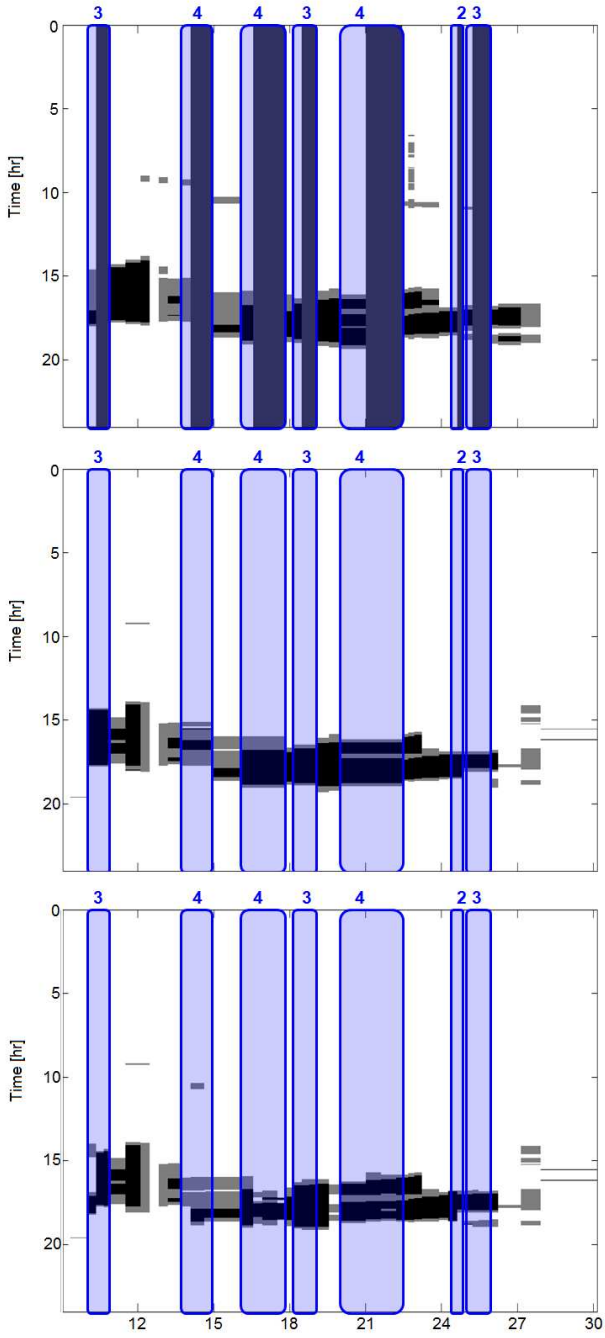

Figure 5.25: I-80E Speed Contours: Measurements (top), Mega-cells intact (middle), Mega-cells split (bottom)

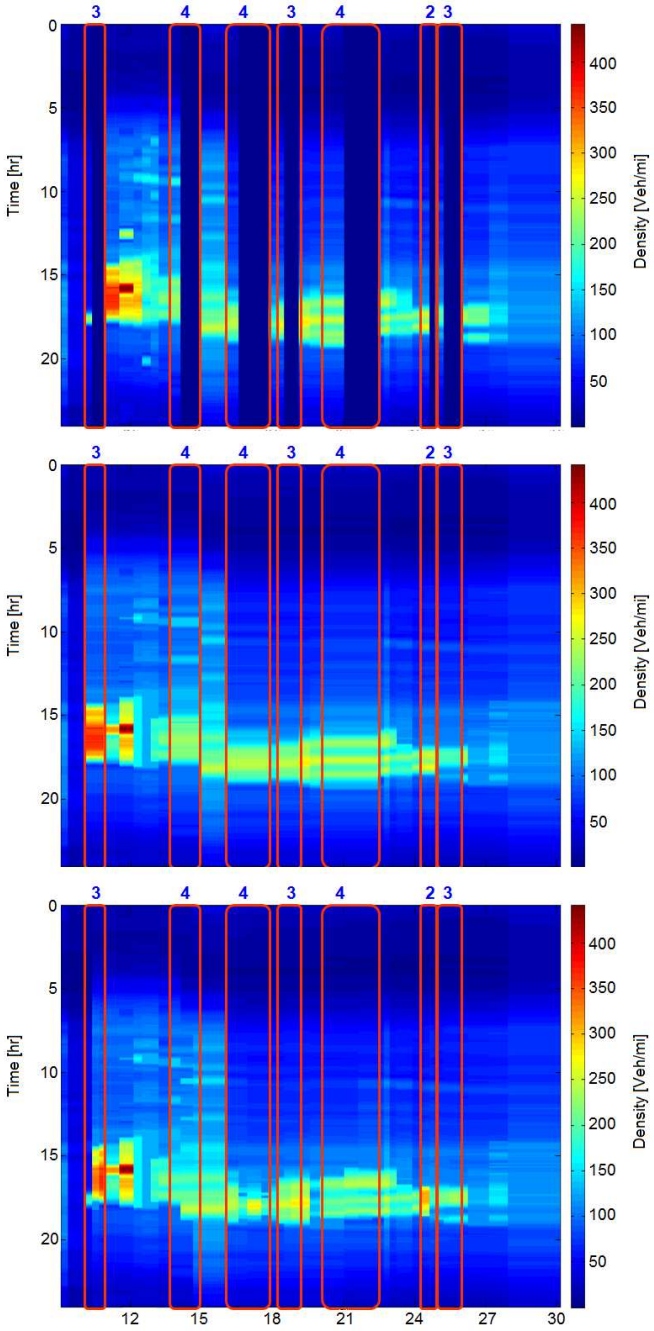

Figure 5.26: I-80E Density Contours: Measurements (top), Mega-cells intact (middle), Mega-cells split (bottom)

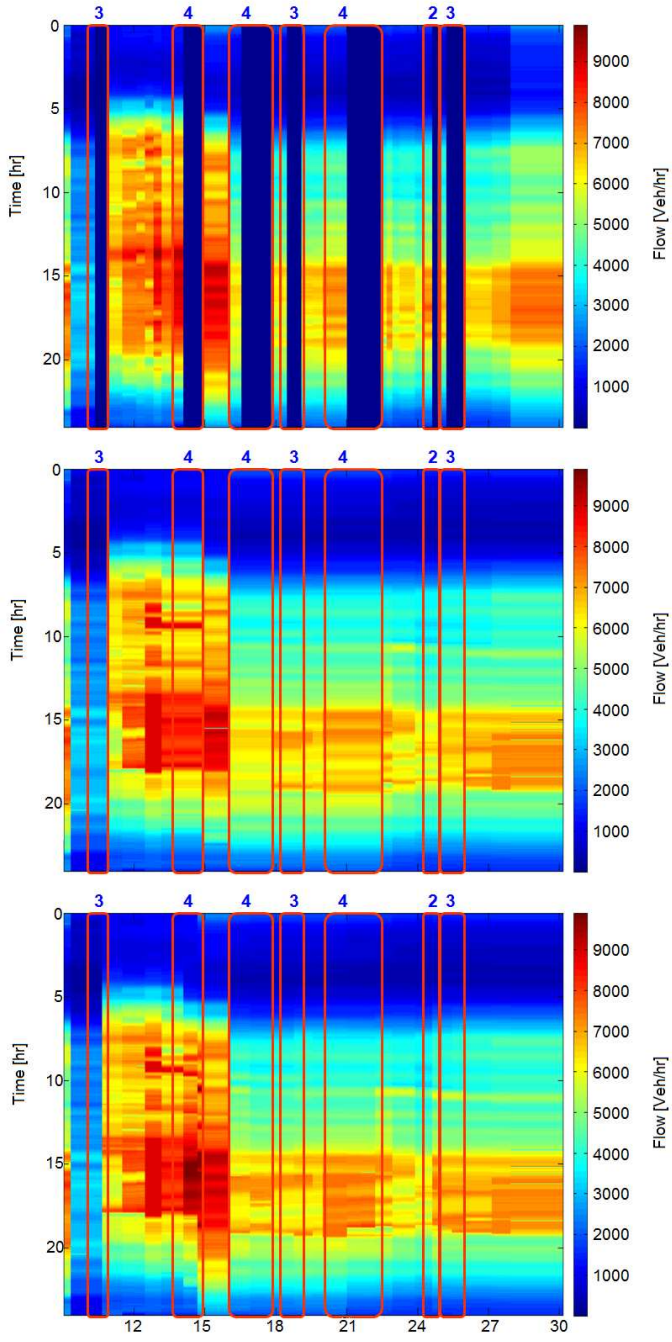

Figure 5.27: I-80E Flow Contours: Measurements (top), Mega-cells intact (middle), Mega-cells split (bottom)

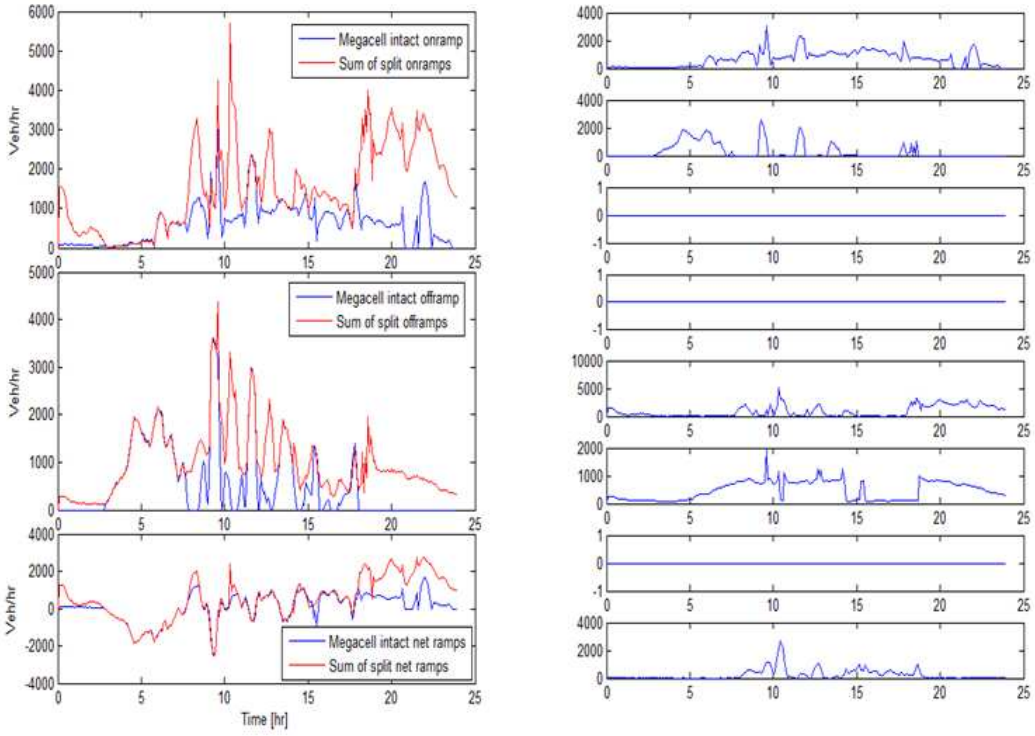

Figure 5.28: I-80E Ramp Flow and Demand Profiles for Mega-cell 2

In summary, the constructed model goes through the following steps:

1. Imputation of unknown ramp flows: In this step, freeway links are bunched together to build mega-cells due to PeMS-reported malfunctioning detectors

2. Fault detection: The fault detection algorithms flags and removes further bad detectors not reported by PeMS

3. Re-imputation of unknown ramp flows: Mega-cells are extended due to additional detectors being removed from analysis

4. Splitting of mega-cells: This step is done in two steps as explained in section 4.4. First, surrogate density measurements are found using the LP approach and second, imputing the unknown ramp flows for a third and last time using the pseudo measurements.

Table 5.1 summarizes the faulty detector information and the errors calculated for each of the models presented above, after each iteration.

Freeways	I-210E	I-80E
Total Number of detectors	36	51
Number of bad detectors reported by PeMS	6	14
Number of bad detectors removed by the fault detection algorithm	5	4
Density Error after first pass with mega-cells	5.63%	6.22%
Flow Error after first pass with mega-cells	10.53%	8.39%
Density Error after fault detection with mega-cells	1.96%	3.08%
Flow Error after fault detection with mega-cells	6.33%	6.80%
Density Error after mega-cell splitting	3.02%	3.69%
Flow Error after mega-cell splitting	5.82%	6.92%

Table 5.1: Model Calibration Results

Chapter 6

CONCLUSIONS AND DISCUSSION

This dissertation explored system identification, fault detection and fault handling methodologies developed for the automated freeway model building and calibration software, as part of the Tools for Operations Planning (TOPL) project. A macroscopic flow model is adopted in order to simplify the model building and provide very fast simulation performance to facilitate live decision support for traffic operators and engineers.

In chapter 2 we have presented the current data collection infrastructure that feeds the data required by the data driven algorithms presented in this work. It has been shown that despite extensive collection and archival capabilities, the quality and availability of data still remains the biggest challenge for the accurate and completely autonomous model construction effort. The main power of the macroscopic modeling approach is its swiftness of operation. Therefore, the challenges posed by poor data have to be tackled efficiently to leverage this convenient feature. Two imputation algorithms, explained in sections 3.3

and 4.2, are used to estimate the ramp flows which are usually not observed in California freeways. Fault detection and handling algorithms that further tackle low quality data problems have been introduced in sections 4.3 and 4.4. A look-up table logic automatically assesses the health of individual detectors along the modeled freeway and the removal of the faulty detectors from the analysis reduces mean absolute errors in density and flow by 40-50% and 20-40%, respectively. For extended portions of freeway with poor detection, a mega-cell decomposition methodology produces substitute measurements that agree with the rest of the healthy mainline detectors in order to keep the automated model building procedure functional even when good data are unavailable. Obviously, this solution should not be interpreted as anything more than a temporary fix. As shown in section 4.4, these pseudo-measurements do not provide estimates of the inherently unobservable state. The correct fix for this problem should be to revive, repair and install the required detectors or collect the necessary data by other means. Yet another algorithm introduced in section 3.6 was the Fundamental Diagram identification, which again addressed the wavering nature of the observed data by utilizing statistical tools to estimate the parameters.

Future research into the model-based sensor fault detection should be exploring the extension of the algorithm presented herein to more complex networks rather than unidirectional freeways. Obviously, this is a shortcoming of the underlying model-based imputation scheme rather than the fault detection algorithm itself. If the imputation algorithm were to be extended to more complex networks, the fault detection scheme would readily carry over. Advancements to the fault detection method can be made in terms of threshold setting. If

the model building process were to be inserted within a consolidated software framework, complete with continuous database access, within a few months there would be enough calibrated days to start establishing a threshold learning scheme for individual detectors. This scheme can subsequently be configured to automatically update and tune the thresholds as more calibrated days become available. Another improvement can be made towards removing the assumption that ramp measurements are fault-free whenever they are available. This is a simplifying assumption made while developing the fault detection framework and due to this assumption, any mismatch between the simulated and measured profiles is attributed to the mainline measurement, while in reality this may not necessarily be the case.

Bibliography

[1] Bhl website. http://bhl.path.berkeley.edu/project, accessed 08/24/2012.

[2] PeMS website. http://pems.dot.ca.gov, accessed 08/24/2012.

[3] TOPL network editor. http://vii.path.berkeley.edu:8097/NetworkEditor, accessed 08/24/2012.

[4] J.H. Banks. Freeway speed-flow concentration relationships: More evidence and interpretations. In *Transportation Research Record 1225*, pages 53–60, 1989.

[5] J.H. Banks. The two-capacity phenomenon: Some theoretical issues. In *Transportation Research Record 1320*, pages 83–90, 1991.

[6] J.H. Banks. Review of empirical research on congested freeway flow. *Transportation Research Record*, 1802:225–232, 2002.

[7] Transportation Research Board. *Highway Capacity Manual 2000*, December 2000.

[8] C. Chen, J. Kwon, J. Rice, A. Skabardonis, and P. Varaiya. Detecting errors and imputing missing data for single loop surveillance systems. In *Transportation Research Record*, volume 1855, pages 160–167, Washington DC, 2003. TRB, National Research Council.

[9] Cassidy, M. and R. Bertini. Some traffic features at freeway bottlenecks. *Transportation Research Part B*, 41B:82–95, 2007.

[10] C. Claudel and A. Bayen. Minimal error certificates for detection of faulty sensors using convex optimization. 2009. Allerton Conference on Communication and Control, Allerton, IL.

[11] D. Cleghorn, F. Hall, and D. Garbuio. Improved data screening techniques for freeway traffic management systems. In *Transportation Research Record 120*, pages 17–23, Washington, DC, 1991. TRB, National Research Council.

[12] G. Dervisoglu, G. Gomes, J. Kwon, R. Horowitz, and P. Varaiya. Automatic calibration of the fundamental diagram and empirical observations on capacity. Transportation Research Board, 88th Annual Meeting, 2009.

[13] X.S. Ding. *Model-Based Fault Diagnosis Techniques*. Springer, 2008.

[14] L.C. Edie. Car following and steady-state theory for non-congested traffic. In *Operations Research 9*, pages 66–76, 1961.

[15] L.C. Edie. *Flow Theories in Traffic Science*. Wiley, New York, 1974.

[16] A. Chow et al. TOPL: Tools for operational planning of transportation networks. 2008. ASME Dynamic Systems and Control Conference, Ann Arbor, Michigan, USA.

[17] W. Mauser F. Siebel. On the fundamental diagram of traffic flow. *ArXiv Condensed Matter e-prints*, 2008.

[18] J.H. Banks F.L. Hall, V.F. Hurdle. Synthesis of recent work on the nature of speed-flow and flow-occupancy (or density) relationships on freeways. In *Transportation Research Record 1365*, pages 12–18, Washington, DC, 1992. TRB, National Research Council.

[19] M.A. Gunter F.L. Hall. Further analysis of the flow-concentration relationship. In *Transportation Research Record 1091*, pages 1–9, 1986.

[20] M.A. Gunter F.L. Hall, B.L. Allen. Empirical analysis of freeway flow-density relationships. In *Transportation Research 20A*, pages 197–210, 1986.

[21] G. Gomes. *Optimization and microsimulation of on-ramp metering for congested freeways*. PhD thesis, University of California, Department of Mechanical Engineering, Berkeley, CA, 2004.

[22] G. Gomes and R. Horowitz. Optimal freeway ramp metering using the asymmetric cell transmission model. *Transportation Research, Part C*, 14(4):244–262, 2006.

[23] B. Greenshields. A study of traffic capacity. 1934. Proceedings of the 14th annual meeting of the Highway Research Board.

[24] J. Huper and et al. Macroscopic modeling and simulation of freeway traffic flow. 2009. 12th IFAC Symposium on Control in Transportation Systems.

[25] R. Isermann. Model-based fault detection and diagnosis - status and applications. *Annual Reviews In Control*, 29:71–85, 2005.

[26] J. Kwon, C. Chen, and P. Varaiya. Statistical methods for detecting spatial configuration errors in traffic surveillance. *Transportation Research Record*, 1870:124–132, 2005.

[27] J.A. Myers J. Treiterer. The hysteresis phenomenon in traffic flow. In *Proceedings of the 6th International Symposium on Transportation and Traffic Theory.*

[28] L. Jacobson, N. Nihan, and J. Bender. Detecting erroneous loop detector data in a freeway traffic management system. In *Transportation Research Record 1287*, pages 151–166, Washington, DC, 1990. TRB, National Research Council.

[29] Z. Jia, B. Coifman, C. Chen, and P. Varaiya. The PeMS algorithm for accurate, real-time estimates of g-factors and speeds from single loop detectors. In *2001 IEEE Intelligent Transportation Systems Proceedings*, pages 536–541, Oakland, CA, 2001.

[30] A.D. May J.S. Drake, J.L. Schofer. A statistical analysis of speed density hypotheses. In *Highway Research Record 154*, pages 53–87, 1967.

[31] M.J. Cassidy K. Chung, J. Rudjanakanoknad. Relation between traffic density and capacity drop at three freeway bottlenecks. *Transportation Research B*, 33B:25–42, 1999.

[32] R. Rajagopal P. Varaiya K. Kwong, R. Kavaler. Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors. *Transportation Research Part C: Emerging Technologies*, 17:586–606, 2009.

[33] B.S. Kerner. Dependence of empirical fundamental diagram on spatial-temporal traffic patterns features. *ArXiv Condensed Matter e-prints*, 2003.

[34] R. W. Koenker. *Quantile Regression*. Cambridge U. Press, 2005.

[35] A. Kurzhanskiy. *Modeling and Software Tools for Freeway Operational Planning*. PhD thesis, University of California, Berkeley, 2007.

[36] D.R.P. Gibson L.A. Klein, M.K. Mills. Traffic detector handbook: Third edition. Technical report, Federal Highway Administration, 2006.

[37] M. Lighthill and G. Whitham. On kinematic waves I: Flow movement in long rivers. II: A theory of traffic flow on long crowded roads. *Proc. Royal Society of London, Part A*, 229(1178):281–345, 1955.

[38] A. Muralidharan. *Tools for modeling and control of freeway networks*. PhD thesis, University of California, Berkeley, 2012.

[39] A. Muralidharan, G. Dervisoglu, and R. Horowitz. Freeway traffic flow simulation using the link node cell transmission model. American Control Conference, 2009.

[40] A. Muralidharan, G. Dervisoglu, and R. Horowitz. Probabilistic graphical models of fundamental diagram parameters for simulations of freeway traffic. *Transportation Research Record*, 2011.

[41] A. Muralidharan and R. Horowitz. Imputation of ramp flow data using the asymmetric cell transmission traffic flow model. ASME Dynamic Systems and Control Conference, 2009.

[42] C.F. Daganzo N. Gerolimnis. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B*, 42:759–770, 2008.

[43] H.J. Payne, E.D. Helfenbein, and H.C. Knobel. Development and testing of incident detection algorithms. Technical Report FHWA-RD-76-20, Federal Highway Administration, Washington DC, 1976.

[44] P. Balle R. Isermann. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice 5*, pages 709–719, 1997.

[45] R. Rajagopal, X. Nguyen, S. C. Ergen and P. Varaiya. Distributed online simultaneous fault detection for multiple sensors. In *Proceedings of the 2008 Intl. Conf. on Information Processing in Sensor Networks (IPSN 2008)*, 2008.

[46] P. Richards. Shock waves on the highway. *Operations Research*, 4(1):42–51, 1956.

[47] Schrank, D. and T. Lomax. The 2011 Annual Urban Mobility Report. Technical report, Texas Transportation Institute, 2011. http://mobility.tamu.edu.

[48] A.D. May S.M. Easa. Generalized procedures for estimating single- and two-regime traffic flow models. In *Transportation Research Record 772*, pages 24–37, 1980.

[49] R. E. Turochy and B. L. Smith. A new procedure for detector data screening in traffic management systems. In *Transportation Research Record 1727*, pages 127–131, Washington, DC, 1991. TRB, National Research Council.

[50] R. Horowitz A. Skabardonis X.Y. Lu, P. Varaiya. Fundamental diagram modeling and 374 analysis based on ngsim data. 2009. 12th IFAC Symposium on Control in Transportation Systems.

[51] H.M. Zhang. A mathematical theory of traffic hysteresis. *Transportation Research Part B*, 33:295–300, 1999.

Appendix A

Table of Fault Signatures

The table below lists the fault signatures corresponding to the four modes of fault considered:

1. Positive Bias in Density
2. Negative Bias in Density
3. Positive Bias in Flow
4. Negative Bias in Flow

The fault signatures are enumerated as follows:

1. Density mismatch during free flow
2. Density mismatch during congestion
3. Flow mismatch during free flow

4. Flow mismatch during congestion

5. "Jump" in estimated demand when there is a switch in flow conditions

The ramp configurations are encoded as binary numbers from 0 to 63. Each digit in the 6 digit number stands for a ramp. Read from left to right (which is the direction in which traffic is assumed to be flowing), the first entry corresponds to the onramp in the first cell, the second entry to its offramp, the third entry to the onramp of the second cell and so on. The fault is assumed to be isolated to the middle cell, i.e. the second cell only.

Ramp Configuration	Fault Modes	Fault Signatures	
0 0 0 0 0 0 0 0 0 0 0 1	(1) Positive Bias in Density	Cell 1	2,4
		Cell 2	2
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	2
		Cell 2	2
		Cell 3	1
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 0 0 0 1 0 0 0 0 0 1 1	(1) Positive Bias in Density	Cell 1	2,4
		Cell 2	2
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	2
		Cell 2	2
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 0 0 1 0 0 0 0 0 1 0 1	(1) Positive Bias in Density	Cell 1	2,4
		Cell 2	3,5
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	2
		Cell 2	1,3
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 0 0 1 1 0 0 0 0 1 1 1	(1) Positive Bias in Density	Cell 1	2,4
		Cell 2	3,5
		Cell 3	-
	(2) Negative Bias in Density	Cell 1	2
		Cell 2	1,3,5
		Cell 3	-
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 0 1 0 0 0 0 0 1 0 0 1	(1) Positive Bias in Density	Cell 1	2,4
		Cell 2	4
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	2
		Cell 2	3,5
		Cell 3	1
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 0 1 0 1 0 0 0 1 0 1 1	(1) Positive Bias in Density	Cell 1	2,4
		Cell 2	4
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	2
		Cell 2	3,5
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 0 1 1 0 0 0 0 1 1 0 1	(1) Positive Bias in Density	Cell 1	2,4
		Cell 2	5
		Cell 3	-
	(2) Negative Bias in Density	Cell 1	2
		Cell 2	3,5
		Cell 3	-
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 0 1 1 1 0 0 0 1 1 1 1	(1) Positive Bias in Density	Cell 1	2,4
		Cell 2	5
		Cell 3	-
	(2) Negative Bias in Density	Cell 1	2
		Cell 2	3,5
		Cell 3	-
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 1 0 0 0 0 0 1 0 0 0 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	2
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	2
		Cell 2	2
		Cell 3	1
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	1,3
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	2,3,5
		Cell 2	1,3
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 1 0 0 1 0 0 1 0 0 1 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	2
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	2
		Cell 2	2
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	1,3
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	2,3,5
		Cell 2	1,3
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 1 0 1 0 0 0 1 0 1 0 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	3,5
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	2,4,5
		Cell 2	1,3,5
		Cell 3	-
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	1,3
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	2,3,5
		Cell 2	1,3
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 1 0 1 1 0 0 1 0 1 1 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	3,5
		Cell 3	-
	(2) Negative Bias in Density	Cell 1	2,4,5
		Cell 2	1,3,5
		Cell 3	-
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	1,3
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	2,3,5
		Cell 2	1,3
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 1 1 0 0 0 0 1 1 0 0 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	4
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	2,4,5
		Cell 2	1,3,5
		Cell 3	1
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	5
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	2,3,5
		Cell 2	5
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 1 1 0 1 0 0 1 1 0 1 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	4
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	2,4,5
		Cell 2	2
		Cell 3	-
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	5
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	2,3,5
		Cell 2	5
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 1 1 1 0 0 0 1 1 1 0 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	5
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	2,4,5
		Cell 2	3
		Cell 3	1
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	5
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	2,3,5
		Cell 2	5
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
0 1 1 1 1 0 0 1 1 1 1 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	5
		Cell 3	-
	(2) Negative Bias in Density	Cell 1	2,4,5
		Cell 2	3
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	5
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	2,3,5
		Cell 2	5
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 0 0 0 0 0 1 0 0 0 0 1	(1) Positive Bias in Density	Cell 1	2,4,5
		Cell 2	2
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	4,5
		Cell 2	2
		Cell 3	1
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 0 0 0 1 0 1 0 0 0 1 1	(1) Positive Bias in Density	Cell 1	2,4,5
		Cell 2	2
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	4,5
		Cell 2	2
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 0 0 1 0 0 1 0 0 1 0 1	(1) Positive Bias in Density	Cell 1	2,4,5
		Cell 2	3,5
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	4,5
		Cell 2	1,3,5
		Cell 3	1
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 0 0 1 1 0 1 0 0 1 1 1	(1) Positive Bias in Density	Cell 1	2,4,5
		Cell 2	3,5
		Cell 3	-
	(2) Negative Bias in Density	Cell 1	4,5
		Cell 2	1,3,5
		Cell 3	-
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 0 1 0 0 0 1 0 1 0 0 1	(1) Positive Bias in Density	Cell 1	2,4,5
		Cell 2	4
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	4,5
		Cell 2	3
		Cell 3	1
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 0 1 0 1 0 1 0 1 0 1 1	(1) Positive Bias in Density	Cell 1	2,4,5
		Cell 2	4
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	4,5
		Cell 2	3
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 0 1 1 0 0 1 0 1 1 0 1	(1) Positive Bias in Density	Cell 1	2,4,5
		Cell 2	5
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	4,5
		Cell 2	3,5
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 0 1 1 1 0 1 0 1 1 1 1	(1) Positive Bias in Density	Cell 1	2,4,5
		Cell 2	5
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	4,5
		Cell 2	3,5
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	3,4
		Cell 2	-
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 1 0 0 0 0 1 1 0 0 0 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	2
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	-
		Cell 2	2
		Cell 3	1
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	1,3
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	-
		Cell 2	1,3
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 1 0 0 1 0 1 1 0 0 1 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	2
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	-
		Cell 2	2
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	1,3
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	-
		Cell 2	1,3
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 1 0 1 0 0 1 1 0 1 0 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	3,5
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	-
		Cell 2	1,3,5
		Cell 3	-
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	1,3
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	-
		Cell 2	1,3
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 1 0 1 1 0 1 1 0 1 1 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	3,5
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	-
		Cell 2	1,3,5
		Cell 3	-
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	1,3
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	-
		Cell 2	1,3
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 1 1 0 0 0 1 1 1 0 0 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	2
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	-
		Cell 2	2
		Cell 3	1
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	5
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	-
		Cell 2	5
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 1 1 0 1 0 1 1 1 0 1 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	4
		Cell 3	5
	(2) Negative Bias in Density	Cell 1	-
		Cell 2	2
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	5
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	-
		Cell 2	5
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 1 1 1 0 0 1 1 1 1 0 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	5
		Cell 3	1
	(2) Negative Bias in Density	Cell 1	-
		Cell 2	3,5
		Cell 3	5
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	5
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	-
		Cell 2	5
		Cell 3	-

Ramp Configuration	Fault Modes	Fault Signatures	
1 1 1 1 1 0 1 1 1 1 1 1	(1) Positive Bias in Density	Cell 1	4,5
		Cell 2	5
		Cell 3	-
	(2) Negative Bias in Density	Cell 1	-
		Cell 2	3,5
		Cell 3	-
	(3) Positive Bias in Flow	Cell 1	4,5
		Cell 2	5
		Cell 3	-
	(4) Negative Bias in Flow	Cell 1	-
		Cell 2	5
		Cell 3	-

Table A.1: Fault Signatures for Different Ramp Configurations