12th WCTR
World Conference on Transport Research
July 11-15, 2010 - Lisbon

# Experimental observations in freeway merges with ramp metering control

Nikolas Geroliminis[1, 2*], Anupam Srivastava[2], Panos Michalopoulos[2]

[1] Urban Transport Systems Laboratory, *Ecole Polytechnique Fédérale de Lausanne*
[2] *Department of Civil Engineering, University of Minnesota*

**Abstract**
An accurate density monitoring along a stretch of a freeway can be a useful piece of information to evaluate congestion levels, understand multiple traffic phenomena and develop efficient control strategies. While values of density near capacity are quite stable, bottleneck capacity has stochastic variations and a control strategy based on constant predetermined flow thresholds is likely to underload the freeway or, conversely, lead to traffic congestion. Following a different approach that attracted strong interest from Regional Transportation Management Center (RTMC) in Minnesota, we explore alternatives in developing the Next Generation strategy for the Twin Cities Metropolitan area freeway systems by focusing on density rather than flow. In the first part of the paper, we show empirical evidence in favor of the capacity drop phenomenon, we provide a methodology based on phase diagrams to quantitatively estimate the level of the drop, we investigate whether implementation of control strategies has an effect on the value of this capacity drop. In the second part of the paper, we develop a methodology to estimate densities with space and time based on data from loop detectors. The methodology is based on solving a flow conservation differential equation (using LWR theory) with intermediate (internal) freeway mainline boundaries, which is faster and more accurate from previous research using only external boundaries. To capture the capacity drop phenomenon into a first-order model we utilize a fundamental diagram with two values of capacity and we provide a memory-based methodology to choose the appropriate value in the numerical solution of the problem. Results compared with microsimulation of a long freeway stretch show that this model produces more reliable and accurate results than previous theories.

*corresponding author, e-mail: nikolas.geroliminis@epfl.ch

## 1. INTRODUCTION

An accurate density monitoring along a stretch of a freeway can be a useful piece of information to evaluate congestion levels, understand multiple traffic phenomena and develop efficient control strategies as ramp metering or variable speed limits. Nevertheless, this would require knowledge of a detailed density profile along a freeway section, especially at the location of the bottleneck. Most freeways have traffic state monitoring setups at specific locations along its stretch, but often do not monitor the bottleneck itself due to technical difficulties (loop detectors tend to have high errors when placed close to merge locations). This persuades the need for an effective traffic model that can efficiently predict the traffic states along a stretch of the freeway.

Lighthill and Whitham (1955) and Richards (1956) provided the first traffic flow approximation models that compared flow of traffic to compressible fluid flow. This model came to be famously known as the LWR model, and has since been extensively used as the preferred model for representing the flow dynamics from a macroscopic perspective. The LWR theory, although quite capable in providing a coarse description of main traffic features (e.g., formation and dissolution of shockwaves), is inadequate in describing some more complex traffic patterns such as stop-and-go waves, capacity drop phenomena, traffic oscillations etc. The LWR model is based on a hyperbolic partial differential equation of first order, which describes the conservation of cars in time and space. The LWR model assumes that the relation between speed (or flow) and density observed under steady state conditions holds at all times, even when flow and density vary with time and space. In other words there is always a fundamental diagram of speed (or flow) vs. density. This assumption suppresses all other traffic states and phase transitions not belonging on this curve. Thus, as described in Zhang (2003) "phase curve obtained through statistical averaging suppresses finer traffic dynamics represented by the fine structures in the scatter". According to the theory, the traffic density is predicted to be piece-wise smooth, with transitions between stable regions approximated by discontinuous shocks. This is described by the generation of shock waves between two neighbor states in a time-space plane, with characteristic speeds equal to the change in flow over the change in density between the states.

To overcome these deficiencies of LWR theory, higher order models have been developed, which usually contain an additional equation describing the spatiotemporal evolution of speed (Payne, 1971, Whitham, 1974). For a review of different models the reader can refer to Helbing (2001). The validity of this type of second order models have been questioned by many researchers (Michalopoulos, 1987, Daganzo, 1994 etc). In a seminal paper, which has created strong debates thereafter (e.g. Papageorgiou, 1998, Zhang, 2003, Helbing and Johansson, 2009), Daganzo, (1994) described many flaws of second order models. The most important is that characteristic speeds can be faster than the speed of traffic, which means that drivers are affected by phenomena occurring behind them.

The scope of this paper is two-fold. We firstly observe empirical data of macroscopic traffic phenomena at freeway merges, while later we provide an extension of LWR theory to capture capacity drop phenomena without the need to introduce higher order models. In the experimental part of this paper we initially study the capacity drop phenomenon in MN freeways for different time periods and control conditions with no change in geometry. Instead of using traditional cumulative curves of input and output flow at different locations of a freeway the analysis is based on phase diagrams, where traffic conditions at a merge are expressed in a two-dimensional plane with axes mainline and ramp flows. This type of methodology has been chosen as it provides the ability to follow the trajectory of the intermediate states between high and low capacity values, while cumulative plots can only identify the two levels of capacity.

By carefully analyzing empirical data of active bottlenecks in the Twin Cities Metropolitan Area we noticed that (i) there are many cases where capacity is underutilized, because of inefficient ramp metering control and (ii) the system once congested is unable to return to a state of flow near capacity for *too long*. One of the main reasons for the above inefficiencies is that capacity is considered constant by the control logic during all times at all bottlenecks. This is concluded based on two empirical findings: (i) a significant capacity drop after the breakdown in many locations (varying 10-20%) and (ii) the development of congestion does not only depend on the total flow (sum of mainline+on ramp), but it is a function of the ratio of the two flows, especially at times close to the breakdown occurence. More specifically, when ramp flows are higher breakdown can occur at smaller total flow. This phenomenon is observed often in MN ramps because of violations of queue ramp constraints. Another interesting observation is that capacity before and after the breakdown is quite similar (i) in 2000; ramp metering strategy did not have any ramp delay constraints, (ii) in 2001; ramp metering was out of operation and (iii) in 2008; ramp constraint was active. In all cases capacity drop phenomena were of the same magnitude not affected by the different type or ramp control strategy.

Based on these empirical findings, we suggest a segmented LWR modeling to predict density profile against space and time along a freeway. A segmented LWR utilizes all possible internal boundaries with known traffic states by breaking down the entire site into smaller sections. This reduces propagation of any erroneous estimation. To incorporate capacity drop phenomena in our formulation, we propose a memory based step-wise-linear approximation of the flow-density relation that accounts for capacity drop effects. Our results show that such a methodology provides an increased accuracy and reliability over the standard LWR model that utilizes only the external boundaries.

## 2. EMPIRICAL OBSERVATIONS

The capacity of freeway sections is most commonly defined as the maximum flow possible at the bottleneck under the current circumstances. Bottlenecks are the cause of congestion on road networks. A bottleneck is a phenomenon where the full performance level (capacity) of an entire system cannot be realized due to an abnormality at a single component of the system. The performance at one location thus brings down the performance of the entire system. An 'active' bottleneck is a bottleneck whose performance is not affected by any bottlenecks occurring downstream, and has free-flow conditions downstream. The capacity at a bottleneck can be defined as the maximum throughput possible at the bottleneck or the net flow exiting the bottleneck.

While this maximum possible flow was traditionally considered to be of fixed value for a given location, many studies have revealed that there is some stochastic nature to bottleneck capacities. Banks (1991) and Hall (1991) first suggested that discharge flow at bottlenecks diminish once queues start forming upstream of the location, thus marking the onset of congestion. The phenomenon is now best known as bottleneck 'capacity-drop'. Thus, the congested capacity of the bottleneck can be distinguished from the bottleneck's free flow capacity, with the difference being termed as the capacity drop. While some studies of the bottleneck capacity drop have suggested that the drop is non-noticeable or nonexistent (Persuad 1986), others place the drop ranging from about 3% (Hall 1991, Banks 1991) all the way to up to 12% (Cassidy 1998, Cassidy 2006). Further, studies aimed at understanding the bottleneck breakdown phenomenon suggest that the breakdown itself does not always occur at a fixed flow rate and is actually stochastic. Elefteriadou (2007) suggests that capacity can therefore only truly be defined as a function of breakdown probability. Cassidy (2006) reports that while capacity at a bottleneck might have large variations; the critical density associated with the breakdown tends to be a more stable with a smaller range of variation.

While most traditional freeway control mechanisms (including the Stratified Zone Algorithm of Minnesota freeways - Feng et al., 2006) utilize the capacity and flow measurements as the governing parameters, the higher reliability of breakdown density presents itself as the better choice as a control parameter. Further, classic capacity based control strategies do not account for the capacity drop and thus either underestimate pre congestion capacity, or overestimate post congestion capacity. Our own research on the subject in Minnesota's freeways confirms the above findings and indicates a capacity drop of as high as 20% resulting in substantial miscalculation of the optimal metering rates. This suggests that a control strategy based on flow thresholds is likely to underload the freeway or lead to traffic congestion.

We investigate an active bottleneck to understand the capacity drop phenomenon, and to estimate the extent by which capacity might fall post congestion. We further show that the capacity (when defined as the total discharge at a bottleneck) is not independent of the ratio between the mainline and on-ramp merge flow at the bottleneck. Lastly, we observe that the capacity drop witnessed at a location, is independent of any control strategies in place, and is also independent of the demands upstream and at the merging on-ramp.

The study site is a 12-mi segment of Trunk Highway 169 northbound (TH-169 NB), starting from the I-494 interchange and ending at 63rd Avenue North (Figure 1). This site is a circumferential freeway traversing the Twin Cities west metropolitan region. It includes 10 weaving sections, 4 HOV bypass ramps, 24 entrance ramps (17 metered), and 25 exit ramps. Among the metered ramps, 15 local access ramps and two freeway-to-freeway ramps connect TH-62 and I-394, respectively. The upstream and downstream boundaries are uncongested.

***Fig. 1:*** *The selected test site (TH-169 NB).*

### 2.1 Data Analysis

We choose an active bottleneck along the site of US Highway 169 Northbound at Plymouth Avenue on-ramp (closely downstream of the highway to highway connection with TH55) for our analysis. Recurring congestion is evident during the evening hours (approximtately 16:00-18:00) while the station downstream of the bottleneck does not register congestion levels with speeds close to free-flow, thus confirming that the chosen site is an active bottleneck site. Various traffic state data was collected for this study site for various years: 2000 (with the previous incarnation of Minnesota’s ramp metering strategy: Zone Metering under implementation), 2001 (with no metering strategy active), and 2008 (with the latest implementation of metering: SZM in place). The traditional way to observe capacity drops (e.g. Chung et al., 2006) is by using time series of output or cumulative output at the bottleneck (approximated most commonly as the sum of the flow just upstream of the bottleneck and flow at the ramp involved in the bottleneck or as the flow just downstream of the bottleneck if the data is available). The capacity drop, could then be identified either as the fluctuation in the value of flow pre- and post-congestion in case of measuring output, or as a change in slope of the cumulative output curve. Another way is using a phase diagram to plot a relation between the flow and the density (or occupancy) at the bottleneck site. Such a plot is more useful than cumulative curves as the trajectory of ramp/mainline flow can be observed. We present a representative throughput time series plot (sum of volumes at the upstream mainline detector station and volumes at the on ramp merge) in Figure 2, which shows a capacity drop of roughly 16.5% (from ~4200 pre congestion between around 16:15 to ~3500 during congestion as seen after 16:30). Speed profiles have significantly higher values before the occurrence of the breakdown, at 16:20. The figure also shows the ramp demand separately in the same plot so as to provide an estimate of the demand at the bottleneck. Note that ramp flow significantly increases a few minutes before the breakdown (black arrow in fig. 2), while similar total demand at 16:05 did not create a breakdown because on-ramp flow was lower. The Flow versus Density plot for the bottleneck (using flow as the total output flow at the bottleneck as defined earlier, and density at the upstream mainline detector), shown in Fig. also shows the 16.5% capacity drop (from a peak at ~4200 during the uncongested regime to ~ 3500 during congestion). We further show in the following portion of the paper that these capacity drop values remain constant across a vast time horizon (2000 – 2008) and under varying ramp control mechanisms.

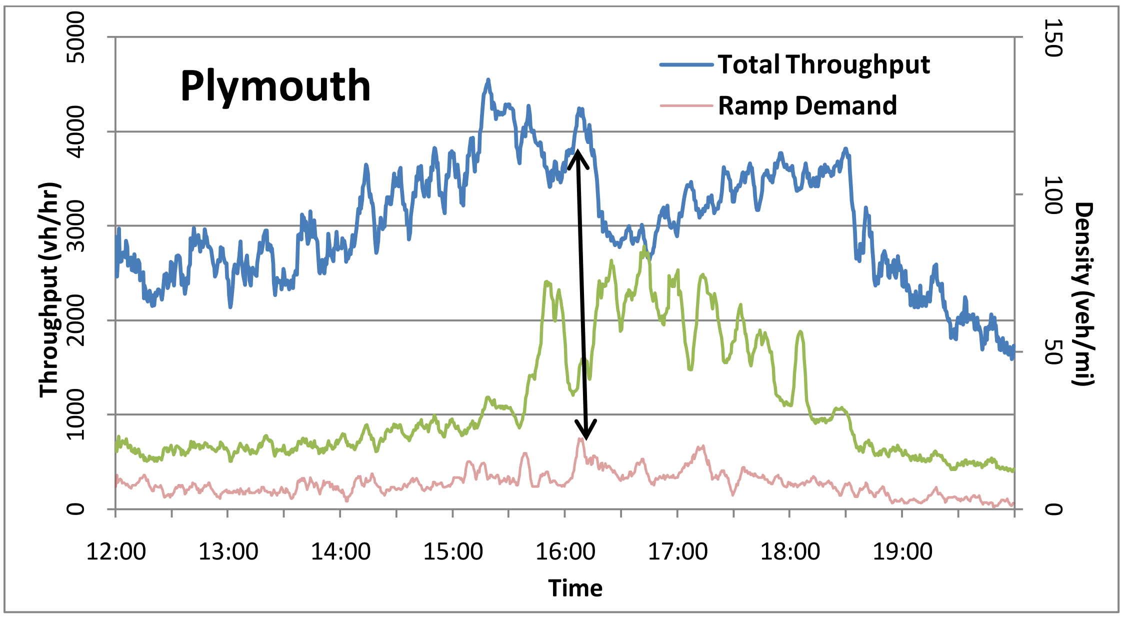



***Fig. 2:*** *Time Series Plot of Throughput at Bottleneck (5 min averages)*

*The graph shows the 5-minute-time-average time series plots of total throughput at the bottleneck, along with the demand at the corresponding on-ramp, and the mainline density (along secondary axis). The total throughput can be seen here to dip from a maximum of ~4200 v/hr between 15:15-16:15 hrs, to ~3500 v/hr between 16:30-18:15 hrs. The corresponding mainline density plot clearly shows the congested period, while the ramp demand plot shows that the drop in throughput cannot be attributed to low demand.*

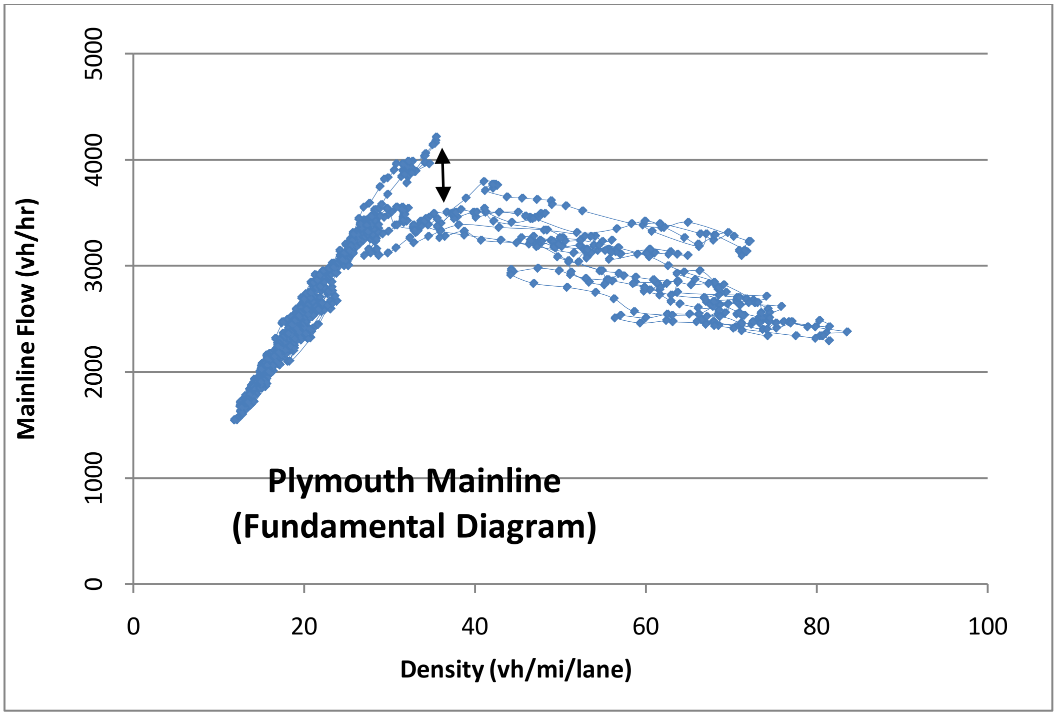



***Fig. 3:*** *The fundamental diagram depicting the flow-density relationship at the bottleneck. The fundamental diagram for the US169 NB - Plymouth bottleneck location (based on 5 minute time average values) shows a similar 'gap' in capacity from ~4200 v/hr during free flow conditions to ~3500 v/hr post congestion.*

In order to better understand the capacity drop and to test how it changes with time (varying demand / varying control strategy implementations), we make use of a bottleneck flow contribution plot, a phase diagram. By plotting the flow volumes at the ramp associated with the bottleneck against the flow volumes observed at the upstream mainline detector one can better understand the breakdown behavior of the bottleneck, and how capacity changes during the duration of the breakdown (onset of congestion). In order to incorporate the time element into the graphs, we distinguish between the various phases: pre-congestion free flow regime, onset of congestion when speeds continue to decrease, and post breakdown congested regime, each represented by a different shade of the plot. Thus we can follow the behavior as the breakdown happens. Further, we support this plot with a time series plot of the demands on the ramp (using traffic states both at the merge detector and at the queue detector whenever available) which would help identify demand fluctuations on the ramp. Figure 4 show the bottleneck flow contribution plots for various days in 2000 (simple Zone Metering), 2001 (No metering), and 2008 (Stratified Zone Metering). The left hand side plots of figure 4 show the phase diagrams between on-ramp and mainline flow for different days. The right hand side plot show the time series of metering rates (flow) at the downstream on-ramp detector and the occupancy measures at the upstream queue on-ramp detectors. For earlier years (2000,2001) occupancy is for downstream ramp detectors, as queue detectors were not instrumented at that time.

An isoquant in the phase diagrams, can be defined as the line connecting all flow contribution points on the plot that sum up to a certain total (ramp volume + mainline volume = constant). Since the scale along the 'ramp volume' axis is much smaller than the corresponding 'mainline volumes' axis, the isoquants are along lines with a very steep negative slope. The horizontal separation between portions of the curve would thus represent the change in capacity. Certain spikes in the demands on the ramp are highlighted, both in the flow contribution plots as well as the demand time series graphs for better understanding.

Fig. 4a represents roughly a 60 minute span of data between 1500 – 1600 hours (30 minutes of free flow conditions, 10 minutes of onset of congestion, and 20 minutes of congestion). The figure shows the mainline volume versus ramp volume plot for September 17$^{th}$ 2008, along with the time series plot for ramp volumes (the ramp demand at Queue detector location, and the ramp supply at the Merge detector location). The flow of time for the graph is from the blue segment representing the period before onset of congestion, the green segment that depicts the time when the congestion starts building and then to the red segment when the location is under consistently congested phase. We can see from the 2 plots that the queue detector starts registering high demands starting approximately at 15:08. This is followed by high discharge rate at the ramp, thus, initially increasing the throughput at the bottleneck (along the blue segment) to 36 vehicles per 30 seconds. This throughput is sustained for approximately 5 minutes before the consistently high ramp discharge rate causes a breakdown at 15:15 which corresponds to the marked peak in the flow distribution graph. Once the breakdown happens, the high volumes are no longer sustainable and the capacity falls by ~15%. The plot in the flow contribution graph shows the fall of throughput along the blue segment and leading into the green and red phases. It is clearly seen here, that once the breakdown happens, the throughput stagnates at a capacity of about 32 vehicles per 30 seconds, a 15% decrease from the initial peak.

Fig. 4b narrates a similar story on September 10$^{th}$ 2008. High demands are first registered between 15:08 and 15:20. The flow contribution plot here, follows the blue segment up to a peak (due to increased discharge due to high demand on ramps), and then down to where it merges with the red segment (due to restrictive metering following the high volumes at the mainline). The restrictive metering in this case (supported by low demands post 15:20) however, is able to avoid the onset of congestion. The throughput thus once again increases to its peak levels (merge or the green and red segments) when the demand rises at 15:35. The sustained high demand however causes a breakdown this time and the throughput falls to the congested capacity along the green segment where it stabilizes this time.

The remaining days all show similar behavior. The breakdown is always triggered by an increase in ramp discharge volumes, causing a fall in capacity by 15%. The uncongested capacity throughput is almost always sustained for duration of at least 5 minutes. Fig. 4e illustrates that after breakdown, the throughput stabilizes close to the post-congestion capacity with very slight variations which are distinguishably smaller than the drop in capacity during breakdown. We claim that this provides sufficient proof that the pre-breakdown and the post-breakdown capacities are sustainable and that the capacity drop is consistent across time and variations in control strategies or demands.

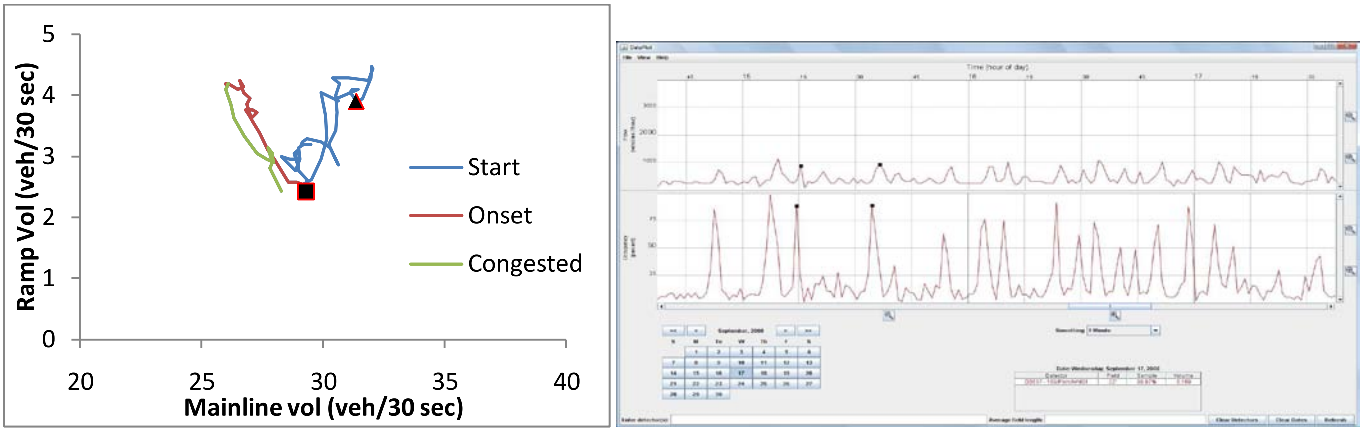



Fig. 4a: September 17th, 2008

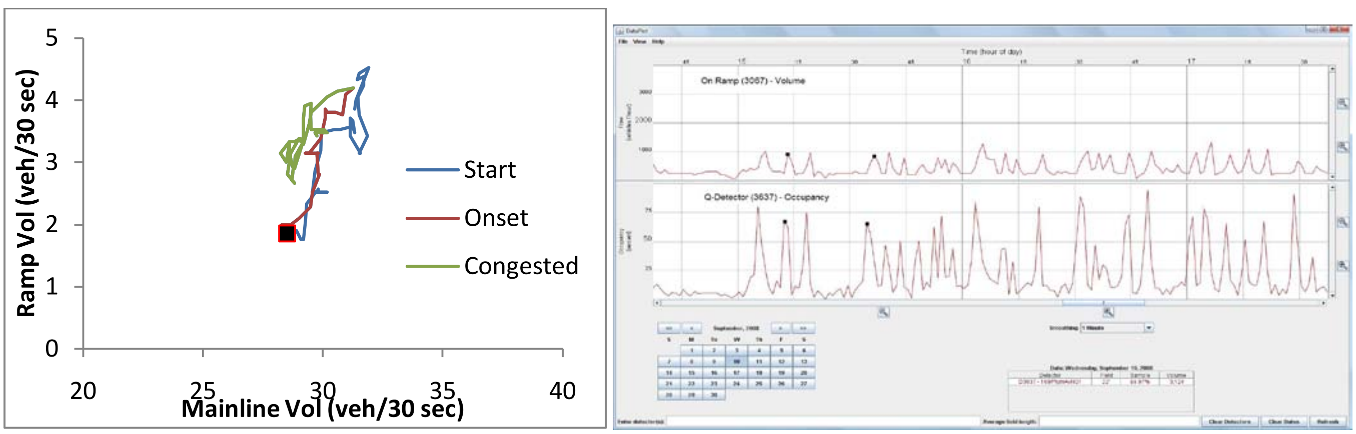



Fig. 4b: September 10th, 2008

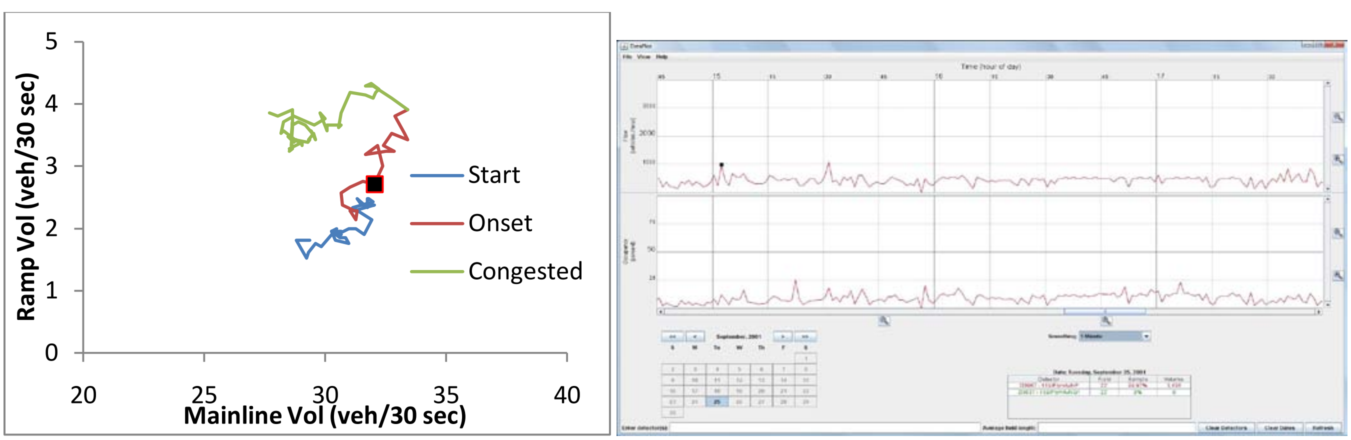



Fig. 4c: September 25th, 2001

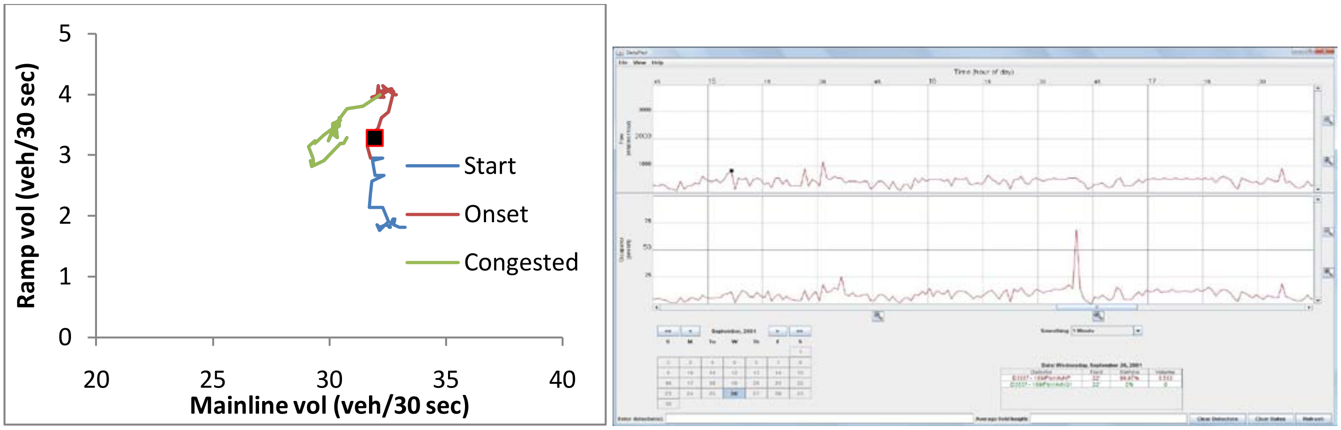



Fig. 4d: September 26th, 2001

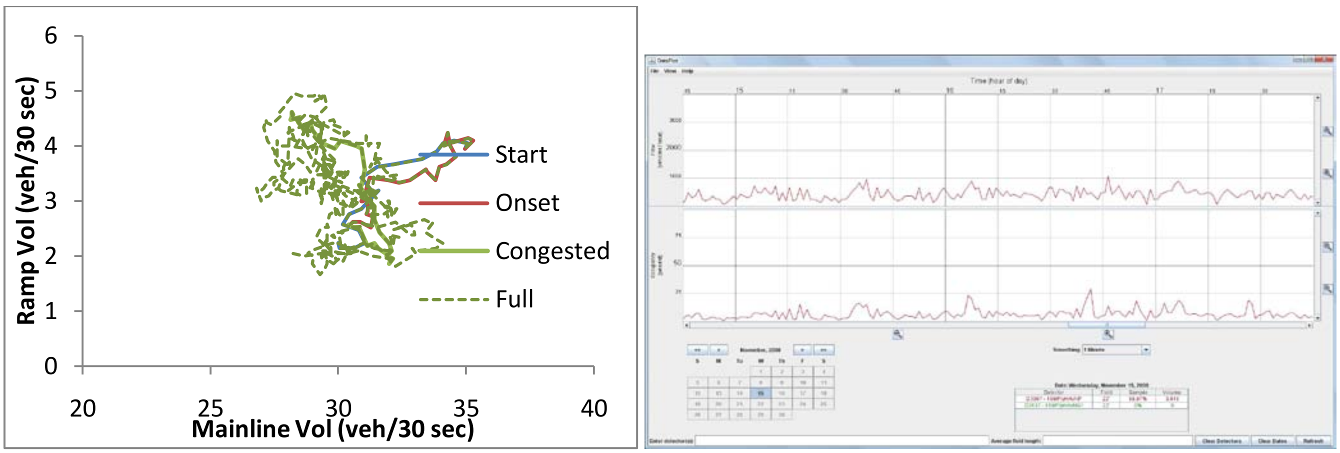



Fig. 4e: November 15th, 2000

*Fig. 4: Bottleneck Flow Contribution Plots showing Mainline Volumes Vs Ramp Volumes.*

*The plots 4a through 4e show the flow contribution plots for the bottleneck for various candidate dates in 2000, 2001 and 2008. The mainline volume is plotted against the ramp volumes at the site of the bottleneck. Also shown are the time series plots of ramp volumes: Q detector demand volumes, and Merge detector supply volumes (The queue detectors were not in place in 2000 and thus the densities at merge detector location are instead shown for the corresponding cases). The plots show a consistent pattern across the dates and suggest a consistent capacity drop of ~15% for the location.*

It can be shown from 4, that the pre-congestion (~72 vehicles/minute) and post-congestion (~62 vehicles/minute) capacities at the bottleneck, and hence also the drop in capacity (~15%), are roughly the same for all three years considered. This leads to the argument that the capacity drop is independent of the demand seen at the bottleneck, and of the control strategy in place. This implies that changing the ramp control metering can be expected to not affect the capacity drop noticed at the bottleneck. There is also an interesting observation that is derived directly from some of the flow contribution plots by observing the inclination of the curves to the two axes.

The inclination (during the congested phase) from Fig. 4a suggests that the volumes stabilize along an inclination that represents a ratio roughly equal to 2:1 between the mainline and ramp. This would imply that the isoquants are not equally distributed between the mainline and the ramp, and that an additional vehicle on the ramp is twice as detrimental to the congestion level at the bottleneck, as two additional vehicles on the mainline. However, this is not an easy attribute to be observed (since such an observation can only be made if the bottleneck remains close to full capacity, and thus in the same isoquant, for an extended duration) more bottlenecks need to be explored and with a vaster time horizon in order to be able to support such a proposition. The various plots however, give consistent and convincing proof towards the existence of capacity drops and their independence from demand pattern variations.

## 3. DENSITY MODELING

After describing the behavior of capacity drop, we next attempt to model the density distribution along a stretch of a highway section, using traffic data collected at the available detectors along the mainline and along the ramps. Current metering strategies based on density information use either a linear interpolation of density between known detector locations, or simple assuming the maximum / minimum at either ends of a section, for estimating the congestion level within the extent of a section. The actual bottlenecks are however usually likely to form closer to where the ramp merges into the freeway and thus often considerably away from the location of the detectors. The detectors are often not placed close to the actual merge intentionally since such locations witness a lot of lane changing movements and detectors are often not capable of accurately catching the right counts / densities under such situations. Thus, the densities that are observed at the two ends (upstream and downstream) of a section of freeway are often lower than the peak density witnessed along the section. Using simulation data obtained for densities along a freeway section, we first show that linear interpolation approximations, from known detector locations, can have high errors (both positive and negative). Once a need for a better profiling of density along the section has been established, we propose a simple model that can be efficiently used to predict densities along the stretch and then follow it with some comparison studies done against simulation data.

For the purpose of the study, we use the stretch of US Highway 169 Northbound, between its intersection with Valley View Road (at the upstream end) and with County Road 10 (at the downstream end). We specifically look at the congestion section along the highway between its intersection with Excelsior Boulevard at the upstream end and with Plymouth Avenue at the downstream end for some of the analysis. The Plymouth Avenue on ramp to US 169 was a site of an active bottleneck consistently.

We model the traffic behavior along the highway stretch using AIMSUN (microscopic traffic simulation software). The simulator uses demand values at all access points to the network, as well as using turning movement percentages at any decision point (such as an off ramp) as input. Demand data have been obtained from real measurements from detector data. Additional detectors were placed along the stretch of freeway being studied in order to catch the traffic states between actual detector locations. The density profile obtained through the simulation is compared against a simple linear interpolation from available mainline detector readings, is shown in figure 5. Fig. 5a and 5b show density values (in veh/mile) vs. distance (in feet) for two different times. The figure clearly shows the error levels in a simplistic linear estimation model, thus demanding a better estimation model in order to more effectively predict realistic congestion levels within a section.

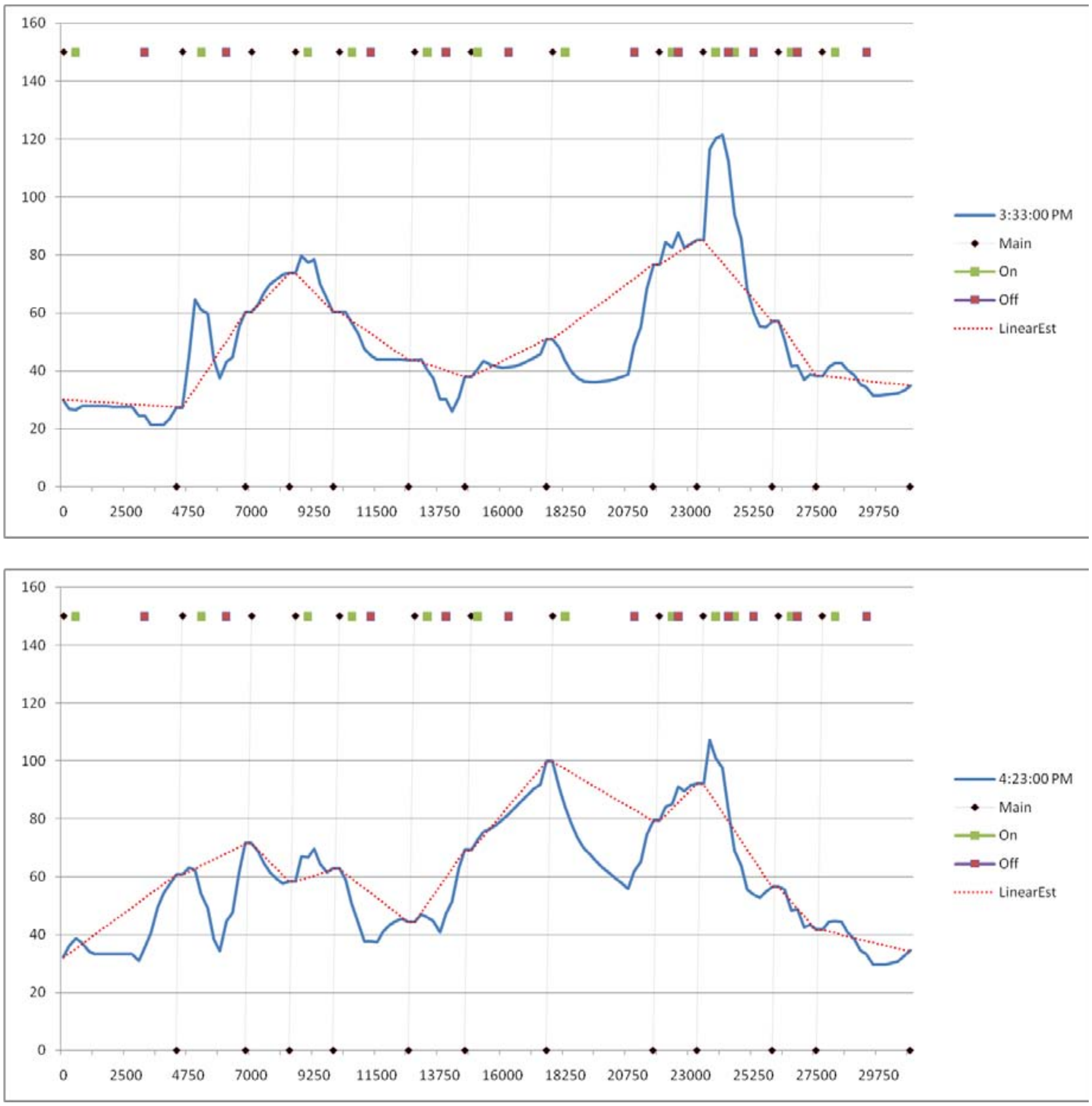



*Fig. 5: Simulation Obtained Density Vs Linear Approximation Model.*

*The figure shows two plots for random candidate time positions, showing the density profile as reported by AIMSUN Simulation along US169-NB. The vertical lines mark the location of mainline detector stations located along the stretch, while the red and green markers on top represent the location of off-*

*ramps and on-ramps along the stretch. The linear interpolation density profile constructed using the real detector location values is shown to contrast the non-linear nature of density profile along any section of freeway. The plot illustrates that the mid-section densities can be considerably higher (or lower) than the linear interpolation estimates.*

### 3.1 Extended first-order modeling

The Lighthill Whitham Richards (LWR) theory is a well established continuum theory for traffic flow. The LWR model is based on hydrodynamics analogy treating traffic flow as similar to fluid flow. The theory's strength lies in its simplistic representation of traffic flow as a continuous fluid, inherently treating traffic in an equilibrium state. The first order continuum model employs only the flow conservation equation (eq. 1) and a known relation between flow and density, depicted through the representation of stream speed as a function of density (eq. 2). Traditionally a known state equation (such as the Greenshields model) is used to determine this stream speed – density dependence. Michalopoulos et al. (1984, 1987) provide a numerical solution for the first order approximation model assuming a known flow-density relation. Though the first order models have been known to have certain drawbacks: inability to account for stop-and-go behavior resulting in unstable traffic under congestion, inability to incorporate the capacity drop phenomenon, abrupt transitions between states thus suggesting infinite acceleration etc.

We use a first order numerical solution to the LWR model to predict the temporal and spatial distribution of the traffic state (flow characteristics) along a section of the freeway (with a defined geometry) similar to Michalopoulos et al. (1984). Traffic state at the boundaries (time and space boundaries) along with a flow-density relationship for the section is provided to the model as input. The first difference in our approach is that instead of modeling the whole section of a freeway, we utilize the information of intermediate mainline loop detectors. We incorporate this data by segmenting the freeway into segments between actual loop detectors and applying detector measurements as internal boundaries in the formulation. Thus, the numerical solution is much faster and as we will show, the obtained solutions much more accurate.

The entire stretch of the freeway network is first divided into sections (segments) bounded by presence of detector locations where the traffic states are measurable. The Segment LWR model is applied to each such segment separately. Each segment is broken down into distance steps of length $\Delta \boldsymbol{x}$ and similarly time is discretized into $\Delta \boldsymbol{t}$ steps. The input data consists of traffic state information at the two segment boundaries ($k$, $Q$), and generation rates at any source of entry or exit within the segment $g(q)$. Ideally the relation between $\Delta \boldsymbol{x}$ and $\Delta \boldsymbol{t}$ should be such that $\Delta \boldsymbol{x}/\Delta \boldsymbol{t} > \boldsymbol{u_f}$ *(free-flow speed).*

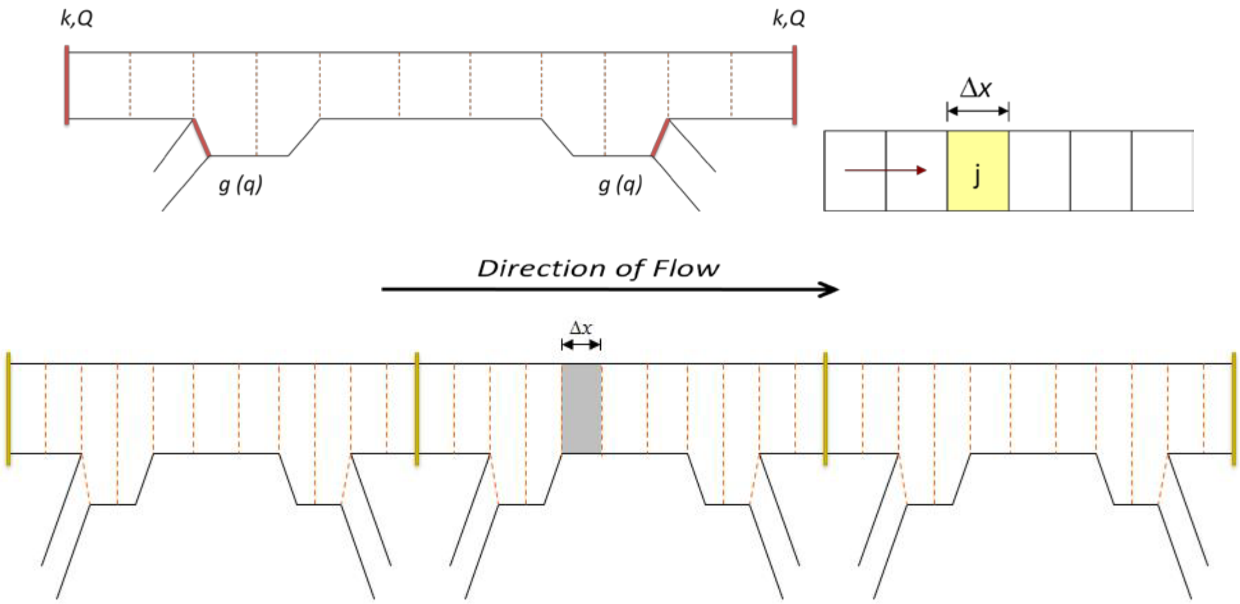



***Fig. 6:*** *Application of LWR Model to a section of freeway.*

Equations (1) and (2) describe the traditional LWR theory formulation, while equations (3) and (4) provide the numerical approximation. Equation (1) is a mass conservation equation between flow $q$ and density $k$ in time and space, expressed as a hyperbolic partial differential equation, while (2) expresses the fundamental diagram of speed $u$ vs. density $k$ for steady state conditions. The term $g(x,t)$ is a generation or termination flow, e.g. at on-ramps, off-ramps and at the downstream and upstream mainline boundaries. The set of equations (3) and (4) provide the first order numerical calculations for obtaining the various traffic states (density and flow) at any time instant at any given location along the section. A profile of the density distribution is thus created along the time and space dimensions.

$$\frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = g(x,t) \qquad (1)$$

$$u = u_e(k) \qquad (2)$$

$$k_j(n+1) = \frac{1}{2}\left[k_{j+1}(n) + k_{j-1}(n)\right] - \frac{\Delta t}{2\Delta x}\left[q_{j+1}(n) - q_{j-1}(n)\right] + \frac{\Delta t}{2\Delta x}\left[g_{j+1}(n) + g_{j-1}(n)\right] \qquad (3)$$

$$q_j(n+1) = k_j(n+1).\,u_j(n+1) = k_j(n+1).\,u_e\left[k_j(n+1)\right] \qquad (4)$$

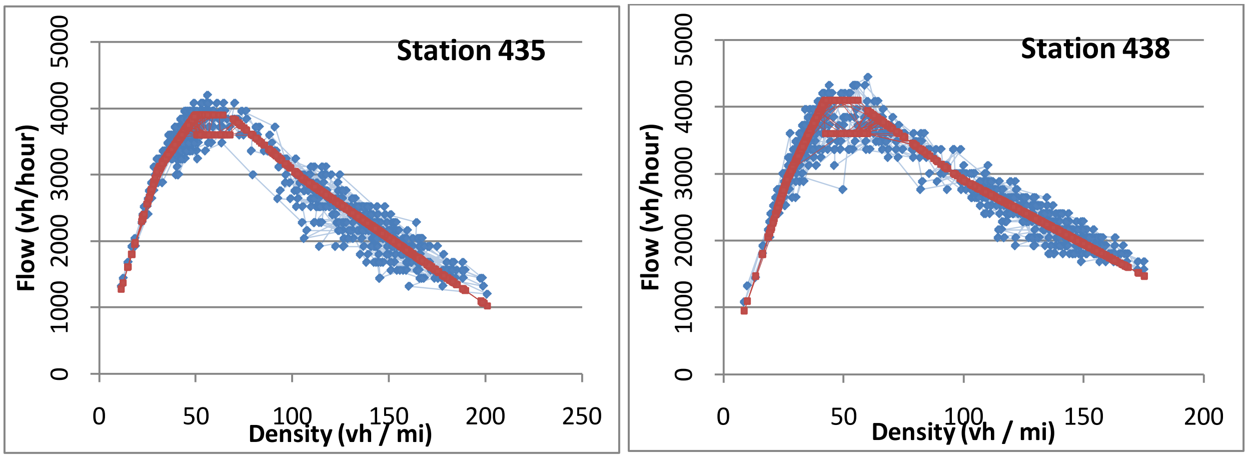
To capture the capacity drop phenomenon into a first-order model we utilize a fundamental diagram with two values of capacity and we provide a methodology to choose the appropriate one in the numerical solution of the problem.We introduce a state parameter used to add a memory-based decision. It utilizes the previous 3 minutes of density data collected at the location to predict whether the location is currently under free flow or congested regime. This aspect is then used in order to incorporate the effect of capacity drop. Thus, for the same density value, a higher flow is estimated (corresponding to free flow capacity) if the densities observed in the previous 3 minutes has been lower than the critical density for the location, while a lower flow is estimated (corresponding to congested capacity) if the location has been in congestion.

The given section is divided into sub-parts each of a length $\Delta$x which form the location/distance steps, and the time horizon is similarly divided into time steps each $\Delta$t seconds of length. Ideally the relation $\Delta$x/$\Delta$t should be close to the free flow speed on the section being studied. The inputs to the model then consist of: (1) the flow and the density information at the boundaries of the freeway section (at the upstream boundary and the downstream boundary) at all time steps, (2) the flow and density at each location at time zero along the length of the section, (3) flow at all sources (on-ramps), and sinks (off-ramps) along the section, and (4) the flow-density relation for traffic in the given section. Fig. 6 shows a representation of how the model is constructed.

Previous studies have been able to utilize such a first order estimation model to predict the density / flow profile along a section of freeway. These studies have used the entire freeway stretch as a single unit and applied the model on the entire stretch thus using the traffic state data only at the upstream and the downstream extremes as input. A segmented LWR model, that treats each section (stretch of freeway between two consecutive available detector stations) as a separate unit, thus computing the distribution of density separately for each section, can more effectively utilize all the available detector information. Such a method is not only bound to have a higher accuracy (since errors can no longer propagate along space), but is also computationally more efficient (since each model is applied for a single section, and hence the computational load can potentially be distributed). Furthermore, the LWR model needs an intrinsic relationship definition between the flow and the density at all locations along the stretch being modeled. We propose here, that using a simple stepwise linear estimation of the flow-density relation, while accounting for the capacity drop phenomenon, can greatly help increase the accuracy of the modeling while keeping the computational effort light. Figure 7 shows the MFD (flow density relationship) estimation used for this purpose for a few sample locations along the freeway.

The fundamental flow – density relation for any section is approximated as a 5-step piecewise linear model. Each step is represented by a conditional block when defining the flow-density relation in the LWR model. The model thus computes the correct flow value corresponding to a given density, by sequentially going through all the conditional blocks that are used to define the relation. Each block $i$ is defined by a set of 4 values: density boundary between block $i$ and $i$+1, $k_i$; congestion state, $s_i$ (with value 0 for uncongested and 1 for congested); flow axis intercept, $c_i$ and slope, $m_i$ (together define the shape of the linear relation pertaining to the current step).

The first two blocks blocks typically represent the uncongested phase of traffic. Block 1 is for light conditions where vehicles run at free-flow speed. Block 2 is undersaturated, but the effect of vehicle interactions slightly decreases speed below free-flow. The next two blocks represent the pre-congestion and post-congestion capacities, while block 5 represents the behavior in the congested phase at the location. The proposed Segmented LWR model thus incorporates both the segmented nature, and the memory based flow-density approximation model described.




| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $k_i\ (vh/mi)$ | 30 | 50 | 68 | 68 | |
| $s_i$ | 0 | 0 | 0 | 1 | 1 |
| $m_i\ (km/hr)$ | 95 | 42 | 0 | 0 | -25 |
| $c_i\ (vh/hr)$ | 200 | 1800 | 3900 | 3600 | 5580 |

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $k_i$ | 27 | 42 | 58 | 60 | |
| $s_i$ | 0 | 0 | 0 | 1 | 1 |
| $m_i$ | 110 | 75 | 0 | 0 | -26 |
| $c_i$ | 0 | 930 | 4100 | 3600 | 5500 |

***Fig. 7:*** *"5-step" Stepwise Linear Flow-Density Relation.*

*The figure shows the stepwise linear approximation functions used for the flow-density relationship for 2 locations along the study site. The curve is typically broken down into 5 steps: 2 for the free flow regime of the relation, 1 each for the pre-congestion and post-congestion capacities, and 1 for the congested regime of the relation. The memory retaining nature of the estimation allows for use of 2 different values of capacity flow based on whether the conditions have been congested during the previous 3 minutes at the location or not. The tables accompanying the figures show the values of the various parameters used to construct the flow-density relation that serves as input to the LWR model.*

### 3.2 Simulation results

For the purpose of the study, we use the stretch of US Highway 169 Northbound, between its intersection with Valley View Road (at the upstream end) and with County Road 10 (at the downstream end). The 15-min-interval traffic demand data used in the simulation were extracted from the Minnesota DOT loop detector database. For the purpose of this study, only one peak period was tested, specifically the p.m. peak on September 25th, 2008 between 2pm and 8pm.

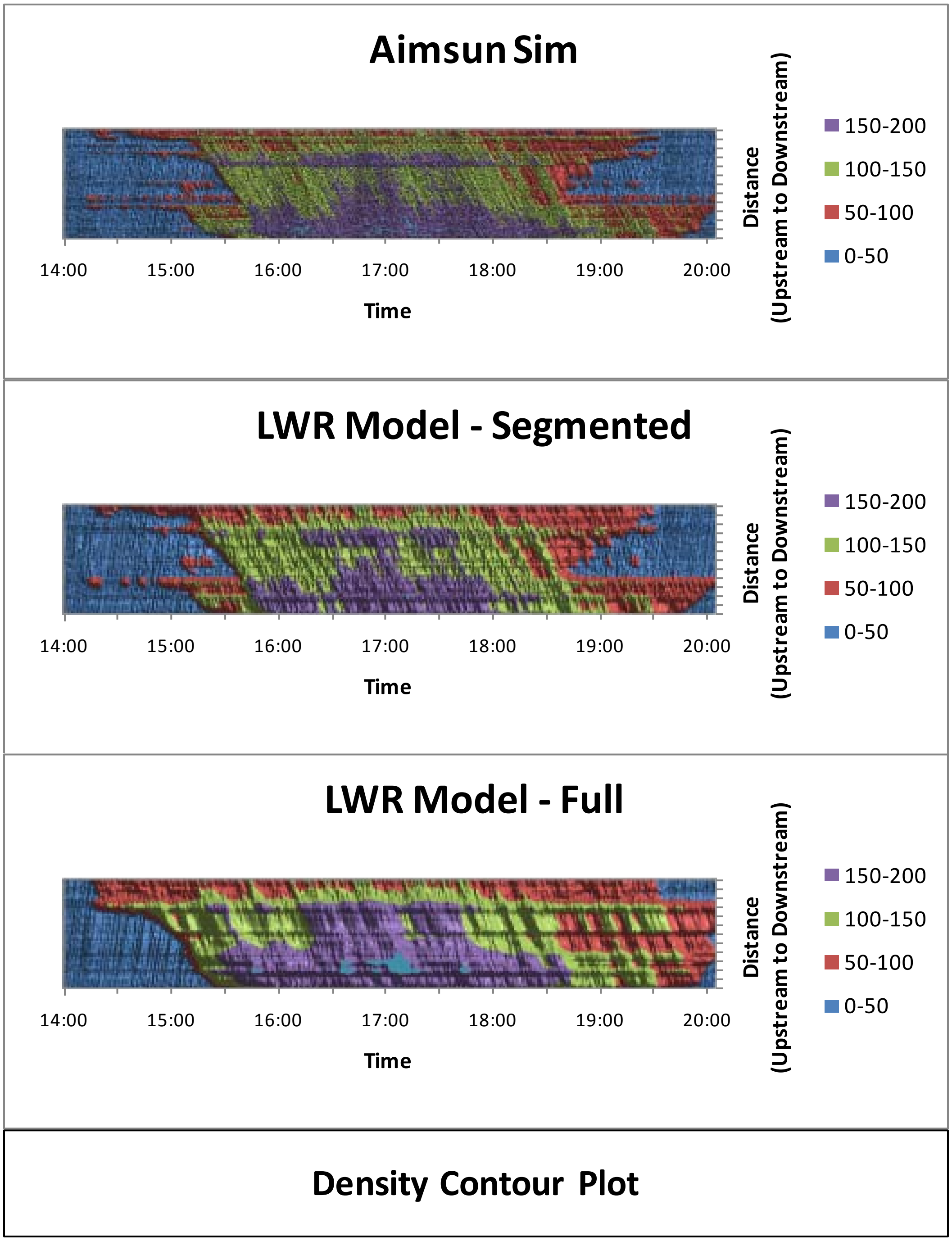



***Fig. 8:*** *Density Contour Plots comparing AIMSUN vs Traditional LWR vs Segmented LWR.*

*The figure compares the Density contour plots (against space and time) obtained from AIMSUN simulation, against those obtained from a traditional full length LWR model and the proposed Segmented LWR model with capacity drop phenomena. The Segmented LWR model clearly outperforms the full length LWR model as is illustrated above. The full length model shows error predicting the onset of congestion early on for some of the locations, as well as considerably delayed offset predictions. The segmented LWR model's predictions remain considerably truer to the simulation reported values of density throughout the study duration.*

Fig. 8 shows a contour plot depicting density profile against time (horizontal axis) and space (vertical axis) for the three cases being compared: Aimsun simulation data, the Segmented LWR Model as proposed in the paper, and the simple Full Span LWR Model estimation results. The Segmented LWR model shows considerable improvement over the Full Span LWR Model in estimating the density profile as verified against the simulation generated profile. In order to further substantiate the estimation model's strength, Binary Contour Plots are created for the three scenarios using different threshold density values. The Binary Contour Plot depicts the contour using a specific value of density (50 vehs/mile) which closely approximates the critical density along the stretch, and 30 vehs/mile as shown in Figure 9. Such a plot gives a clearer picture of how well a model can estimate the congestion boundaries (onset and offset) both in space and in time.

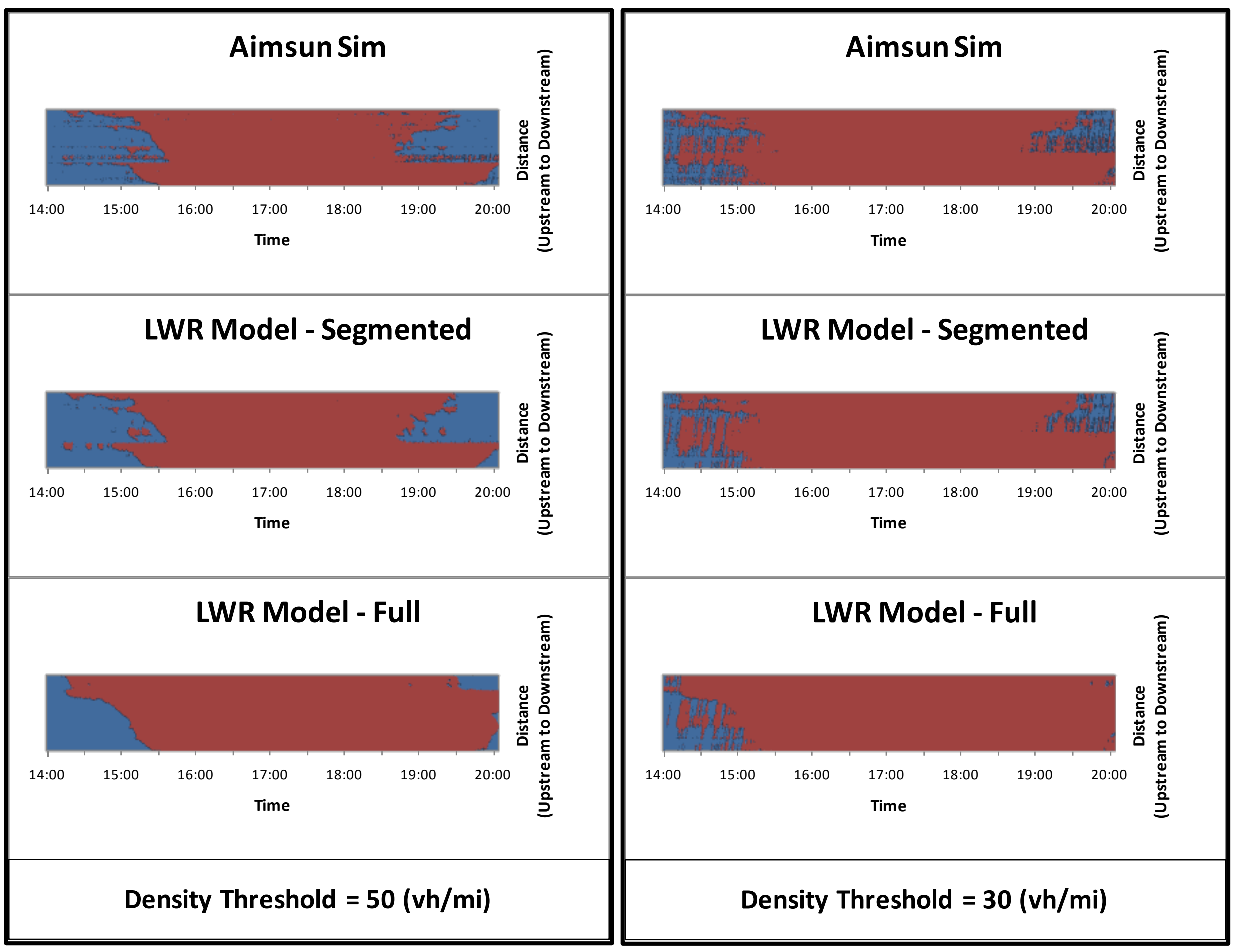



*Fig. 9: Binary Density Contour plots for AIMSUN vs. Segmented LWR vs. Full Length LWR Model.*

*The figure shows two binary plots of density that depict the density contour for two specific thresholds. These plots are an extension of the previous figure that showed the entire density contour plot. The binary plots are useful in representing contours for specific values of density with a higher level of clarity. For both the density thresholds used, one can see that the Segmented LWR performs much better than the Full*

*Length LWR as is seen especially during the offset of congestion along the site where the full length model over-estimates the congestion duration. Similarly, the segmented LWR shows improvements over the full length model during the onset hours as well, more closely modeling the actual onset conditions.*

## 4. CONCLUSIONS

In the first part of the paper, we show empirical evidence for the capacity drop phenomenon using data from freeways in the Twin Cities Metropolitan Area. We provide a methodology based on phase diagrams to quantitatively estimate the level of the drop and we investigate whether implementation of control strategies has an effect on the value of this capacity drop. An interesting observation is that capacity before and after the breakdown is quite similar (i) in 2000; ramp metering strategy did not have any ramp delay constraints, (ii) in 2001; ramp metering was out of operation and (iii) in 2008; ramp constraint was active. In all cases capacity drop phenomena were of the same magnitude not affected by the different type or ramp control strategy. In the second part of the paper, we develop a methodology to estimate densities with space and time based on data from loop detectors. The methodology is based on solving a flow conservation differential equation (using LWR theory) with intermediate (internal) freeway mainline boundaries, which is faster and more accurate than previous research using only external boundaries. To capture the capacity drop phenomenon into a first-order model we utilize a fundamental diagram with two values of capacity and we provide a memory-based methodology to choose the appropriate one in the numerical solution of the problem. Results compared with microsimulation of the Highway H-169 stretch show that this model produces more reliable and accurate results than previous theories.

## ACKNOWLEDGMENTS

Support for the study described here was provided by the ITS Institute of the Center for Transportation Studies at the University of Minnesota. The authors acknowledge Brian Kary and Doug Lau of MN/DoT’s Regional Traffic Management Center and Dr. John Hourdos of Un. of Minnesota, for their valuable comments, support, and continuous cooperation.

## REFERENCES

Banks, J. H. (1991). Two-capacity phenomenon at freeway bottleneck. Transportation Research Record, 1320, 83-90.

Cassidy, M. J., Bertini, R. L. (1999). Some traffic features at freeway bottlenecks. Transportation Research Part B, 33, 25-42.

Chung, K., Rudjanakanoknad, J., Cassidy, M. J., (2007). Relation between traffic density and capacity drop at three freeway bottlenecks. Transportation Research Part B, 41, 82-95.

Daganzo, C. F. (1995). Requiem for second-order fluid approximations of traffic flow. Transportation Research Part B, 29, 277-286.

Feng, B., Hourdos, J., and Michalopoulos, P., (2006). Improving Minnesota's Stratified Ramp Control Strategy *Trans. Research Record*, 1959, 77-83

Hall, F. L., Agyemang-Duah, K. (1991). Freeway capacity drop and definition of capacity. Transportation Research Record, 1320, 91-98.

Helbing, D., (2001). Traffic and related self-driven many-particle systems,Rev. Mod. Phys.,73 (4), 1067-1141

Helbing, D. and Johansson, A.F., (2009). On the controversy around Daganzo's requiem for and Aw-Rascle's resurrection of second-order traffic flow models, Eur. Phys. J. B 69, 549–562

Lighthill, M. J., Whitham, G. B. (1955). On kinematic waves II: a theory of traffic flow on long crowded roads. Proceedings of the Royal Society, A229, 317-345.

Lorenz, M. R., Elefteriadou, L., (2007). Defining freeway capacity as function of breakdown probability. Transportation Research Record, 1776, 43-51.

Michalopoulos, P. G., Beskos, D. E., Lin, J. (1984). Analysis of interrupted traffic flow by finite difference methods. Transportation Research Part B, 18, 409-421.

Michalopoulos, P. G., Lin, J., Beskos, D. E. (1987). Integrated modelling and numerical treatment of traffic flow. Appl. Math. Modelling, 11, 447-457.

Papageorgiou, M., (1998). Some remarks on macroscopic flow modeling. Transportation Research A 32 (5), 323–329.

Payne, H.J., (1971). Models of freeway traffic and control. In: Bekey, G.A. (Ed.), Mathematical Models of Public Systems. Simulation Councils Proc. Ser., vol. 1, pp. 51–60.

Persuad, B. N. (1987). Study of a freeway bottleneck to explore some unresolved traffic flow issues. PhD Dissertation, University of Toronto, Canada.

Richards, P. I. (1956). Shockwaves on the highway. Operations Research, 4, 42-51.

Whitham, G.B., (1974). Linear and Nonlinear Waves. Wiley, New York.

Zhang H.M., (2003) On the consistency of a class of traffic flow models, Transportation Research Part B 37, 101–105.