



Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model



Anupam Srivastava^a, Nikolas Geroliminis^{b,*}

^a Department of Civil and Environmental Engineering, University of California, Irvine, USA

^b School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ARTICLE INFO

Article history:

Received 12 April 2011

Received in revised form 11 February 2013

Accepted 13 February 2013

Keywords:

Capacity drop
First-order models
Freeway merges
Ramp metering

ABSTRACT

An accurate density monitoring along a stretch of a freeway, especially under congested time-variant conditions is necessary to evaluate congestion levels, understand complex traffic phenomena and develop efficient control strategies. In the first part of the paper (i) we show empirical evidence from freeway-ramp merges in Twin Cities freeway system, in favor of the capacity drop phenomenon, (ii) we provide a methodology based on phase diagrams to quantitatively estimate the level of the drop, (iii) we show that the level of the drop depends on the ratio of mainline vs. ramp flow and (iv) we investigate whether implementation of control strategies has an effect on the value of 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, by integrating the capacity drop. 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 approaches using only external boundaries. To capture the capacity drop phenomenon into the 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 with a Godunov scheme. Results compared with real data and micro-simulation of a long freeway stretch show that this model produces more reliable and accurate results than previous theories.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

An accurate density monitoring and modeling along a stretch of a freeway is important while evaluating congestion levels, understanding multiple traffic phenomena and developing efficient control strategies through ramp metering or using 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 US 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 often tend to have high errors when placed close to merge locations, e.g. due to double counting of vehicles in regions with high lane changing behaviors). The placement of the detectors (upstream, at the merge or downstream) around the merge areas is also determined by the implemented ramp metering algorithm. Thus, there is a 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 (LWR). The LWR model has since been extensively used as the preferred model

* Corresponding author. Tel.: +41 21 6932481.

E-mail addresses: srivast1@uci.edu (A. Srivastava), nikolas.geroliminis@epfl.ch (N. Geroliminis).

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 and traffic oscillations. 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 has been questioned by many researchers (Michalopoulos et al., 1987; Daganzo, 1995). In a seminal paper, which has created strong debates thereafter (e.g. Papageorgiou, 1998; Zhang, 2003; Helbing and Johansson, 2009), Daganzo (1995) 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 twofold. 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 Minnesota 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. In the modeling part of this paper we integrate the capacity drop (a property with memory), in a first order model, which is traditionally memory-less (history does not affect the shape of the fundamental diagram). 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 also show that utilizing a single value of capacity (before or after the drop or the average of the two) creates significant errors in the model.

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. These limitations of the metering control are identified through two important empirical findings: (i) a significant capacity drop after the breakdown is observed at many locations (varying 10–20%) and (ii) the development of congestion does not only depend on the total flow (sum of mainline plus on ramp), but it is a function of the ratio of the two flows, especially at times close to the breakdown occurrence. This value of capacity drop is consistently larger than the ones observed in other studies for close locations (e.g. Zhang and Levinson, 2004). This phenomenon is observed often in Minnesota (MN) ramps because of violations of queue ramp constraints and the associated overreaction of the existing strategy with fast alternations of high and low ramp metered flows (like a bang–bang strategy). Another interesting observation is that capacity drop (and the uncongested/congested capacities) is quite similar across spans of years: (i) in 2000 when ramp metering strategy did not have any ramp delay constraints, (ii) in 2001 while ramp metering was out of operation and (iii) in 2008 when the ramp maximum waiting time constraint was active. In all cases capacity drop phenomena were of similar magnitude, not significantly affected by the different type of ramp control strategy.

Based on these empirical findings, in the second part of the paper 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. The main difference of the model when compared to other segmented LWR models (e.g. Cell Transmission Model; Daganzo, 1994) is that we incorporate capacity drop phenomena in the formulation. To do this, we propose a memory based step-wise-linear approximation of the flow–density relation that accounts for capacity drop effects. A Godunov numerical scheme is introduced to solve the formulated problem. Our results from real data implementation and a detailed micro-simulation for the aforementioned study sites 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 of capacity drop

The capacity of freeway sections is most commonly defined as the maximum flow possible at a specific location under the current circumstances. Active bottlenecks can influence the maximum flow upstream or downstream and cause congestion on

road networks. A bottleneck is a location where the full performance level (capacity) of an entire system cannot be achieved due to an abnormality at a single or multiple components 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 and congested conditions upstream. The term “active bottleneck” refers to the traffic conditions that the discharge rates measured downstream of a queue are not affected by conditions from further downstream (Daganzo, 1997). The capacity at a bottleneck can be defined as the maximum throughput possible at the bottleneck or the net maximum outflow exiting the bottleneck (Cassidy and Bertini, 1999).

2.1. An overview of previous work

While this maximum possible flow was traditionally considered a fixed value for a given location, many studies have revealed that there is a stochastic nature of bottleneck capacities. Hall and Agyemang-Duah (1991) and Banks (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. Some studies of the bottleneck capacity drop have suggested that the drop is non-noticeable or non-existent (Persuad, 1987); others place the drop ranging from about 3% (Banks, 1991) all the way to up to 12% (Cassidy and Bertini, 1999; Chung et al., 2007). 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. Many researchers suggest that capacity can therefore only truly be defined as a function of breakdown probability (Lorenz and Elefteriadou, 2007; Brilon et al., 2005). Cassidy and Bertini (1999) reports that while capacity at a bottleneck might have large variations; the critical density associated with the breakdown tends to be more stable with a smaller range of variation. A recent study in many freeways in the Twin Cities area, has observed capacity drops ranging from 3% to 12% (Zhang and Levinson, 2004). Average capacity drops have been estimated for different locations across many days, which does not allow to investigate dynamic characteristics of traffic flow phenomena and connect these drops with ratio of mainline over ramp flows.

Recently, many researchers have tried to identify the causes of capacity drop by looking at microscopic phenomena. Lane-changing maneuvers, vehicles entering a merge at low-speed and heterogeneous lane behavior have been considered to enlarge the capacity drop phenomenon (Cassidy and Ahn, 2005; Laval et al., 2005; Treiber et al., 2006). One of the reasons is that the aforementioned phenomena generate variations between different levels of saturated traffic states at merges (see for example Mauch and Cassidy, 2002; Ahn and Cassidy, 2007; Li et al., 2010) that prevent the system to reach the full free-way capacity before the breakdown.

Several models have been proposed to account for capacity drops in an exogenous way. For example, Koshi et al. (1983) and Hall and Hall (1990) propose 1st order models that are based on inverse lambda shape fundamental diagrams while Siebel et al. (2009) assume second order models derived from Greenberg (2001). Furthermore, most of these models are hard to implement and to calibrate in practice (Leclercq et al., 2011). A few efforts to integrate in a model local traffic interactions related to lane-changing maneuvers and their effect on the capacity can be found in Laval and Daganzo (2006) and Leclercq et al. (2011). The objective of our work is not to model or explain microscopically the phenomenon of capacity drop, but to integrate empirical findings of capacity drop magnitudes at different freeway locations in an efficient and accurate first order model.

While many 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 (see for example Papageorgiou and Kotsialos, 2002). 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 empirical investigation of many locations in Twin-Cities freeway system confirms the above findings and indicates a capacity drop in some critical locations of as high as 10–20% resulting in substantial miscalculation of the optimal metering rates. This suggests that a control strategy based on flow thresholds is likely to under-load 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 very similar for time periods with different control strategies in place.

The study site is a 12-mile segment of Trunk Highway 169 northbound (TH-169 NB), starting from the I-494 interchange and ending at 63rd Avenue North (Fig. 1). This site is a two lane per direction circumferential freeway traversing the Twin Cities west metropolitan region. It includes 10 weaving sections, 4 High Occupancy Vehicle (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 and severe congestion is experienced frequently at multiple locations.

2.2. Empirical observations in MN freeways

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, marked with a blue square in Fig. 1) for our analysis.

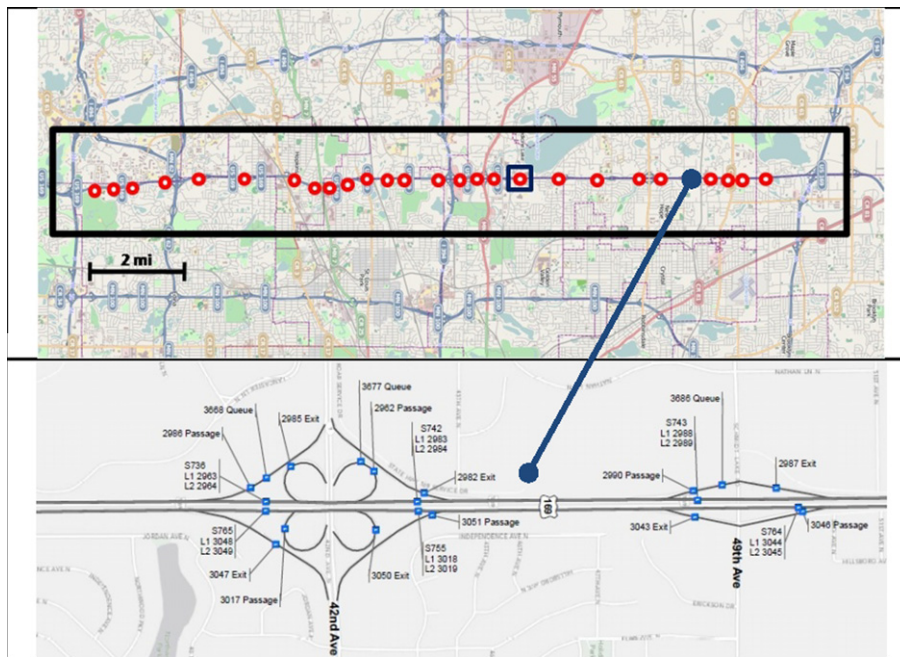


Fig. 1. The selected test site (TH-169 NB). TH 169 is a two lane freeway along the entire stretch with extended regions of acceleration lanes close to the two freeway exchanges. The red dots mark the location of mainline detectors. The lower portion of the figure shows a sample layout map made available by MnDOT that maps all detectors along stretches of the highway (credit for the map: ©OpenStreetMap contributors, credit for the detector layout: MnDOT All Detector Report). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Recurring congestion is evident during the evening hours (approximately 16:00–18:00) while the station downstream of the bottleneck is uncongested with speeds close to free-flow, confirming that the chosen site is an active bottleneck. 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 – Stratified Zone Metering in place). The traditional way to observe capacity drops (e.g. Cassidy and Bertini, 1999) 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. This way does not allow differentiating the effect of mainline vs. ramp flows as these are aggregated.

Another way is to plot a relation between the flow and the density (or occupancy) at the bottleneck site. 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 Fig. 2a. The graph shows the 30 s-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 density measure is an estimate based on converting the measured detector occupancies at the locations into density by calibrating each detector for an average field length factor as done by Minnesota Department of Transportation. The average field length for each detector combines the average vehicle length and the sensor sensitivity into a single factor. This is indeed not a true measure of density, but a reasonable approximation technique for locations where effective length does not vary much with time. Direct measurement of traffic density can occur only with cameras or if many vehicle trajectories are known.

The total throughput can be seen here to decrease from a maximum of ~ 4200 veh/h between 15:15–16:15 h, to ~ 3500 veh/h between 16:30–18:15 h (about 16.5% decrease). 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. Speed profiles have significantly higher values before the occurrence of the breakdown, at 16:20. The figure also shows the ramp demand 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. 2a), while similar total demand at 15:05 did not create a breakdown because on-ramp flow was lower. The flow vs. 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) is shown in Fig. 2b. High values of densities are observed because the location of the detector is slightly upstream of the merge location. Flow is measured as an estimate of the total flow (on ramp + mainline) at the bottleneck. This plot is useful to understand the value of capacity before and after the occurrence of the breakdown. Downstream conditions are always uncongested (smaller density and speed close to free flow). It also shows a capacity drop of the same magnitude. We further investigate in the following

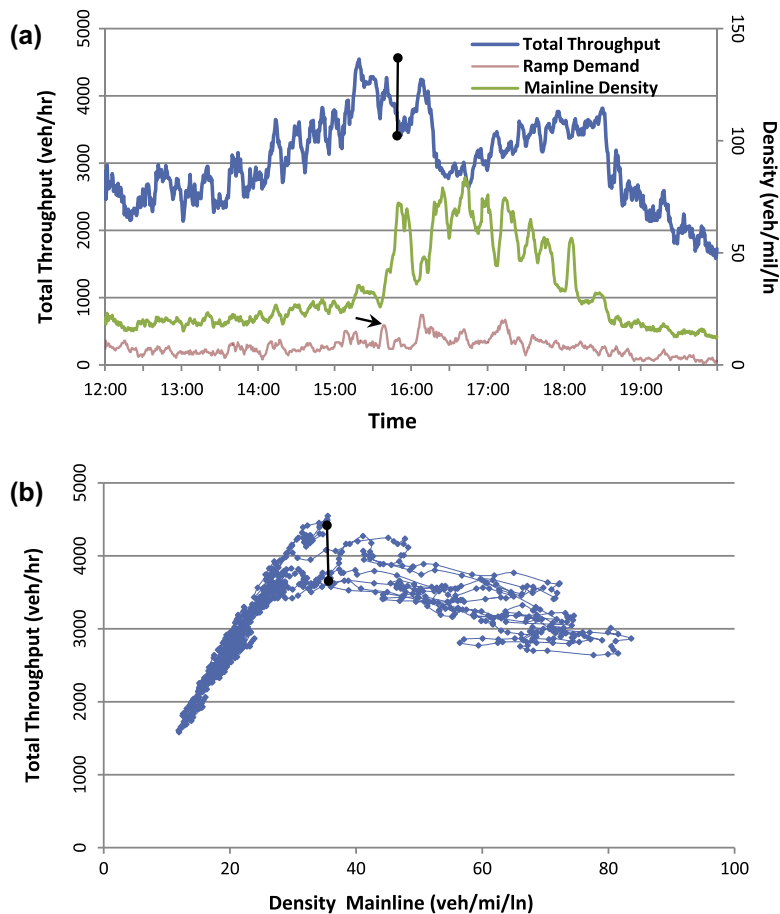


Fig. 2. Capacity drop identification at Plymouth ramp (a) time series plot of throughput at bottleneck; (b) fundamental diagram of total throughput flow (slightly downstream of the bottleneck) vs. mainline density (slightly upstream of the bottleneck). All values shown are 5 min averages taken every 30 s on 23rd September 2008.

portion of the paper the magnitude of these capacity drop values across a vast time horizon and under varying ramp control mechanisms.

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. The onset is defined as the duration between the time that speed starts decreasing monotonically (below the stable free-flow speed observed while location is uncongested), and the time when it reaches the critical density. Additionally, we plot a time series of (i) the ramp flows at the entrance of the freeway and (ii) occupancy of the ramp queue detector, to identify the instances that the ramp queue constraint is violated and metering priority switches to ramp clearance. Fig. 3 shows the phase diagrams for various days in 2000 (simple Zone Metering without ramp queue constraints), 2001 (no metering), and 2008 (Stratified Zone Metering – SZM). The SZM is a coordinated algorithm, which uses a flow equation to determine the metering rate for different zones of the freeway (with a predetermined constant value of capacity). Total ramp volume is distributed over all metered ramps in the zone in proportion to their demands. A maximum queue waiting time at the on-ramps has been integrated, which releases vehicles from the on-ramps at much higher rates, when the constraint is violated.

The left hand side plots of Fig. 3 show the phase diagrams between on-ramp and mainline flow for different days. The right hand side plot shows the time series of metering rates (flow) at the downstream on-ramp detector (measuring the actual ramp flow entering the mainline) and the occupancy measures at the upstream queue on-ramp detectors. For earlier years (2000, 2001) occupancy measurements are from downstream ramp detectors, as queue detectors were not instrumented at that time.

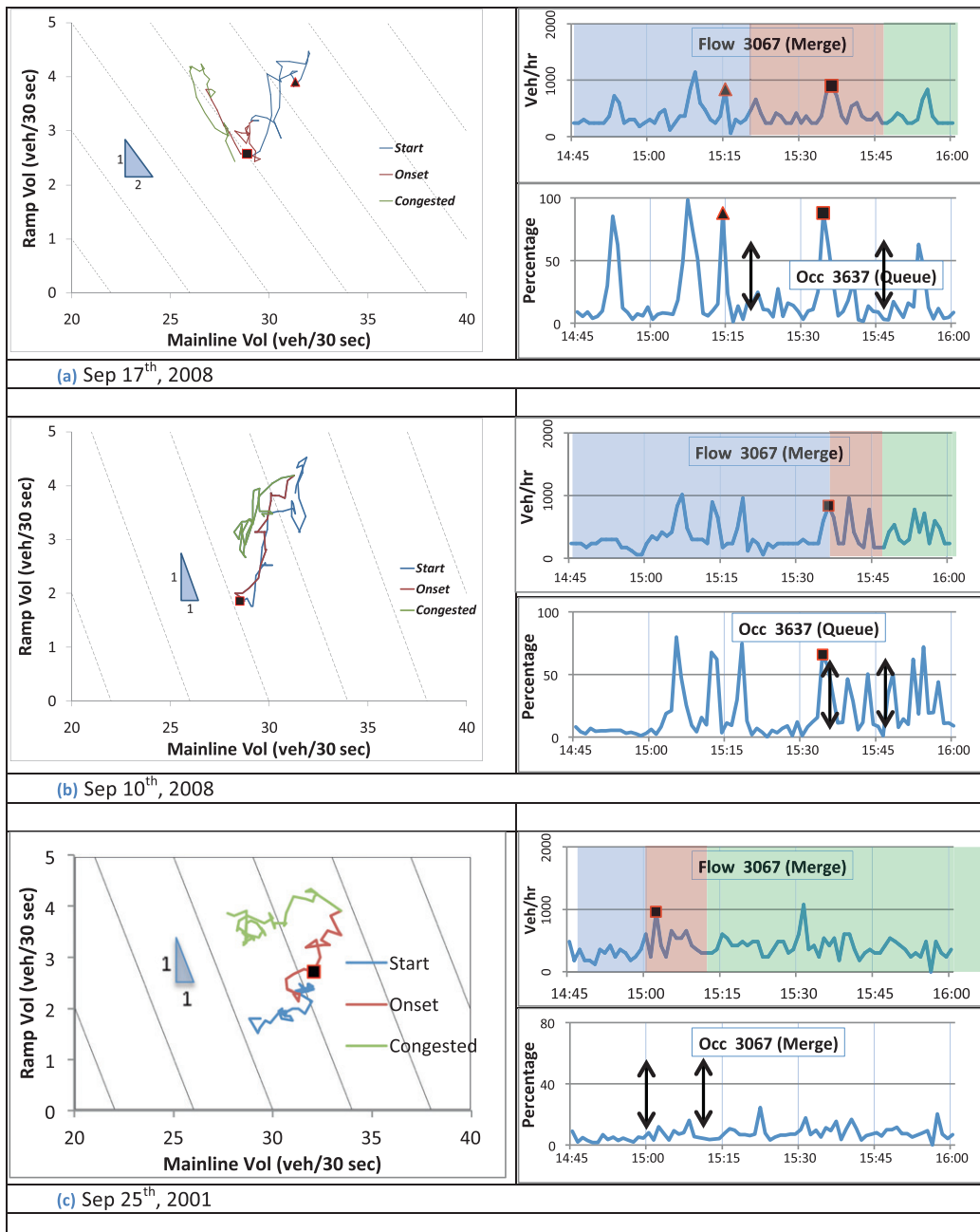


Fig. 3. Bottleneck flow contribution plots showing mainline volumes vs. ramp volumes for different days. (a) September 17th, 2008; (b) September 10th, 2008; (c) September 25th, 2001; (d) September 26th, 2001; (e) November 15th, 2000. The plots show a consistent pattern across the dates and suggest a consistent capacity drop of ~15% for the location.

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). 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 phase diagrams (left) as well as the demand time series graphs (right) for better understanding. All plots of Fig. 3 show 1:1 flow isoquants (negative 45° slope if axes were equally scaled) as overlay, with exception of Fig. 3a (slope 1:2), which will be explained later in the final part of the current section.

Fig. 3a represents roughly a 60 min span of data between 15:00 and 16:00 h (30 min in free flow conditions, 10 min in onset of congestion with decreasing speeds, and 20 min of congestion with significantly low speeds). The left side of the figure shows the mainline volume vs. ramp volume plot for September 17th 2008, along with two time series plots for ramp

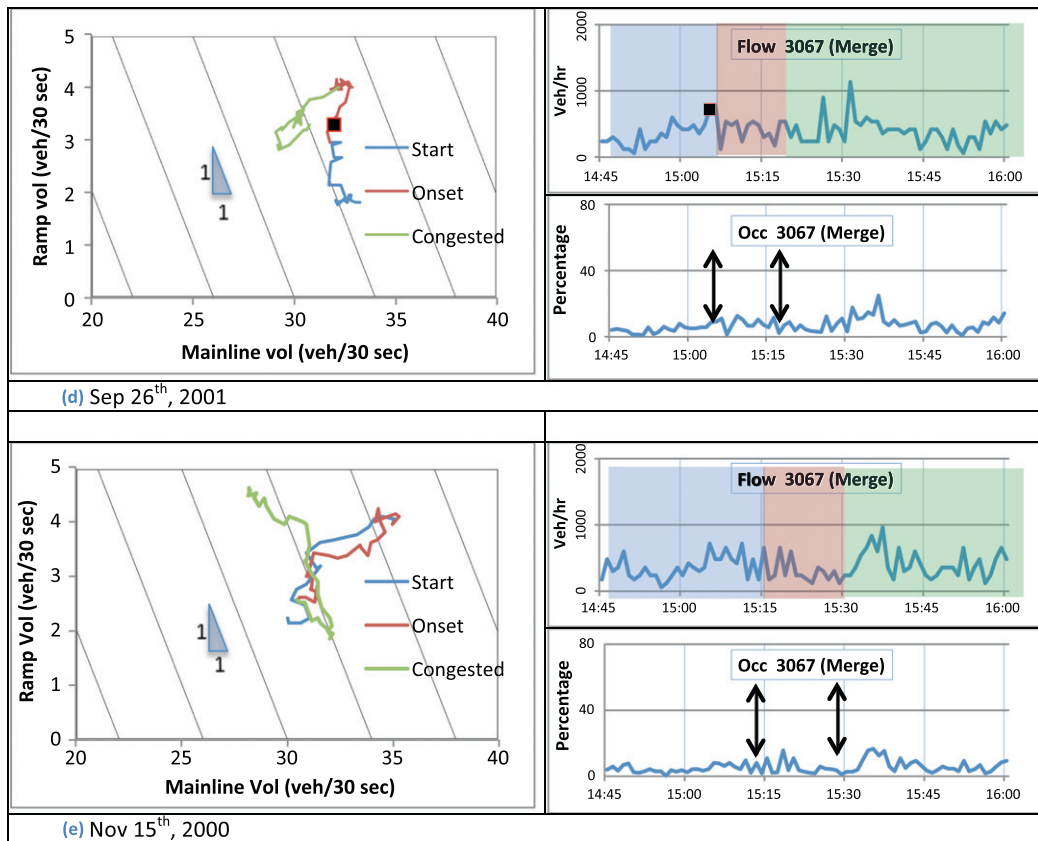


Fig. 3. (continued)

measurements, (i) the occupancy at the upstream on ramp detector (queue detector) identifies when long queues occur at the on-ramp, which will activate the maximum waiting time ramp constraint and release more vehicles to the freeway, and (ii) the ramp supply at the merge detector location, which shows at what rate vehicles are entering the freeway from the ramp. Note that when occupancy at the queue detector is not large, then the queue detector measures the demand and the merge detector measures the supply, while when this value is large, queue exceeds the detector and demand is unknown. The existing ramp metering strategy will significantly increase the metering rate to avoid long ramp queues, when queue exceeds the upstream detector. The phase diagram plots are however smoothed out using 3 min moving average time windows observed every 30 s. This smoothing is done in order to observe the trend of volumes more clearly and to reduce the high disturbances in 30 s data that might reflect neither steady nor mean states (and thus not be representative of the capacity of bottleneck). The ramp detector data still represent 30 s resolution to catch the spikes in the ramp volumes, which are expected to be the prime reason for the breakdowns and the strong capacity drop. The smoothing however has two effects on the plotting which need to be kept in mind while analyzing the graphs: (i) there is a small inherent time lag introduced because of the moving average, and (ii) the smoothed values of volume in the phase diagrams might not match perfectly the instantaneous values reported on the time series plot.

The flow of time for the graph is from the blue segment representing the period before onset of congestion (Start), the red¹ segment that depicts the time when the congestion starts building (Onset) and then to the green segment when the location is under consistently congested phase (Congested). We can see from the two right plots that the queue detector starts registering high demands 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 36.2 veh/30 s (4320 veh/h). This throughput is sustained for approximately 5 min before the consistently high ramp discharge rate causes a breakdown at 15:15, which corresponds to the marked triangular peak in the flow distribution graph. Once the breakdown happens, the high volumes are no longer sustainable and the capacity falls by approximately 15%. The plot in the phase diagram shows the fall of throughput along the blue segment and leading into the red and green phases. It is clear that once the breakdown happens, the throughput stagnates at a capacity of about 31.8 veh/30 s, a 15% decrease from the initial peak.

¹ For interpretation of color in Figs. 3, 4, 8 and 10, the reader is referred to the web version of this article.

Fig. 3b narrates a similar story on September 10th 2008. High demands are initially observed between 15:08 and 15:20. The phase diagram here, follows the blue segment up to a peak (due to high demand on ramp), 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 is followed by lower mainline flows past 15:20 (with periods of low ramp demand that allow the buildup to clear out) and is able to avoid the onset of congestion. The throughput once again increases to its peak levels (merge of the red and green segments) at 15:35. The consistent high demand this time causes a breakdown at the location and the throughput first rises along the red line, reaches breakdown, and finally falls to and stabilizes at the congested capacity along the green segment. 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 min. This figure 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.

It can be observed from Fig. 3, that the pre-congestion (~ 72 veh/min) and post-congestion (~ 62 veh/min) capacities at the bottleneck, and hence also the drop in capacity ($\sim 15\%$), are roughly the same for all 3 years considered. Nevertheless, the fact that the capacity drop is the same with and without ramp metering is contradicting the main purpose of ramp metering which is to improve mainline conditions. Improved ramp metering strategies could be expected to create smaller capacity drops, when compared with the no control case. The high stop-and-go effect created due to an inadequately designed ramp metering (as seen in Fig. 4 below), suggests that an improved strategy could possibly decrease the magnitude of capacity drop. However, this is a future research direction and the present work cannot provide a full conclusion towards this hypothesis. We might conjecture that the specific ramp metering strategy may not be very efficient for conditions close to breakdown. This value of capacity drop is consistently larger than the ones observed in similar locations (see for example Zhang and Levinson, 2004). The main reason for this high drop is the extremely high values of ramp flows a few minutes before the occurrence of the breakdown, due to the overreaction of the existing ramp metering strategy once long queues occur at the on ramps. Our recent findings (Geroliminis et al., 2011) show that a ramp metering strategy with a smoother reaction to long queues can significantly improve total travel times.

As seen in Fig. 4 (time series of total throughput in red and ramp rates in blue), capacity before breakdown at 3:35 pm is around 4400 veh/h. The high increase in ramp flows (~ 1000 veh/h) creates a breakdown and the flow after 5 min decreases to 3600 veh/h followed by a speed decrease (< 35 mph). Downstream of this location speed is still close to 60 mph. Demand decreases after this time and speed returns close to free-flow value around 4:05. At this time queue ramp constraint is violated and ramp rates increase again. A 2nd breakdown occurs and we observe approximately the same flow decrease (from 4400 to 3600 veh/h). Very high ramp rates continue to occur and the result is a further decrease in the bottleneck output (~ 2800 veh/h). The reason for this inefficiency is that the number of vehicles entering the freeway is very large and density increases at values which the bottleneck cannot discharge at capacity (as it can be seen in Fig. 2). We do not consider the latter as capacity drop because density is much higher than the critical value of density. Nevertheless, it is clear that despite the fact that there is no restriction downstream vehicles cannot discharge at higher values of flow as their speed is very small (~ 10 mph). After 5:15 pm the ramp rates decrease and the discharge flow of the bottleneck increases again to 3600 veh/h (bottleneck is still active after 5:15 pm).

To further support the relation between capacity drop and on-ramp flows, we present in Fig. 5 data from another congested day, but with smaller ramp flows. It is clear that capacity drop is much smaller (6–8%) and ramp flows never increase more than 600 veh/h (compared to 1200 veh/h in the previous cases). We have analyzed additional locations and the same patterns occur, which connect the magnitude of the capacity drop with the ramp rates.

There is also an interesting observation that is derived directly by observing the inclination of the curves to the two axes, in the phase diagrams. The inclination (during the congested phase) from Fig. 3a suggests that the volumes stabilize along an

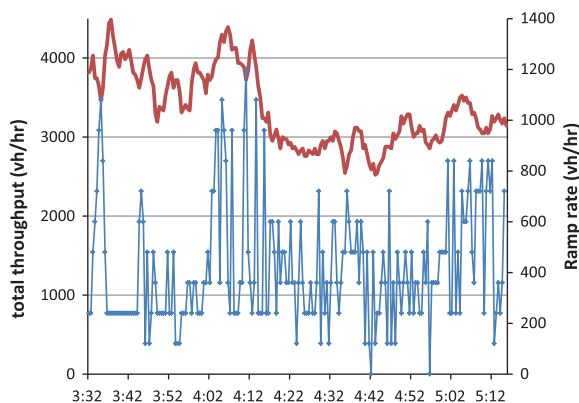


Fig. 4. Time series of total throughput and ramp rates at Plymouth location during the onset of congestion.

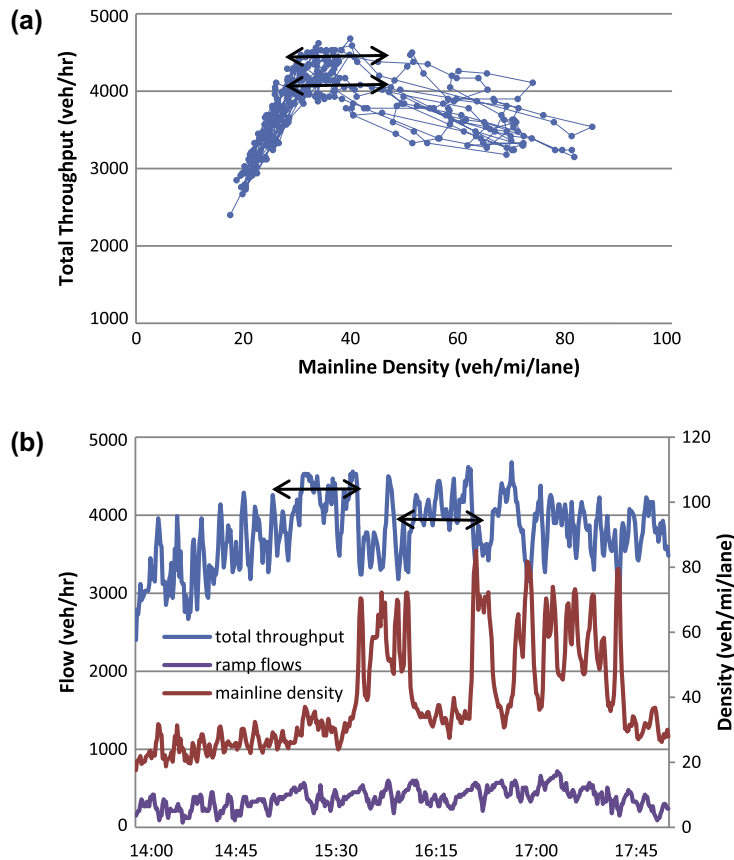


Fig. 5. Empirical data for a smaller capacity drop at Plymouth location: (a) total bottleneck throughput vs. mainline density upstream of bottleneck; (b) time series of total throughput, on-ramp flows and mainline density (red double arrows mark the estimate of uncongested and congested capacities). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 an additional vehicle on the mainline. This observation could possibly be attributed to the merging behavior and its effect on the flow of traffic. 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). Multiple bottlenecks need to be explored, and in a vaster time horizon, in order to be able to support such a proposition. Exploring this phenomenon can be a future research direction. The various plots however, give consistent and convincing proof towards the existence of capacity drops and their independence from demand pattern variations.

3. A first order traffic model with internal boundaries and capacity drop

After describing the phenomenon 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 at the ramps. Current metering strategy in the Twin-Cities network and in other locations estimate densities within a section using a linear interpolation between known detectors at section boundaries. 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 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 density profile within the interior of the section often peaks at values higher than the densities observed at the boundaries (upstream and downstream ends). Following these observations, we propose an extended first order model, which integrates capacity drop phenomena and that can be efficiently used to estimate densities along the stretch. Afterwards, we present some comparison studies against real traffic measurement as well as simulation data.

For the purpose of the study, we use the same stretch of US Highway 169 Northbound shown in Fig. 1, 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 is a site of an active bottleneck consistently.

As the existing data cannot give full information about the density profile, we first provide empirical validation of the model, given the limitations of data availability and then we also model the traffic behavior along the highway stretch using microscopic traffic simulation software (AIMSUN). The simulator uses demand values at all access points to the network, and 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. The parameters of the simulator for this network have been calibrated in previous studies (Feng et al., 2006). Additional detectors were placed in the simulation along the stretch of freeway in order to observe 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 (as currently utilized by SZM metering algorithm), shown in Fig. 6. Fig. 6a and b shows 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 density estimation model.

3.1. Extended first-order modeling of traffic flow

The LWR theory is a well-established continuum theory for traffic 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 and a known relation between flow and density, depicted through the representation of stream speed as a function of density. Traditionally a known state equation (a fundamental diagram) is used to determine this stream speed–density dependence. As highlighted before, 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) and we integrate capacity drop in the model. Traffic state at the boundaries (time and space boundaries) along with a flow–density relationship for the section is provided as input. The main methodological contribution of our model is the integration of the capacity drop phenomena in a first order model, using a memory-based approach. Another 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 the obtained solutions 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 segmented LWR model is applied to each such segment separately. Each segment is broken down into distance steps of length Δx and similarly time is discretized into Δt 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 $g(q)$. The Courant–Friedrichs–Lewy's, 'CFL' (Courant et al., 1928) condition to guaranty stability and convergence of solution require that Δx and Δt should be chosen such that $\Delta x/\Delta t \geq u_f$ (free-flow speed).

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 that helps to incorporate a memory-based decision. It utilizes the previous interval of density data collected at the location to predict whether the location is currently under free flow or congested regime. Thus, for the same density value, a higher flow is estimated (corresponding to free flow capacity) if the densities observed are lower than the critical density for the location, while a lower flow is estimated (corresponding to congested capacity) if the location is congested.

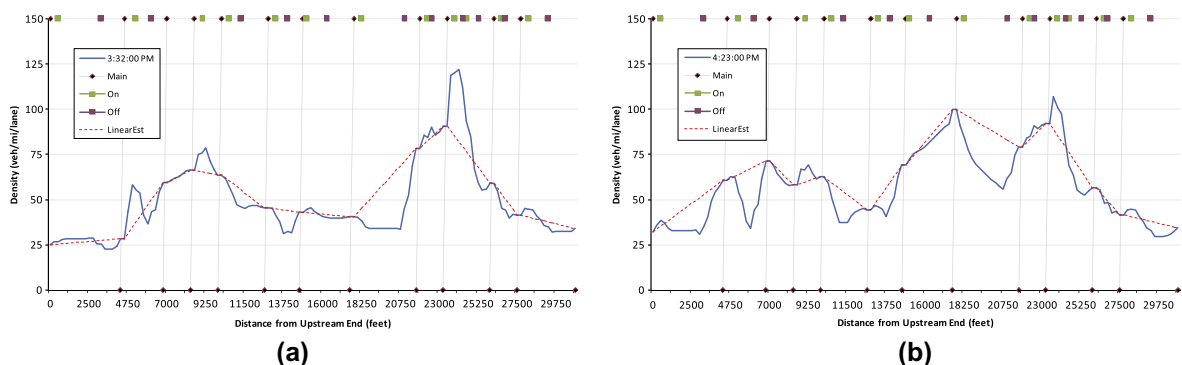


Fig. 6. Simulation obtained density vs. linear approximation model for two different times.

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 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.

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 (e.g. Michalopoulos et al., 1984). 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 is smaller).

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 a stepwise linear estimation of the flow–density relation, while accounting for capacity drop. Fig. 7 shows the flow density relationship estimation used for this purpose for a few sample locations along the freeway. We have to point out that if the intermediate locations experience high errors in the detector measurements, the segmented LWR model can lead to similar or larger errors. Thus, an analysis in the quality of the data should be investigated before the implementation of the model and possibly exclude detectors with faulty or biased measurements.

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. 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 .

The first two blocks typically represent the uncongested phase of traffic. Block 1 is for light conditions where vehicles run at free-flow speed. Block 2 is under-saturated, but the effect of vehicle interactions 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 incorporates both the segmented nature, and the memory based flow–density approximation model described. The memory retaining nature of the estimation utilizes different values of capacity based on whether the conditions have been congested at the location or not. The fundamental diagram is expressed by a set of 5 lines, each of them having a constant part (c_i) and a slope (m_i) as shown in Fig. 7 (index of location is omitted from the equation):

$$q(k) = \begin{cases} c_1 + km_1 & \text{if } 0 < k < k_1 \text{ and } s = 0 \\ c_2 + km_2 & \text{if } k_1 < k < k_2 \text{ and } s = 0 \\ c_3 + km_3 & \text{if } k_2 < k < k_3 \text{ and } s = 0 \\ c_4 + km_4 & \text{if } k_2 < k < k_4 \text{ and } s = 1 \\ c_5 + km_5 & \text{if } k > k_4 \text{ and } s = 1 \end{cases} \quad (1)$$

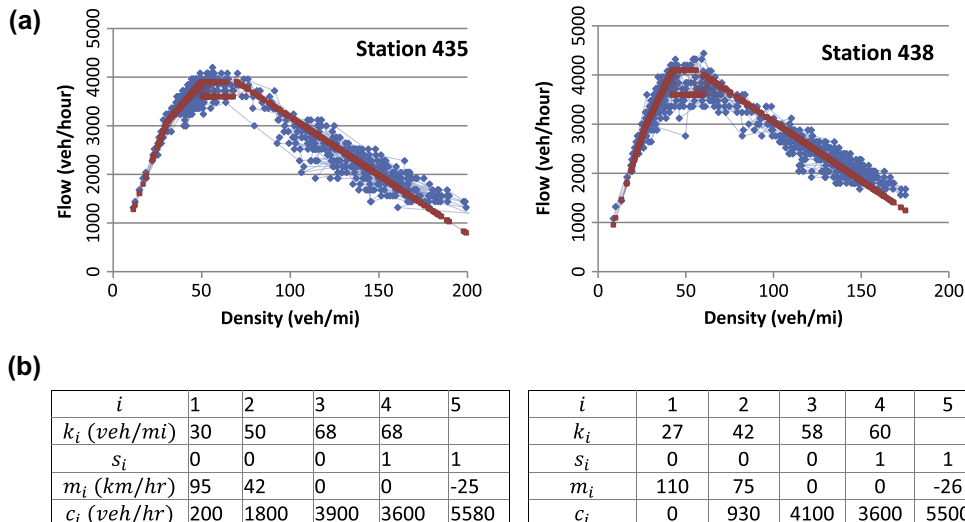


Fig. 7. (a) Application of LWR model to a section of freeway. (b) “5-step” stepwise linear flow–density relation with two-capacity values.

Many numerical methods have been proposed to solve the LWR model. One approach is to solve the Riemann problem and apply a Godunov method for this model (Godunov, 1959). Both solutions to the Riemann problem and the Godunov method are well developed for hyperbolic conservation laws (Smoller, 1983). A different approach is to utilize demand and supply functions (Daganzo, 1995; Lebacque, 1996), which turnout to be variants of Godunov's method (Jin et al., 2009). Eqs. (2) and (3) describe the traditional LWR theory formulation, while Eqs. (4)–(7) provide the numerical approximation of an upwind Godunov scheme. Eq. (2) is a mass conservation equation between flow q and density k in time and space, expressed as a hyperbolic partial differential equation, while (3) 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, while it is zero at all other locations. The set of Eqs. (4) and (5) 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 (similar to Leclercq et al., 2007; Lebacque, 1996). Entrance and exit flows at ramps are considered as local inputs (not distributed over multiple distance steps), thus flow rates at ramps (g_j) are used in place of the generation function. A profile of the density distribution is thus created along the time and space dimensions (*subscripts in Eqs. (4) and (5) represent distance step*). The exit flow of cell i at time step $t + 1$, e_i^{t+1} , is the minimum of the demand function at cell i , $\lambda(k_i^t)$ and supply function at cell $i + 1$, $\mu(k_{i+1}^t)$. Demand and supply functions are given by Eqs. (4) and (5).

$$\frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = g(x, t) \quad (2)$$

$$u = u_e(k) \quad (3)$$

$$e_i^{t+1} = \min(\lambda(k_i^t), \mu(k_{i+1}^t)) \quad (4)$$

$$\lambda(k_i^t) = \min(v_f k_i^t, c_2 + m_2 k_i^t, q_m^h) \quad (5)$$

$$\mu(k_{i+1}^t) = \begin{cases} \min(w(k_{i+1}^t - k_{cr}^+), q_m^h) & \text{if } s_{i,t-1} = 0 \text{ (uncongested)} \\ \min(w(k_{i+1}^t - k_{cr}^+), q_m^l) & \text{if } s_{i,t-1} = 1 \text{ (congested)} \end{cases} \quad (6)$$

$$s_{i,t} = \begin{cases} 1 & \text{if } (k_i^t > k_{cr}^+) \text{ OR } ((s_{i,t-1} = 1) \text{ AND } (k_i^t > k_{cr}^-)) \\ 0 & \text{o.w.} \end{cases} \quad (7)$$

where q_m^h, q_m^l is the high and low value of capacity, and k_{cr}^-, k_{cr}^+ is the range of density for the capacity (block 3 and 4 in Fig. 7), v_f is the free flow speed (slope m_1 in Fig. 7) and w is the congested wave speed (slope m_5 in Fig. 7). Note that $\lambda(k_i^t)$ has three parts because of the piecewise linear assumption of the FD. Also note that $m_3 = m_4 = 0$, and $c_3 = q_m^h; c_4 = q_m^l$. Eqs. (6) and (7) describe the memory-based conditions of the fundamental diagram through a binary variable $s_{i,t}$ to account for congested conditions and capacity drop. To avoid sharp and quick fluctuations between congested and uncongested states, a moving average e.g. every 3 min can be utilized to smoothen the variations in state. Then the density at every time step and cell is updated as follows (g_i^t is the inflow (or outflow) from an on-ramp (or off-ramp) and it has a positive value for the on-ramp and negative for the off-ramp):

$$k_i^{t+1} = k_i^t + (e_{i-1}^{t+1} - e_i^{t+1}) * \left(\frac{\Delta x}{\Delta t}\right) + g_i^t * \left(\frac{\Delta x}{\Delta t}\right) \quad (8)$$

3.2. Validation against real observations

For the purpose of the study, we use a 6 mile 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). Mainline detector measurements were available at 13 locations along the study stretch. The 30 s traffic demand data used in the comparison were extracted from the Minnesota DOT loop detector database. For the purpose of this study pm peak data from September 25th, 2008 between 2 pm and 8 pm was used.

The first validation comparison was to estimate conditions at known detector locations along the freeway when they are not used as section boundaries. In order to perform this test, only a fraction of the measurable locations were selected as boundary conditions, while others were reserved as validation locations. The conditions at these 'validation locations' were then estimated by applying the model to sections defined by the selected 'boundaries' and the estimated densities were compared against the observations measured at these locations (see Fig. 8 for an illustration). Fig. 8 also shows contour plots of density against space and time, for three cases being compared: (a) the actual measured densities at the locations; (b) the density estimates obtained from a segmented LWR model that uses alternate locations (every second detector) as boundaries; and (c) the density estimates from a full span model that only uses the extreme ends as boundaries. The figure clearly shows the advantage to be gained from using segmented models over full-span models. A time series plot of the densities estimated by the models for two locations under different segmentation configurations (Fig. 9) further demonstrates how a segmented model can be expected to be far more accurate than the longer span models. Table 1 presents the comparison of the mean errors in densities predicted under the three segmentation schemes (as shown in Fig. 8b).

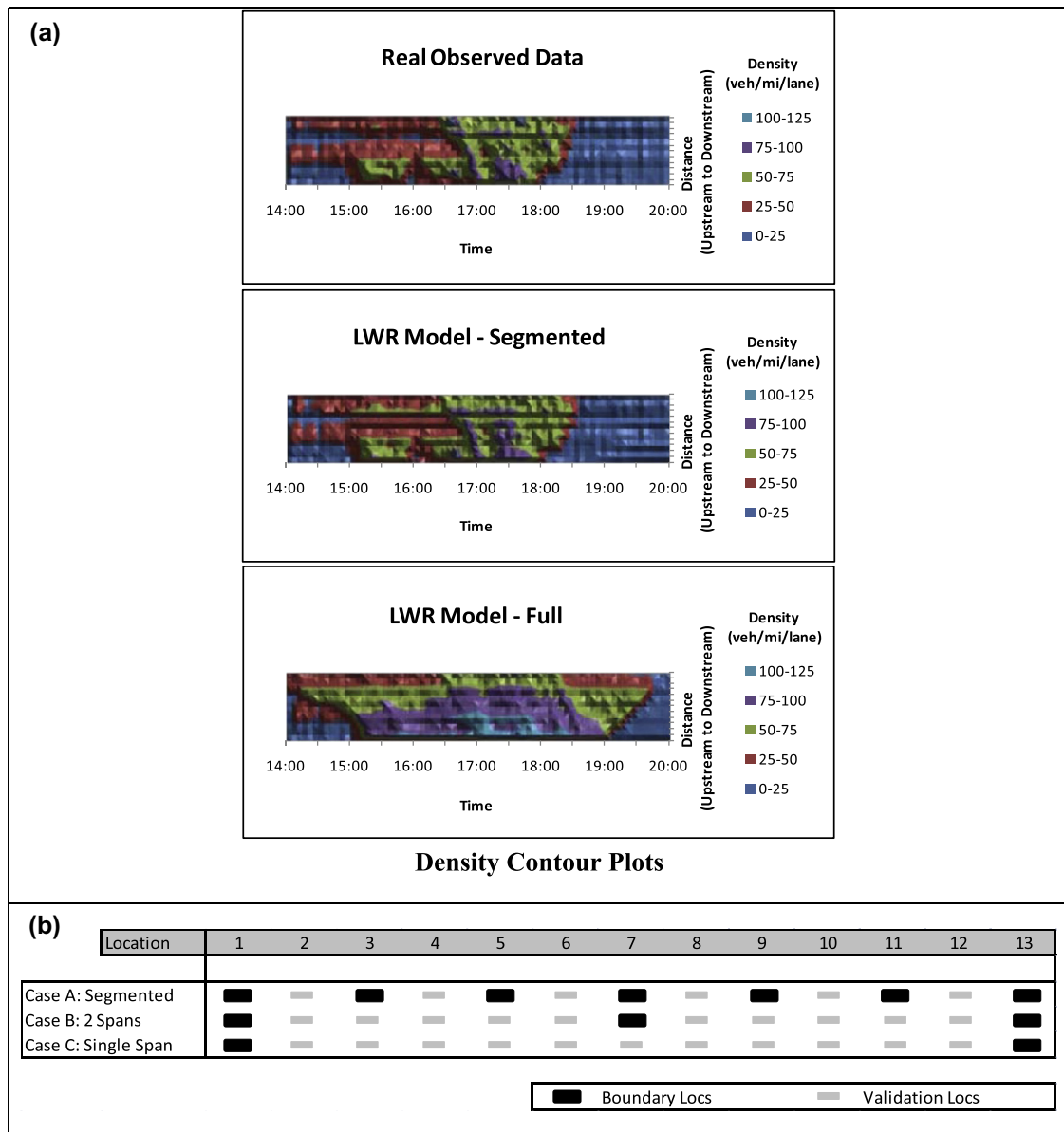


Fig. 8. (a) Density contour plots comparing real observed densities at validation detectors vs. traditional LWR vs. segmented LWR with capacity drops (the plot shows average densities every 5 min period). (b) Detector locations used as boundaries and for validation for the three different case studies.

A second comparison experiment was performed in order to compare the effect of including capacity drop integration into the model as presented earlier in the paper through the 5-step piecewise linear fundamental diagram. The segmented model (Case A) was used under the following configurations of the fundamental diagram: (a) with two-value capacity drop incorporated; (b) using a single capacity equal to the average capacity, and (c) using a single capacity equal to the congested capacity at all locations. Table 2 presents the comparison of the mean errors in predicted densities under different configurations of the fundamental diagrams. The table clearly suggests that there is an increase in accuracy when including the capacity drop in a memory-based model in the model and when higher segmentation is utilized.

3.3. Simulation results

While the comparison against real observations described in the previous section provides for a limited validation process (given that data between detector locations are not available), a simulation-based comparison was also performed to test the model performance at a higher resolution, where density is a continuous function in space. All 13 detector locations were

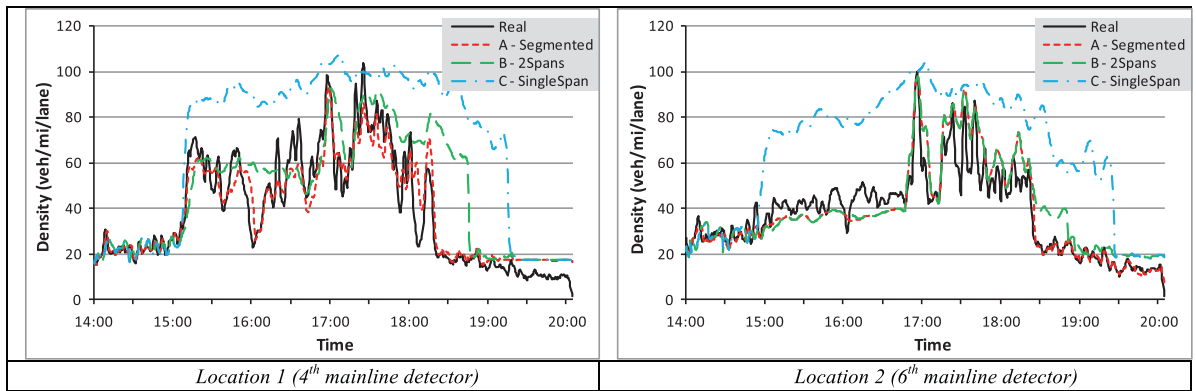


Fig. 9. Time series plot of density at two validation locations comparing real observations vs. LWR model under different segmentation schemes. In Case A, every alternate detector location is marked as a boundary and only half the detectors are used for validation. In Case B, the first, seventh, and thirteenth detectors are used as boundaries, while in Case C, only the first (upstream end) and thirteenth (downstream end) locations are boundaries while all other locations are used as validation locations. Thus, the two reported locations in the figure are validation locations for all three cases.

Table 1

Mean absolute error in (veh/mile) and mean absolute % error for different segmentation levels with capacity drop.

Location	Segmentation level setting (with capacity drop)					
	Case A		Case B		Case C	
	MAE (veh/mile)	MAPE (%)	MAE (veh/mile)	MAPE (%)	MAE (veh/mile)	MAPE (%)
2	10.2	30.3	17.4	51.6	38.2	113.2
4	5.6	14.3	12.7	32.6	30.1	77.6
6	5.2	13.8	7.4	19.6	28.6	76.0
8	9.9	27.5	24.9	69.3	33.3	92.7
10	11.9	30.9	23.1	59.9	27.9	72.4
12	5.9	15.2	8.2	21.1	9.3	23.7
Total	8.1	22.0	15.6	42.4	27.9	75.9

used as section boundaries in the segmented LWR application and the densities at all intermediate locations were compared between the estimation from the LWR model, and the simulation results.

Fig. 10 shows a contour plot depicting density profile against time (horizontal axis) and space (vertical axis) for the three cases being compared: Results from the micro-simulation, the segmented LWR model with capacity drop, and the simple Full Span LWR model. 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. 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 closer to the simulation reported values of density throughout the study duration. 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 veh/mile) which closely approximates the critical density along the stretch, and 30 veh/mile as shown in Fig. 11. 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, that can be integrated in a ramp metering algorithm. Note that a Full Span LWR model without capacity drop integration cannot capture the offset of congestion around time 19:30. These density thresholds are widely utilized in ramp metering strategies (see for example Papageorgiou and Kotsialos, 2002).

Table 2

Mean absolute error in (veh/mile) and mean absolute % error for different configurations of FDs.

Location	Capacity setting in FD (segmented every 2nd detector)					
	2-value capacity		Cong. capacity		Median capacity	
	MAE (veh/mile)	MAPE (%)	MAE (veh/mile)	MAPE (%)	MAE (veh/mile)	MAPE (%)
4	6.6	15.7	13.9	33.2	9.0	21.6
6	6.4	15.2	11.6	27.6	7.4	17.7
12	5.7	15.3	13.3	35.8	12.5	33.7

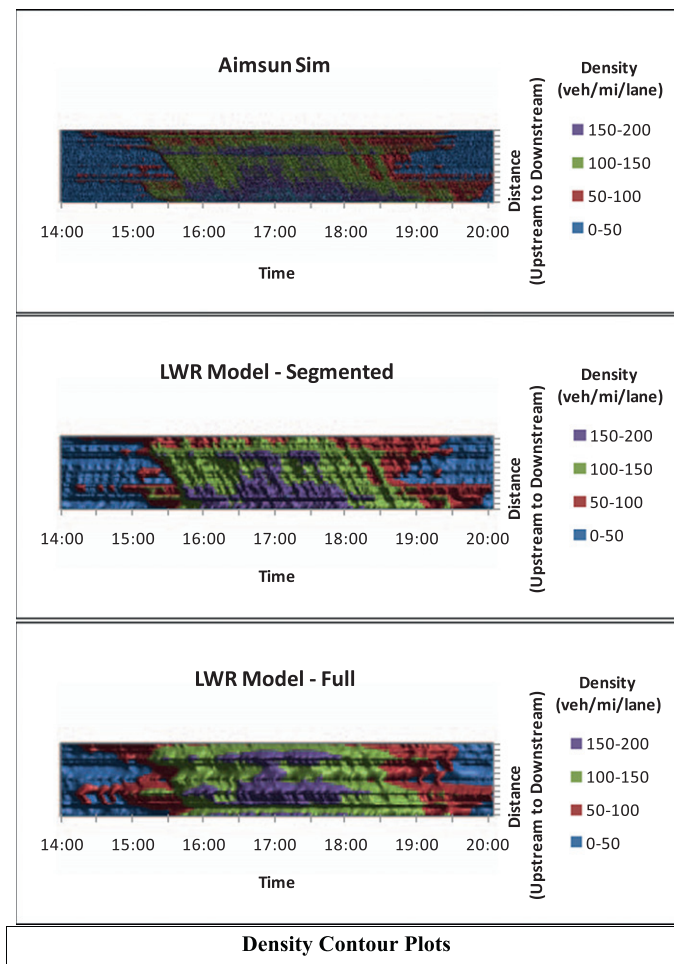


Fig. 10. Density contour plots comparing AIMSUN vs. traditional LWR vs. segmented LWR with capacity drops.

Nevertheless, we have to highlight that most of the existing micro-simulation models have difficulties in expressing with accuracy traffic flow phenomena related to stop-and-go traffic, capacity drop, merge behavior and traffic oscillations. So, a more detailed experiment could be utilized to validate the proposed and other newly developed models of traffic flow.

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 freeway system. 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 queue constraint was active. In all cases capacity drop phenomena were of similar magnitude not affected by the different type or ramp control strategy. Improved ramp metering strategies are expected to create smaller capacity drops, when compared with the no control case. In the second part of the paper, we develop a methodology to estimate densities with space and time based on data from loop detectors by integrating the capacity drop in a first order model. 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

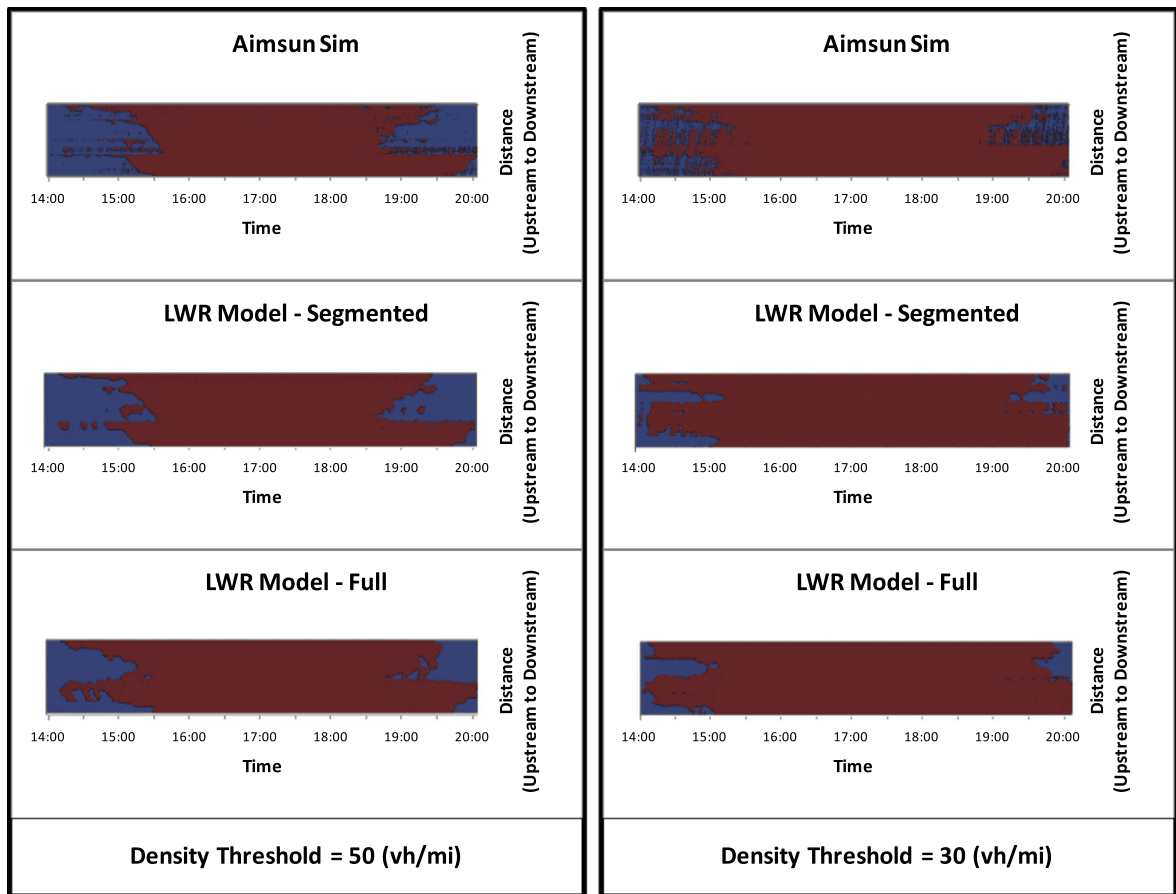


Fig. 11. Binary density contour plots for AIMSUN vs. segmented LWR vs. full length LWR model.

solution of the problem. Results compared with real data and micro-simulation of the Highway H-169 stretch show that the model produces promising results. Future research direction can integrate the proposed model in a ramp metering algorithm and validate its accuracy with detailed empirical data in multiple freeway sections. This additional testing will allow more in-depth evaluation, parameter optimization, and potential improvements for the new model.

Acknowledgments

This study was partially supported by the ITS Institute of the Center for Transportation Studies at the University of Minnesota. The authors acknowledge MN/DoT's Regional Traffic Management Center and Profs. John Hourdos and Panos Michalopoulos of Un. of Minnesota, for their valuable comments, support, and continuous cooperation.

References

- Ahn, S., Cassidy, M.J., 2007. Freeway traffic oscillations and vehicle lane-changing maneuvers. In: *International Symposium of Traffic and Transportation Theory*. Elsevier, Amsterdam, pp. 691–710.
- Banks, J.H., 1991. Two-capacity phenomenon at freeway bottleneck. *Transportation Research Record* 1320, 83–90.
- Brilon, W., Geistefeldt, J., Regler, M., 2005. Reliability of freeway traffic flow: a stochastic concept of capacity. In: *16th International Symposium of Transportation and Traffic Theory*. Elsevier, Maryland, pp. 125–143.
- Cassidy, M.J., Bertini, R.L., 1999. Some traffic features at freeway bottlenecks. *Transportation Research Part B* 33, 25–42.
- Cassidy, M.J., Ahn, S., 2005. Driver turn-taking behavior in congested freeway merges. *Transportation Research Record* 1934, 140–147.
- 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.
- Courant, R., Friedrich, K., Lewy, H., 1967 [orig.: 1928]. On the partial difference equations of mathematical physics. *IBM Journal of Research and Development* 11(2), 215–234.
- Daganzo, C.F., 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B* 28 (4), 269–287.
- Daganzo, C.F., 1995. Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B* 29, 277–286.
- Daganzo, C.F., 1997. *Fundamentals of Transportation and Traffic Operations*. Pergamon-Elsevier, Oxford, UK.
- Feng, B., Hourdos, J., Michalopoulos, P., 2006. Improving Minnesota's stratified ramp control strategy. *Transportation Research Record* 1959, 77–83.

- Geroliminis, N., Srivastava, A., Michalopoulos, P., 2011. A dynamic-zone-based coordinated ramp-metering algorithm with queue constraints for Minnesota's freeways. *IEEE Transactions on Intelligent Transportation Systems* 12 (4), 1576–1586.
- Godunov, S.K., 1959. A difference method for numerical calculations of discontinuous solutions of the equations of hydrodynamics. *Matematicheskii Sbornik* 47, 271–306 (in Russian).
- Greenberg, J., 2001. Extensions and amplifications of a traffic model of Aw and Rascle. *SIAM Journal on Applied Mathematics* 62, 729–745.
- Hall, F.L., Agyemang-Duah, K., 1991. Freeway capacity drop and definition of capacity. *Transportation Research Record* 1320, 91–98.
- Hall, F.L., Hall, L.M., 1990. Capacity and speed-flow analysis at the queen Elisabeth way in Ontario. *Transportation Research Record* 1287, 108–118.
- Helbing, D., Johansson, A.F., 2009. On the controversy around Daganzo's requiem for and Aw–Rascle's resurrection of second-order traffic flow models. *The European Physical Journal B* 69, 549–562.
- Helbing, D., 2001. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics* 73 (4), 1067–1141.
- Jin, W.-L., Chen, L., Puckett, E.G., 2009. Supply-demand diagrams and a new framework for analyzing the inhomogeneous Lighthill–Whitham–Richards model. In: 18th International Symposium on Transportation and Traffic Theory, pp. 603–635.
- Koshi, M., Iwasaki, M., Okhura, I., 1983. Some findings and an overview on vehicular flow characteristics. In: 8th International Symposium on Transportation and Traffic Theory. University of Toronto Press, Toronto, pp. 403–426.
- Laval, J.A., Cassidy, M.J., Daganzo, C.F., 2005. Impacts of lane changes at on-ramp bottlenecks: a theory and strategies to maximize capacity. In: Kühne, R., Poeschl, T., Schadschneider, A., Schreckenberg, M., Wolf, D. (Eds.), *Traffic and Granular Flow '05*. Springer, Berlin, pp. 577–586.
- Laval, J.A., Daganzo, C.F., 2006. Lane-changing in traffic streams. *Transportation Research Part B* 40 (3), 251–264.
- Lebacqz, J.P., 1996. The Godunov scheme and what it means for first order traffic flow models. In: *The International Symposium on Transportation and Traffic Theory*, Lyon, France.
- Leclercq, L., Laval, J., Chevallier, E., 2007. The Lagrangian coordinate system and what it means for first order traffic flow models. In: 17th International Symposium on Transportation and Traffic Theory (ISTTT). Elsevier, New York, pp. 735–753.
- Leclercq, L., Laval, J., Chiabaut, N., 2011. Capacity drops at merges: an endogenous model. *ISTTT – Transportation Research Part B* 45 (9), 1302–1313.
- Li, X., Peng, F., Ouyang, Y., 2010. Measurement and estimation of traffic oscillation properties. *Transportation Research Part B* 44 (1), 1–14.
- 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 A* 229, 317–345.
- Lorenz, M.R., Elefteriadou, L., 2007. Defining freeway capacity as function of breakdown probability. *Transportation Research Record* 1776, 43–51.
- Mauch, M., Cassidy, M.J., 2002. Freeway traffic oscillations: observations and predictions. In: Taylor, M.A.P. (Ed.), 15th International Symposium on Transportation and Traffic Theory. Pergamon, Amsterdam, pp. 653–674.
- 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. *Applied Mathematical Modelling* 11, 447–457.
- Papageorgiou, M., 1998. Some remarks on macroscopic flow modeling. *Transportation Research Part A* 32 (5), 323–329.
- Papageorgiou, M., Kotsialos, A., 2002. Freeway Ramp metering: an overview. *IEEE Transactions on Intelligent Transportation Systems* 3 (4), 271–281.
- 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.
- Siebel, F., Mauser, W., Moutari, S., Rascle, M., 2009. Balanced vehicular traffic at bottleneck. *Mathematical and Computer Modelling* 49 (3–4), 689–702.
- Smoller, J., 1983. *Shock Waves and Reaction-Diffusion Equations*. Springer-Verlag, New York.
- Treiber, M., Kesting, A., Helbing, D., 2006. Understanding widely scattered traffic flows, the capacity drop, and platoons as effects of variance driven time gaps. *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics* 74 (2), 1–10.
- 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.
- Zhang, L., Levinson, D., 2004. Some properties of flows at freeway bottlenecks. *Transportation Research Record* 1883, 122–131.