Title
Critique of a Freeway On-Ramp Metering Scheme and Broader Related Issues

Author
Cassidy, Michael J.

Publication Date
2002-05-01

Institute of Transportation Studies
University of California at Berkeley

Critique of a Freeway On-Ramp Metering Scheme and Broader Related Issues

Michael J. Cassidy

RESEARCH REPORT
UCB-ITS-RR-2002-4

May 2002
ISSN 0192 4095

Critique of a Freeway On-Ramp Metering Scheme and Broader Related Issues

By
Michael J. Cassidy
University of California
Department of Civil and Environmental Engineering
and the Institute of Transportation Studies
416C McLaughlin Hall, Berkeley, CA 94720
(510) 642-7702
cassidy@ce.berkeley.edu

Abstract

Offered here is a critique of a simple scheme recently proposed for metering freeway on-ramps. An earlier report of this scheme's potential for reducing commuter delay is shown to be exaggerated. The discussion makes clear that to reduce delay, metering should increase the rates at which commuters exit the freeway. The scheme critiqued here, as well as other well-known metering algorithms, are shown to have deficiencies in this, particularly when the freeway is plagued by a diverge bottleneck with a congested off-ramp. Other more effective schemes for reducing the delay caused by these types of bottlenecks are described. Suitable ways of field-testing freeway traffic management strategies are discussed as well.

1. Introduction

The following is a critique of a freeway on-ramp metering scheme recently proposed by a research group at the University of California (*Jia, et.al., 2000*). Previous predictions (in *Jia*) that the scheme dramatically reduces commuter delay are shown to stem from analysis that was seriously flawed. Along with a brief description of the scheme itself, the following section of the critique includes a summary of *Jia's* earlier analysis, along with its mistakes. Further discussions of the mistakes, and their consequences, come in sections 4 through 7. These discussions are preceded by some needed background information in section 3.

The above will bring to light certain facts. These facts will explain why significant delay savings come by metering in ways that increase freeway outflows. They will also show why the scheme critiqued here (and others with similar logic) can result in diminished outflows, and thus higher delay, on freeways with diverge bottlenecks.

Ironically, the freeway site used in *Jia's* analysis is evidently plagued by a bottleneck of this type. *Jia's* predictions that the metering scheme would save delay come, in part, from having over-looked this bottleneck. In reality, the scheme would make commute-time travel on this freeway (and its on-ramps) even worse than what presently occurs.

On a more positive note, there are other kinds of schemes that are well suited to diverge bottlenecks and some are described in the conclusions. Methods of field-testing these and other traffic management strategies are discussed there as well.

2. The Scheme and Earlier Analysis

Jia, et. al. predicts impacts of deploying the metering scheme on the 6-mile freeway stretch shown in Fig. 1. The scheme would reportedly eliminate freeway queues (on this or any freeway) using a simple logic. Namely, each on-ramp is separately controlled to keep flows entering its downstream freeway link from exceeding a specified threshold.[1] For *Jia's* analysis, each link's threshold was taken to be 3 percent below the highest flow sampled there during the rush (when there was a queue somewhere on the freeway stretch). The sampling intervals were 5-mins long.

These highest measured flows were said to be the "effective capacities" of the freeway links. Vehicles reportedly traverse a link at free flow speeds when the flow on this link equals its effective capacity; i.e., delays were assumed to occur only in queues.

Vehicle speeds and flows measured by the site's thirteen mainline loop (ML) detectors were used to estimate the vehicle hours traveled (VHT) on the freeway stretch during weekday

---

[1] Although the scheme was given the name *ideal metering*, its logic is no different from that of the so-called demand-capacity algorithm proposed decades earlier (*Wattleworth, 1964*).

mornings. Each day's estimate was compared with the VHT for the delay-free travel freeway traffic would have supposedly enjoyed under the proposed scheme. To keep freeway flows below effective capacities, vehicles were stored on the on-ramps as needed. The resulting ramp delays were included in the predictions and each on-ramp was assumed to have space for storing 30 queued vehicles. Notably, changes in off-ramp outflows that would have resulted from the metering scheme were not estimated in *Jia*.

The predictions resulting from the above analysis are tantalizing ones. *Jia* predicts the period queues persist at the site would drop from more than 2 hours each morning to about 1 hour. Moreover, commuter delay at the site would reportedly diminish by as much as 70 percent. Even further delay savings were predicted if commuters were to alter their travel behavior (e.g. their routes, their schedules or their modes) to keep on-ramp queues created by the scheme from growing long and disrupting surface street traffic. The number of such diversions required for this was reported to be small (i.e., supposedly fewer than 300 commuters would need alter their behavior each morning).

The present critique demonstrates that the above analysis was flawed in a number of fundamental ways and that these flaws invalidate the predictions just noted. This demonstration will be based largely on data provided in *Jia, et. al.* These data describe the morning of June 1, 1998 and they are presumably typical of weekday mornings at the site. The data reveal the following mistakes.

First, *Jia's* claim that the metering scheme would eliminate the freeway's queue is not credible. The data show this queue was triggered by a congested off-ramp, with the problematic ramp being the one located between MLs 10 and 11 (as shown in Fig. 1). Contrary to *Jia's* analysis, the scheme would actually exacerbate freeway queueing and delay by restricting vehicles not destined for the problematic ramp to favor commuters who were directed there. This "first" mistake, in itself, invalidates all predictions in *Jia*.

Second, the queue from the diverge bottleneck grew and eventually propagated beyond ML 1 (see again Fig. 1). Consequently, pent-up demands persisted upstream of the freeway stretch for a full 50 mins. Although these pent-up demands contributed substantially to delay and the duration of the rush, they were ignored. Instead, the lower freeway flows constrained by the queue were used as inputs to *Jia's* analysis. This further contributed to unrealistically small predictions of delay and rush duration.

Third, on-ramp queues predicted to form under the scheme were assumed ultimately to discharge at unrealistically high rates. Ramp discharge flows greater than 2,000 vph were presumed to occur without problem, even when these were inflows to a short weaving section.

Fourth, the amount of commuter diversion needed to manage on-ramp queues was underestimated. This stemmed from an unrealistically short prediction of the rush caused by the previous three mistakes.

Fifth, the assumption of delay-free travel on freeway links flowing close to capacity is contrary to the data. Each of the thirteen occupancy-flow scatter-plots presented in *Jia* indicate that, in uncongested traffic, vehicle speeds diminish (and delays thus arise) as flows approach capacity.

These five mistakes are verified and further discussed in later sections. What follows next is background information to set the stage for this.

3. Background

Issues concerning metering and delay will next be discussed in the context of a very simple and hypothetical queueing system. This system will initially be viewed as one that serves people exiting a sports stadium. But its geometry is similar to that of freeways and important analogies to freeway systems will be made clear. Moreover, the discussion here will eventually shift to one that focuses entirely on freeways.

With the above in mind, Fig. 2(a) illustrates two links that merge to a common stream close to the exit of our sports stadium. Customers in the common stream are served in a first-in, first-out fashion. The dashed lines labeled "off-ramp" in the figure can be ignored for now.

During the rush (at the conclusion of a sporting event), customer arrival rates to Links *A* and *B* exceed μ, the capacity of the stadium exit. Left unattended, the short stretch of the common stream in advance of the exit soon becomes completely queued, such that the flow there is μ. The combined rate that customers from Links *A* and *B* advance to the common stream is then μ as well and we can assume queues propagate backward on both of these links.

The resulting customer delay in this system is given by a queueing diagram, like the one in Fig. 2(b). The curve labeled V displays the cumulative number of customers from both links that would like to have exited the stadium by time t. Since the V(t) are connected with a smooth curve, its slopes are the rates customers would have exited the stadium in the absence of delay. This curve could have been constructed by 1) measuring, for each link, all customer arrival times at some location(s) upstream of all queues; 2) increasing each arrival time by that customer's free flow (or undelayed) trip time from the measurement location to the stadium exit; and 3) sorting the array of these *virtual* arrival times in ascending order and then plotting them cumulatively. Notably, the number of these customers to have actually exited the stadium by t is displayed by the curve labeled D.

It follows that the shaded area in Fig. 2(b) is the total delay collectively incurred by all customers in the system. In this example, the queue(s) and delay persisted from time t_1 to t_2.

The above discussion makes clear that the queueing diagram is a means of displaying data that might have actually been measured. The only conjecture comes from having drawn the D-curve with a fixed slope μ during the rush. This assumes that, if not impeded by some queue from further downstream, customers pass through the stadium exit at its capacity whenever queues form in the system. This maximum rate is independent of the number that are queued and of the proportions of exiting customers supplied by each link.

These assumptions seem reasonable when customers are stadium patrons. More importantly, there is considerable empirical evidence freeway vehicles really do discharge from a merge bottleneck at a nearly constant maximum rate (*Cassidy and Bertini, 1999; Bertini and Cassidy, 2001; Mauch and Cassidy, 2002*).

Suppose a stadium employee acts as a meter by further restricting the rates customers from Link *B* merge into the common stream to favor those from *A*; (see Fig. 2(a)). Travel speeds and flows then increase on Link *A*. Since they now exit the stadium at higher rates, customers from *A* incur lower delays. But this metering does not change the total delay. Fig. 2(b) plainly shows total delay is conserved if the V- and D-curves are unaltered. At best, the scheme would have merely re-distributed delay, with more now going to customers from *B*. (We assume for now that metering does not affect the demand for travel displayed by the V-curve).

Now suppose Link *A* and the common stream is really a freeway stretch; Link *B* is an on-ramp; and μ is the capacity of the freeway link downstream of the merge.[2] Suppose too Link *B* (and any sources feeding Link *A*) are metered using the logic of *Jia's* scheme. During the rush, vehicles now depart the merge at a rate of 0.97μ. Link *A* is completely unqueued, with traffic on this link enjoying higher speeds, as compared with the unmetered case. Yet total delay in the system, and the duration of the rush, both increase.

This is made clear in Fig. 2(b). Since the maximum slope of the D-curve drops to 0.97μ, the shaded area grows. The end of the rush, formerly at time t_2, is postponed and the start of the rush could now occur even earlier than t_1.

The above illustrates that maximizing the cumulative outflows from a system (or more specifically, the integral of these cumulative outflows) is key in holding down delay and should be a primary objective in metering. This is true for the simple system in Fig. 2(a), as well as for freeway systems that include many on- and off-ramps.

---

[2] We consider for the time being a highly idealized case whereby the common stream is a freeway link with only a single travel lane.

Many engineers, unaware of the above, erroneously use higher vehicle speeds and flows on freeway links within a system as evidence a metering scheme has diminished delay (e.g. MnDOT, 2001). The potential flaw in this reasoning was evident in the previous analogy. The metering described above promoted higher speeds and flows on Link *A* without saving delay. *Jia's* (overly restrictive) scheme even increased delay!

On a related note, some engineers have been quick to assume that metering increases the capacities of freeway bottlenecks formed at merges (see, for example, *Papageorgiou and Kotsialos, 2000*). If true, this would mean metering could reduce commuter delay by increasing outflows from these bottlenecks. Claims of this kind, however, have yet to be properly supported by data.[3]

We continue to assume the simple system in Fig. 2(a) is metered using *Jia's* logic and we next add some links to this system. Suppose a diverge link is located along Link *A*. In this case, metering *B* to increase *A*'s flow could mean higher outflows from this "off-ramp" because vehicles might now get to it with less impedance. This would reduce delay in the system.

But suppose that the common stream also has an off-ramp just beyond the merge and that this off-ramp's capacity is μ_o, with $\mu_o < \mu$. Suppose too that the proportion of vehicles entering the common stream that are bound for this off-ramp is $\_{}_o$. We will assume $\_{}_o$ is independent of time. Then the flow that approaches the downstream diverge cannot exceed $\mu_o/\_{}_o$; under this (maximum) flow, vehicles use the off-ramp at its capacity, μ_o.

The diverge becomes the system's bottleneck if the flow directed to its off-ramp exceeds the ramp's capacity; i.e., if $\_{}_o\_0.97\mu > \mu_o$. The off-ramp would be unable to absorb this flow. A queue would form in the common stream and propagate backward past the merge and onto Link *A*. This would mean the flow departing the merge is $\mu_o/\_{}_o$ and this flow would be lower than 0.97μ. Delay in the system would therefore further increase; Fig. 2(b) shows that the shaded area between curves would grow.

The above illustration makes clear that discharge flows through a diverge bottleneck are sensitive to $\_{}_o$. Suppose the proportion of vehicles bound for the downstream off-ramp and originating from Link *A*, $\_{}_{oA}$, equaled the analogous proportion from *B*, $\_{}_{oB}$. Here metering cannot increase flows through the diverge because changing the metering rate (for *B*) would not change $\_{}_o$.

---

[3] There is evidence that merge capacities diminish when these locations become *active* bottlenecks; i.e., when the queues formed by these merges are not affected by traffic conditions further downstream (e.g. *Cassidy and Bertini, 1999*). Whether metering can postpone such capacity reductions is a question requiring further empirical study. Past studies that have not controlled for the effects of downstream queues (i.e., studies not involving *active* bottlenecks) shed no light on this issue.

However, metering can affect a diverge bottleneck (and the delay it creates) if the scheme can affect the _$_{o}$. This can occur for _$_{oA}$ ≠ _$_{oB}$. *Jia's* metering scheme would, for example, promote lower flows through the diverge bottleneck if _$_{oA}$ > _$_{oB}$. This is because metering *B* would foster higher _$_{o}$ in the common stream.

4. The Diverge Bottleneck

Facts presented in this section will show i) the freeway site used in *Jia's* analysis (displayed in Fig. 1) is plagued by a diverge bottleneck near its downstream end; ii) *Jia's* metering scheme would promote higher _$_{o}$ in the freeway queue that forms upstream of this bottleneck and thereby diminish freeway outflows and increase delay; and iii) items i and ii above were notably absent from *Jia's* analysis.

Figs 3(a)-(e) are plots of occupancies (dimensionless measures of density) and flows jointly sampled across all travel lanes by some of the site's ML detectors. These were sampled over 5-min intervals (on June 1, 1998). The occupancies at each ML were evidently averaged across all lanes there.

Fig. 3(a) shows some of the average vehicle speeds dropped a little below those of free flow traffic.[4] But observations of low speeds and high occupancies typical of queues are only visible at locations upstream of ML 11, as evident in Figs. 3(b)-(e).

These five scatter-plots collectively reveal a bottleneck activated between MLs 10 and 11; i.e., unqueued traffic persisted at ML 11 while queueing arose upstream. Fig. 1 displays what is evidently the freeway's only geometric inhomogeneity at this bottleneck location: an off-ramp. This is the first clue freeway queueing was triggered by exiting vehicles at a congested off-ramp.

Additional such evidence is revealed in Figs. 4(a)-(e). These display (5-min) measured vehicle speeds apparently averaged across all lanes at the ML detectors. Fig. 4(a) shows the drop in speeds at ML 11 was indeed modest during the rush; most of these speeds were above 55 mph.[5] Only the ML detectors further upstream measured the more substantial speed reductions, as evident in Figs 4(b)-(e).

Three of these latter figures are annotated with times of day. Those shown in parenthesis indicate when speeds at the ML dropped below 60 mph at the outset of prolonged speed reductions. The start times of these trends occurred later at each upstream ML, with the earliest

---

[4] The average vehicle speed corresponding to any one of these observations is proportional to the slope of the chord passing through the plot's origin and the data point itself.

[5] Measured speeds at ML 12 appeared to be no higher than those at ML 11. This rules-out the possibility freeway queueing was triggered by inflows from on-ramp (OR) 7.

of these occurring some minutes before 7:00 am. These times mark the arrivals of a queue as it propagated against the flow of traffic, as in *Lighthill and Whitham (1955*).

Near the bottleneck at MLs 10 and 9, the queue initially caused small reductions in speeds and these reductions exhibited a downward trend over time. There were brief periods when average speeds fell to 40 mph or lower. These patterns of speed reduction are typical when the proportion of vehicles directed to a congested off-ramp, _$_o$, changes with time (*Muñoz and Daganzo, 2000*).

That speed reductions at MLs 8 and 4 were more pronounced and immediate is likely the result of flow restrictions there caused by inflowing traffic at downstream on-ramps (see *Cassidy and Mauch, 2001*). These restrictions may have been exacerbated by commuters who were bound for the congested off-ramp and who decelerated before squeezing into the shoulder lane's queue (see *Muñoz and Daganzo, 2000*).

Figs. 4(a)-(e) also indicate ML 10 was the only location where average vehicle speeds tended to remain below 60 mph after the rush had ended. (Annotated times of day in Figs. 4(c)-(e) reveal the tail of the queue eventually propagated forward in traffic, bringing an end to the rush some minutes after 9:00 am). It thus appears the congested off-ramp continued to impart small traffic disruptions nearby at ML 10, even after demands for freeway travel had subsided.

Fig. 5 shows the effects of the diverge bottleneck on freeway discharge rates downstream. It displays (5-min) flows at ML 11. The data points connected by the dotted line are the measured values. By about 6:45 am, these no longer tended to increase with time. This indicates the freeway queue started to form at 6:45 am and this is consistent with the onset of speed reductions previously shown in Fig. 4(c).

Fig. 5 also indicates that measured flows tended to diminish gradually from about 7:30 to 9:15 am. This observation confirms what was already revealed in Figs. 4(b) and (c): traffic disruptions caused by the diverge bottleneck persisted, and even intensified, over the rush as _$_o$ changed with time.

Notably, *Jia's* metering scheme would not mitigate the congestion created by this bottleneck. Under the scheme, each ramp's meter is individually tasked with moderating inflows to the freeway link downstream; the burden of this is not shared with neighboring ramps in some coordinated fashion. It follows that on-ramps far upstream of the diverge bottleneck will not necessarily be metered restrictively.[6] Since these are probably the entry points for most

---

[6] This is even evident in the data provided in *Jia, et.al.* Queues at on-ramps upstream of the freeway stretch, collectively designated as on-ramp (OR) 1, display predicted queues much shorter in length than those predicted for most on-ramps within the freeway stretch itself. Figs. 7 and 8(a), shown later in the critique, exemplify this.

commuters destined for the congested off-ramp (as California commute distances tend to be long), their arrivals to the exit queue would be unmitigated. The scheme would promote higher _$_o$ on the freeway stretch by metering commuters not bound for the congested off-ramp to favor those who are directed there. This would create further reductions in outflow at ML 11.

No such reductions are predicted in *Jia*, however. The solid line in Fig. 5 displays the outflows at ML 11 predicted under metering. These seldom fall below their measured counterparts. The reverse is very often the case, particularly from 7:30 to 8:00 am.

Fig. 5 does show a predicted decline in flows after about 8:00 am. It turns out these reductions reflect the mistaken assumption the metering scheme would, by about this time, bring an end to all queueing; i.e., these lower predicted flows are the demands projected after the rush.

This prediction of a shortened rush is not realistic. Additional mistakes that contributed to this prediction are described in the next two sections.

5. Pent-Up Demands

Evidence in this section will show the freeway queue eventually propagated beyond the upstream-most detectors at ML 1. In its attempt to eliminate this freeway queue, the metering scheme would thus have been confronted with pent-up demands from upstream. These added vehicles, and the delay created by having to control them in some fashion, are shown to have been ignored in *Jia's* analysis of the metering scheme.

Figs. 6(a)-(c) present traffic data pertaining to ML 1. The first of these figures shows the measured flows there (dotted line) and those used as input in *Jia's* analysis of the scheme (solid line). Fig. 6(b) displays the measured average vehicle speeds. Finally, Fig 6(c) shows the measured occupancies (dotted line) and those predicted under the metering scheme (solid line) averaged across all lanes at ML 1.

These figures indicate that, by 7:40 am, measured flows and speeds fell sharply while the measured occupancies increased. These conditions persisted until 8:30 am, at which time speeds and occupancies returned to their earlier states.

These are indisputable indications of the queue's presence at ML 1. Consequently, the flows measured there during the 50-min period (from 7:40 am) were not demands for freeway travel, but lower rates constrained by the queue.

Fig. 6(d) is a hypothetical queueing diagram to help clarify issues here. Its V- and D-curves display the desired and actual arrivals to ML 1 by time t. Their sudden displacement at 7:40 am marks the arrival of the queue at this ML; the D-curve's slope between 7:40 and 8:30 am is a constrained flow; the shaded area is the total delay to vehicles that passed ML 1; and the count dimensioned as N_Q is the number of vehicles to have had their arrival times there deferred. These deferments persisted until 8:30 am, at which time the tail of the freeway queue moved downstream of ML 1.

Jia, however, erroneously reports that the metering scheme would bring an end to all queues by 8:00 am. This is partly because a D-curve was, in effect, mistakenly taken for the demand at ML 1. The V-curve (which does describe this demand) was not estimated. The N_Q deferred vehicles, and their delays, were thus ignored in *Jia*.

The evidence of this mistake is clear from the data. Fig. 6(a) shows that, from 7:40 to 8:30 am, the measured and predicted flows at ML 1 were identical. This means that, under the metering scheme, the N_Q deferred vehicles were still not allowed to pass ML 1 undelayed. Yet nor were they stored upstream. This is clear from Fig. 7. It shows the queues that metering would reportedly have generated at all locations upstream of ML 1. Notably, all such queues were predicted to disappear by about 7:40 am.

6. On-Ramp Discharge Rates and Queue Lengths

Jia's analysis included some unrealistically high estimates of on-ramp discharge flows. To demonstrate, Fig. 8(a) displays the queue lengths at on-ramp (OR) 2 that would reportedly result from the metering scheme. The queue there reaches its maximum shortly before 8:00 am and it is supposedly served in only 10 mins.

The predicted discharge flows from this ramp are shown with the solid line in Fig. 8(b). These reach as high as 180 vehicles per 5 mins, a rate of nearly 2,200 vph. This high rate, summed with the freeway flow upstream, was apparently less than the downstream link's effective capacity. But an on-ramp flow this high would still likely produce a tangled mess; i.e., the vehicle merging and lane-changing maneuvers that would accompany this inflow could trigger significant queueing problems (*Moskowitz and Newman, 1963; Cassidy and Rudjanakanoknad, 2001*). This is especially true in that OR 2 resides at the upstream end of a short weaving section where merging and diverging traffic interact (see Fig. 1).

Moreover, such high on-ramp rates might flood a downstream off-ramp with flow and cause problems there. These rates may even exceed the capacity of the on-ramp itself.

J*ia* assumes other nearby on-ramps would discharge their queues at very high rates as well. This, like the previously cited mistakes, contributed to an unrealistically short prediction of the rush. This means the on-ramp queues that would be generated by the metering scheme were underestimated. So were the changes in commuter travel behavior (e.g. the route diversions) needed to hold these queues to manageable lengths.

7. The Relation Between Vehicle Speed and Flow

The discussion in this short section counters *Jia's* contention that free flow vehicle speeds arise on freeway links flowing at their effective capacities. The contention is contrary to data measured on the freeway stretch and this is exemplified in Fig. 9. Shown in this figure are the occupancy-flow data sampled at ML 6. The slope of the chord passing through the origin and those observations of low occupancy and flow can be taken as proportional to the free flow speed. The second chord in Fig. 9 passes through the unqueued data point with flow equal to the link's effective capacity. It has a lower slope; one proportional to the lower speed of vehicles in unqueued, near-capacity flows. Each of the other twelve scatter-plots presented in *Jia, et.al.* are similar in this regard.

By definition, commuter delay arises by traveling below free flow speeds. The claim that near-capacity flows occur without creating delay is therefore false. (Admittedly, the consequences of ignoring speed reductions in unqueued traffic may be small since this kind of delay can be much lower than the delay in a queue).

8. Conclusions

Jia notes it is sometimes possible to save delay if metering alters travel behavior; e.g. if it motivates some commuters to change their routes. What it fails to point-out, however, is that those who divert likely incur more delay than what they had encountered prior to the metering scheme's deployment. They had, after all, previously viewed their new routes as inferior. By the same token, *Jia* does not mention that commuters who divert from a freeway might increase delays for other surface street traffic. (The traffic subjected to these effects could include city buses and the many people they may be carrying). So evidence of diversion does not necessarily mean delay has been saved. Unfortunately, the full effects of diversion can be difficult to evaluate since these effects are often felt at many neighboring surface streets.

Delay is saved, on the other hand, if metering keeps a freeway queue from propagating past an off-ramp and starving it of flow. This was pointed-out using the hypothetical freeway system in section 3. However, changes in off-ramp flows were not estimated in *Jia's* analysis of the metering scheme.

Without fortuitous changes in travel behavior, and absent higher off-ramp flows, metering schemes like *Jia's* can, at best, only transfer freeway delay to on-ramps and surface streets. This would be counter-productive since the freeway has more space for storing delayed vehicles and queue storage space is a commodity that should not be squandered. Things are made worse if the scheme is overly restrictive, such that outflows are diminished (e.g. by 3 percent).

For the freeway site in Fig. 1, moreover, *Jia's* scheme would create even further delay by failing to restrict enough commuters who are bound for the congested off-ramp. Notably, there are other metering schemes that can, in similar fashion, add to problems at freeway diverge bottlenecks. These include well-known metering algorithms such as SWARM (*NET, 1996*), ALINEA and METALINE (*Papageorgiou and Kotsialos, 2000*). Much as in *Jia's* scheme, these can promote higher $\_{o}$ upstream of the diverge bottleneck. An illustration of this using the ALINEA algorithm is offered in another report (*Cassidy, 2002*).

Even if not subjected to ill-suited schemes like those above, diverge bottlenecks can create huge delay (*Muñoz and Daganzo, 2000; Cassidy, et .al., 2000*). Fortunately, there are traffic management strategies that can be very effective in dealing with this type of bottleneck and thus save much delay.

One might, for example, coordinate the metering rates at multiple on-ramps in ways that reduce $\_{o}$. To this end, on-ramps serving higher numbers destined for the problematic off-ramp can be metered more restrictively than others. In the future, it may even be possible to separate vehicles bound for this off-ramp and meter them differently from other on-ramp traffic (*Daganzo, et. al., 2001*).

Moreover, traffic management strategies suitable for a diverge bottleneck are not limited to on-ramp metering. In many cases, the simplest solution for a bottleneck of this type would be to increase the rate vehicles are served by the off-ramp itself. This would commonly entail treating bottlenecks on nearby surface streets, since off-ramp queues often emanate from these.

The above reference (by *Daganzo, et. al.*) includes discussion on a number of different traffic management strategies for diverge bottlenecks. The reference also describes methods of mitigating commuter delay created by other sources.

Suffice to say, no single metering algorithm can suitably address all conditions that arise on different freeways. So a metering scheme, or any traffic management strategy, should be designed only after the freeway of interest has been carefully examined and its sources of delay diagnosed.

Such diagnoses will require data and in this regard, Jia and co-authors deserve much credit. Their work to make freeway traffic data available on the world wide web (at http://transacct.eecs.berkeley.edu) and via a PeMS database is of great value.

These data can also be used when field-testing a scheme's effectiveness in reducing delay. As per previous discussion, tests of this kind need not focus on vehicle speeds and flows on freeway links within a system. They should assess instead the freeway outflows the scheme promotes.

With this in mind, counts can be made of the vehicles exiting at each i^{th} egress point or "destination" along the freeway site. (The freeway stretch shown in Fig. 1, for example, has five such destinations: four off-ramps and the site's downstream boundary at ML 13). These counts should be recorded at specified time t spanning the rush and intervals of 5 mins or so will generally suffice. The counts for all destinations can then be summed together at each of these t, $D(t) = \sum_i D_i(t)$, and the cumulative curve of these summed D can be plotted over time.

One can construct such a cumulative curve resulting from a scheme's deployment and determine the area under this curve bounded by the start and end times of the rush. The same kind of curve can be measured prior to the scheme's deployment and the area determined for the comparable time period. The difference in these areas is an indicator of the scheme's performance (*Lin and Cao, 1997*).

Acknowledgements

The author thanks Z. Jia and A. Skabardonis for providing many of the figures included in this critique. Thanks are also due M. Mauch and S. Ahn for their assistance in preparing these figures for their inclusion here.

References

Bertini, R.L. and Cassidy, M.J. (2002) Some observed queue discharge features at a freeway bottleneck downstream of a merge. *In press, Transpn. Res. A.*

Cassidy, M.J. and Bertini, R.L. (1999) Some traffic features at freeway bottlenecks. *Transpn. Res. 33B,* 25-42.

Cassidy, M.J., Anani, S.B. and Haigwood, J.M. (2000) Study of freeway traffic near an off-ramp. PATH Working Paper UCB-ITS-PWP-2000-10, California PATH Program, University of California, Berkeley. *In press, Transpn. Res. A.*

Cassidy, M.J. and Mauch, M (2001) An observed traffic pattern in long freeway queues. *Transpn. Res. 35A*, 143-156.

Cassidy, M.J. and Rudjananakanoknad, J. (2001) Study of traffic at a freeway merge and roles for ramp metering. DRAFT Final Report, California PATH Program, University of California, Berkeley.

Cassidy, M.J. (2002) Freeway on-ramp metering and the diverge bottleneck. Article submitted to *Access Magazine*, University of California Transportation Center, Berkeley.

Daganzo, C.F., Laval, J. and Muñoz, J.C. (2001) Some ideas for freeway congestion mitigation with advanced technologies. DRAFT Final Report, California PATH Program, University of California, Berkeley.

Jia, Z., Varaia, P., Chen, C. Petty, K. and Skabardonis, A. (2000) Congestion, excess demand and effective capacity in California freeways. Report prepared by the PeMS Development Group, University of California, Berkeley, *available at* http://transacct.eecs.berkeley.edu

Kotsialos, A. and Papageorgiou, M. (1999) Optimal coordinated and integrated motorway network traffic control. In: Ceder, A. (ed.), Transpn. and Traffic Theory, Elsevier, Amsterdam. pp. 621-643.

Lighthill, M.J. and Whitham, G.B. (1955) On kinematic waves I: flood movement in long rivers. II: a theory of traffic flow on long crowded roads. *Procs. Royal Soc. A229*, 281-345.

Lin, W.H. and Cao, J.M. (1997) A system optimal dynamic traffic assignment model with limited system control. Presented at the 76th Annual Meeting of the Transportation Research Board, Washington, D.C.

Mauch, M. and Cassidy, M.J. (2002) Freeway traffic oscillations: observations and predictions. *In press.* In: Taylor, M.A.P. (Ed.), Transpn. and Traffic Theory, Elsevier, Amsterdam.

Minnesota Department of Transportation, MnDOT (2001) Twin cities ramp evaluation. Executive Summary, Cambridge Systematics, Inc, *available on the world wide web at* http://www.dot.state.mn.us/rampmeterstudy

Moskowitz, K. and Newman, L. (1963) Notes on freeway capacity. *Highway Res. Rec. 27*, Highway Res. Board, 44-68.

Muñoz, J.C. and Daganzo, C.F. (2000) The bottleneck mechanism at a freeway diverge. *Accepted, Transpn. Res.*

NET (1996) System wide adaptive ramp metering algorithm – high level design. Final Report prepared by NET for Caltrans and FHWA.

Papageorgiou, M. and Kotsialos, A. (2000) Freeway ramp metering: an overview. *Procs. IEEE Intelligent Transpn. Systs.* Dearborne, MI, USA, pp. 228-238.

Wattleworth, J.A. (1964) System demand-capacity analysis on the inbound Gulf Freeway. Texas Transpn. Inst. Res. Rept. 24-8.

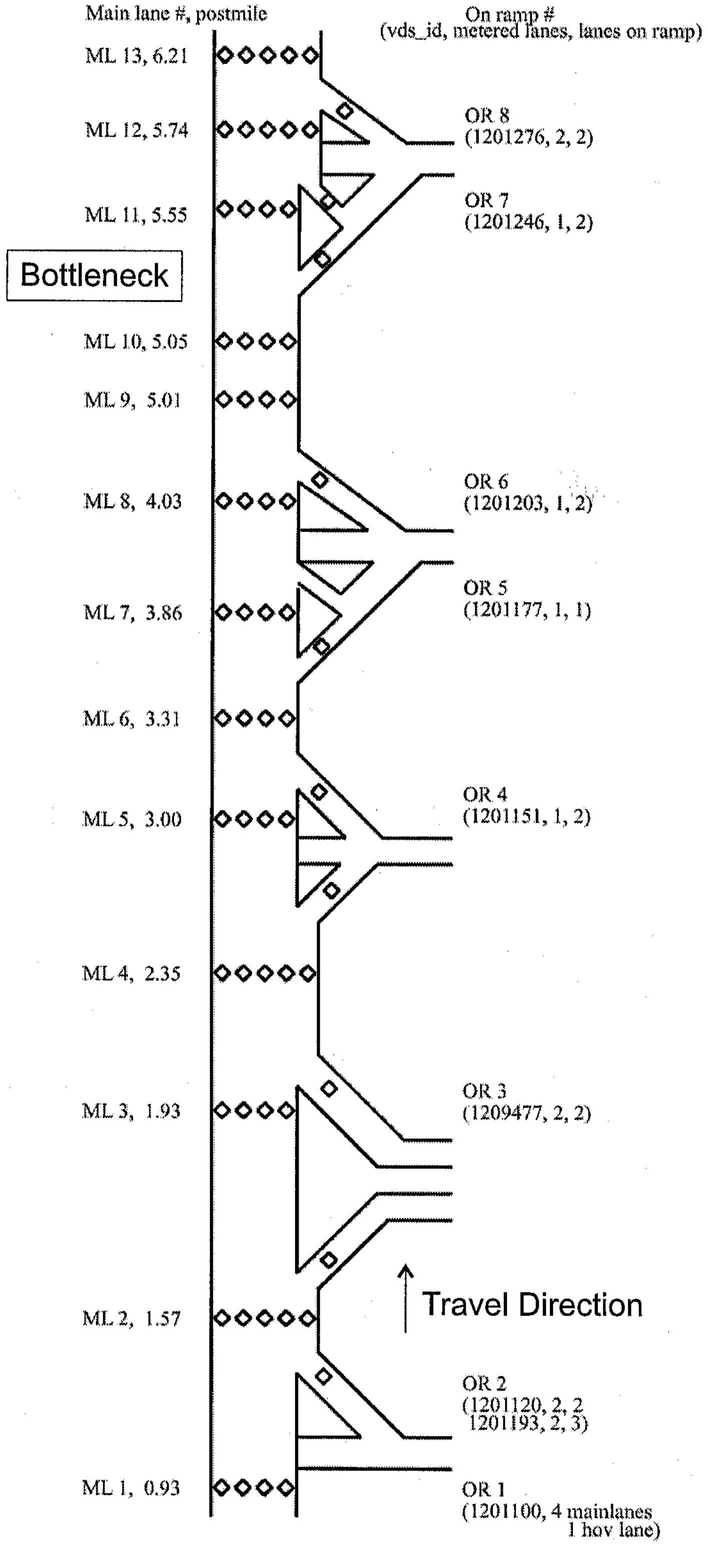


Figure 1

Study Site: Northbound Interstate 405, Orange County, California

(Source: Jia, et.al., 2000; not shown is an HOV lane adjacent to the median; annotations of travel direction and bottleneck were added.)

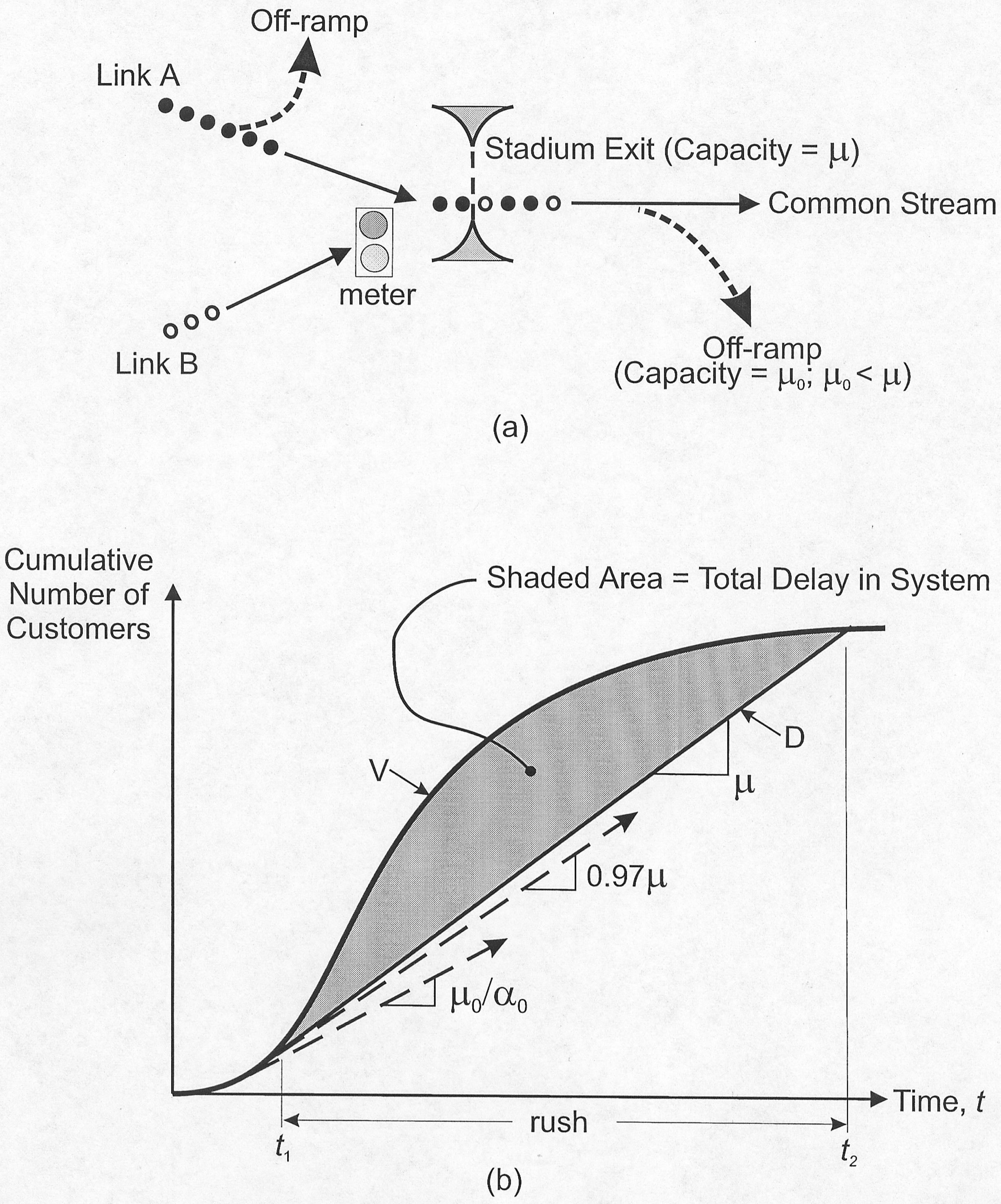
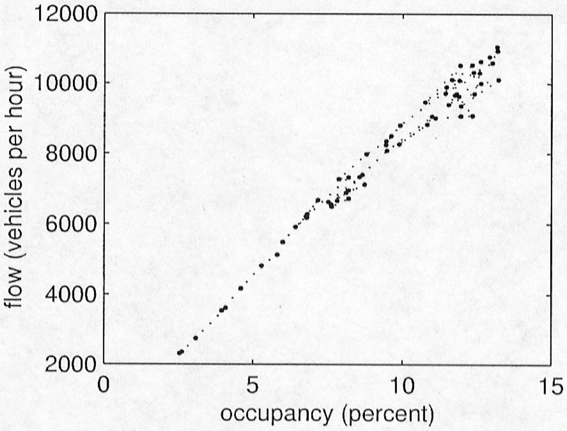
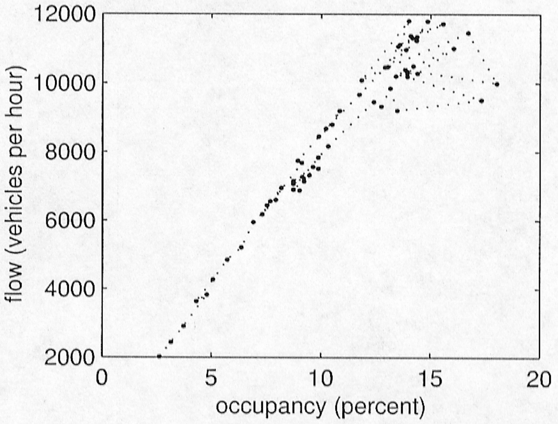


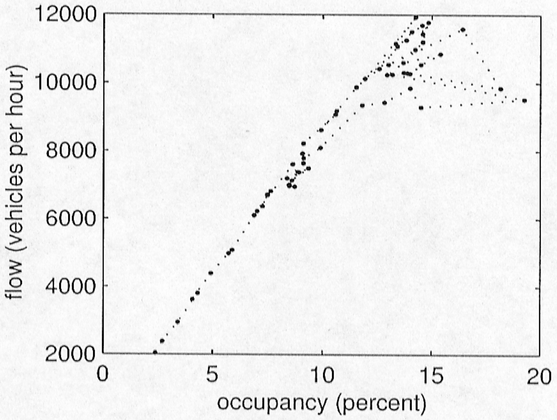
Figure 2
(a) Diagram of simple queueing system;
(b) Example queueing diagram for the system



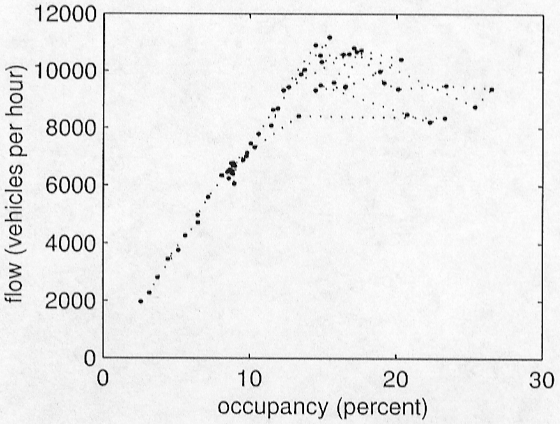
(a), ML11



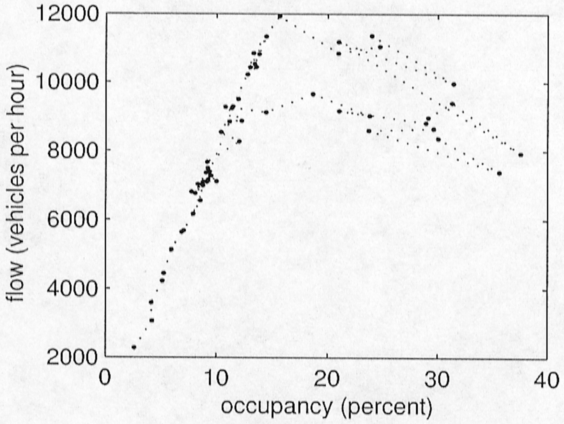
(b), ML10



(c), ML9

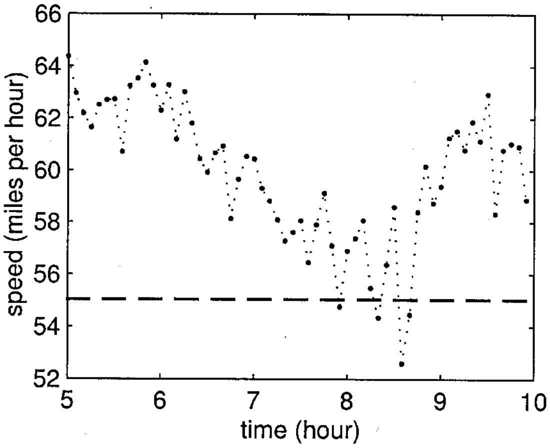


(d), ML8

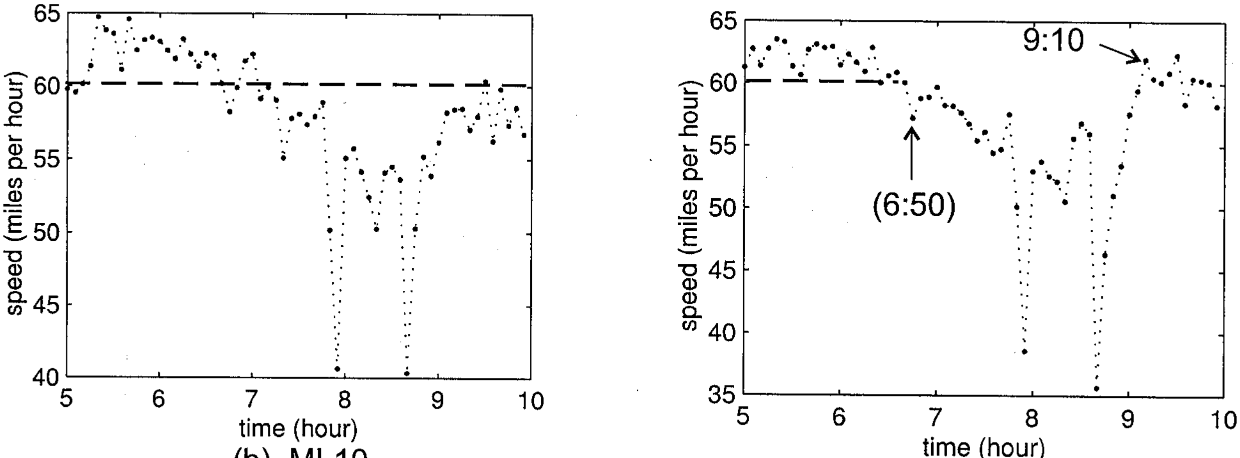


(e), ML4

Figure 3(a) - (e)
Occupancy - flow plots of freeway traffic;
measured across all lanes over 5-min sampling intervals
(source: Jia, et.al., 2000)

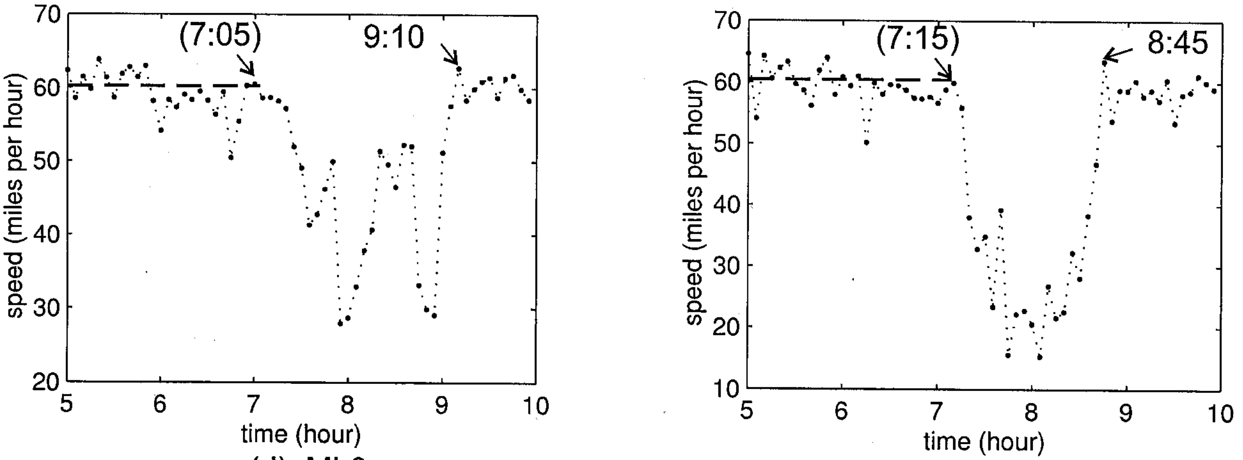


(a), ML11



(b), ML10

(c), ML9



(d), ML8

(e), ML4

Figure 4(a) - (e)
Time-series plots of average vehicle speeds;
averaged across all lanes over 5-min sampling intervals
(source: Jia, et.al., 2000; Dashed line segments and annotated times added)

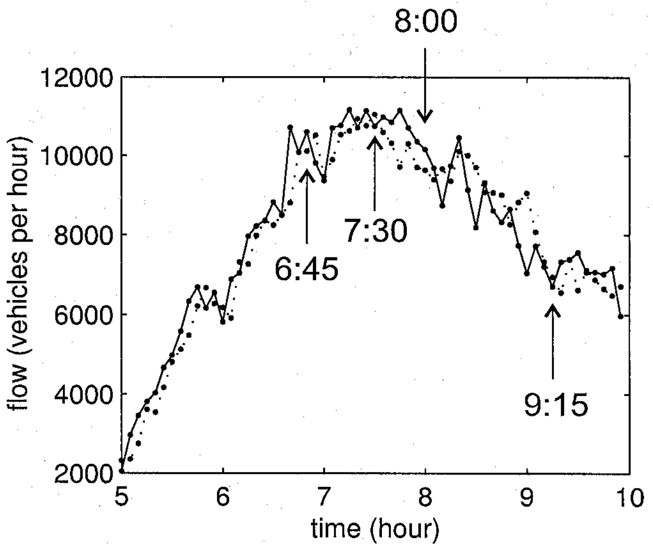


Figure 5
Time-series plots of flow at ML11;
across all lanes over 5-min Intervals;
dotted line displays measured flows, solid line shows predictions.
(source: Jia, et.al., 2000; annotated times added)

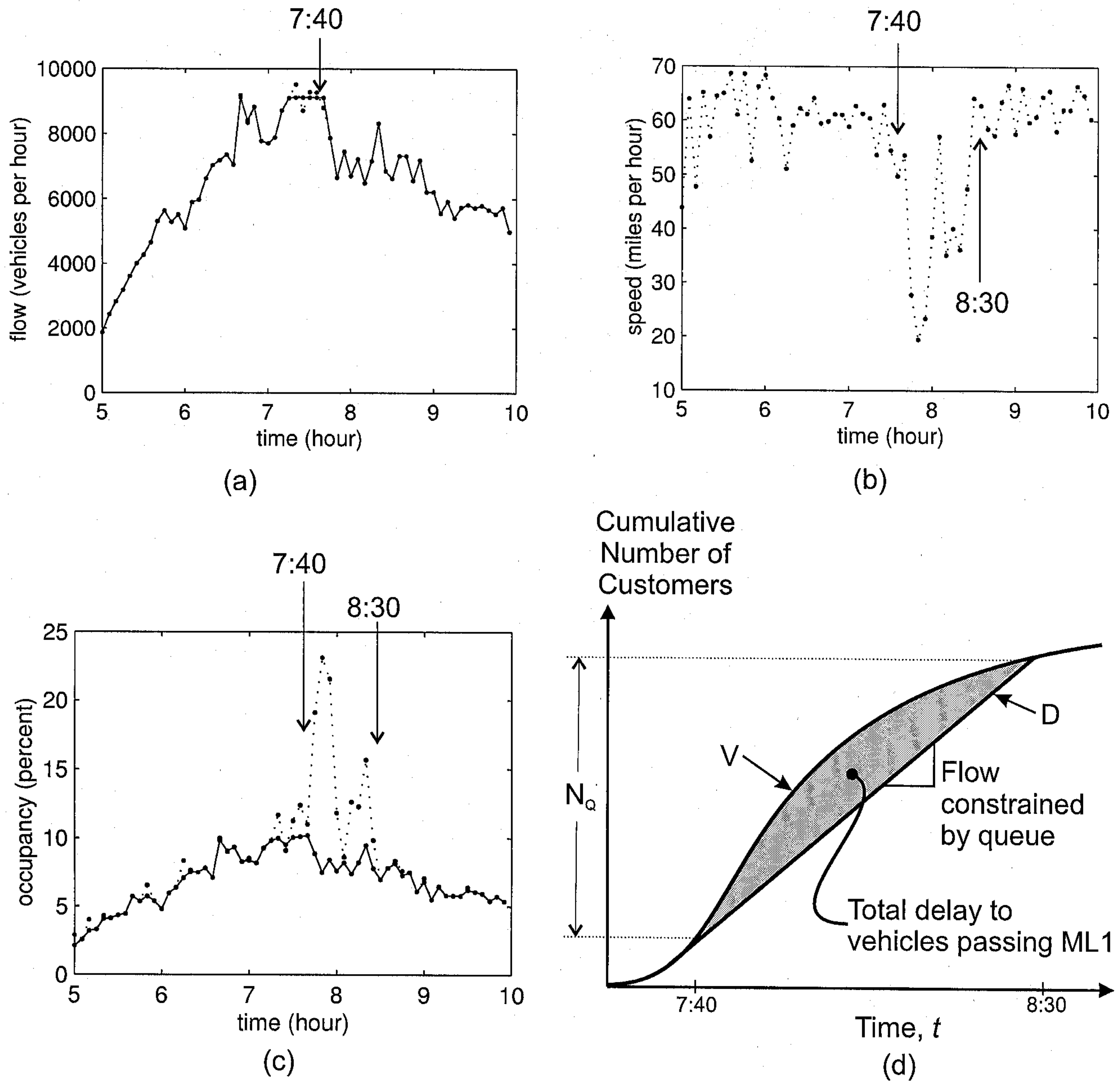


Figure 6
Traffic conditions at ML1

(a) Flows; dotted line displays (5-min) measured values; solid line shows input flows;
(b) Measured vehicle speeds; (5-min) averages over all lanes;
(c) Occupancies; dotted line displays (5-min) measurements averaged over all lanes; solid line shows predictions;
(d) Hypothetical queueing diagram
(source of (a)-(c): Jia, et.al., 2000; annotated times added)

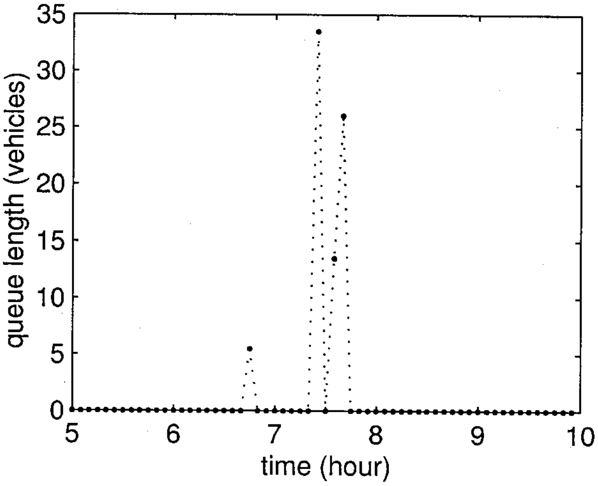
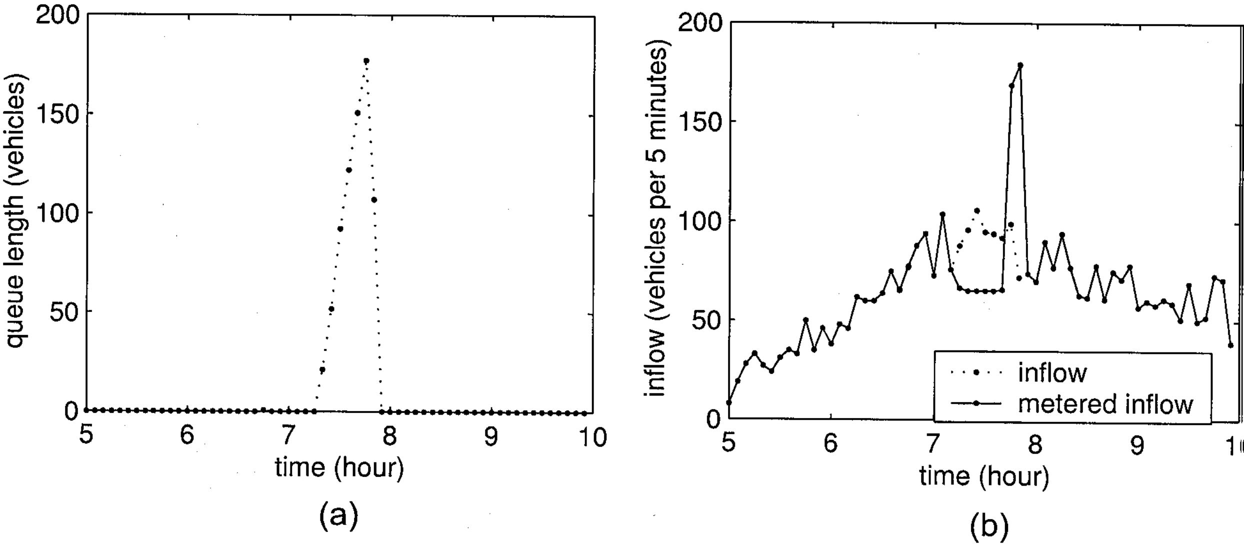


Figure 7
Predicted queue length at OR1
(source: Jia, et.al., 2000)



(a)

(b)

Figure 8
Traffic conditions at OR2
(a) predicted queue lengths,
(b) Measured and predicted flows
(source: Jia, et.al., 2000)

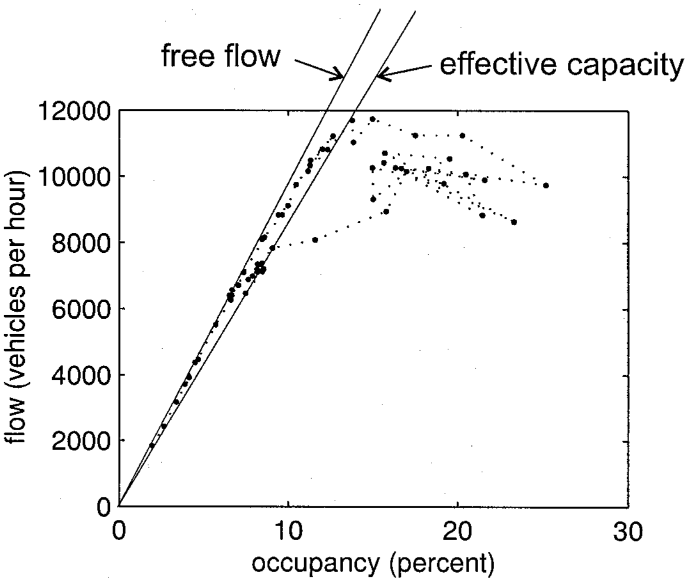


Figure 9
Occupancy-flow plot sampled at ML 6
(source: Jia, et.al., 2000; chords added)