The Pennsylvania State University

The Graduate School

Department of Civil and Environmental Engineering

MODELING UNCERTAINTY IN LARGE-SCALE URBAN TRAFFIC NETWORKS

A Dissertation in

Civil Engineering

by

Xueyu Gao

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2015

The dissertation of Xueyu Gao was reviewed and approved* by the following:

Vikash V. Gayah
Assistant Professor of Civil Engineering
Dissertation Advisor
Chair of Committee

Eric T. Donnell
Associate Professor of Civil Engineering

S. Ilgin Guler
Assistant Professor of Civil Engineering

Minghui Zhu
Dorothy Quiggle Assistant Professor of Electrical Engineering

William Burgos
Professor of Environmental Engineering and Professor in Charge of Graduate Programs

*Signatures are on file in the Graduate School

Abstract

Recent work has proposed using aggregate relationships between urban traffic variables—i.e., Macroscopic Fundamental Diagrams (MFDs)—to describe aggregate traffic dynamics in urban networks. This approach is particularly useful to unveil and explore the effects of various network-wide control strategies. The majority of modeling work using MFDs hinges upon the existence of well-defined MFDs without consideration of uncertain behaviors. However, both empirical data and theoretical analysis have demonstrated that MFDs are expected to be uncertain due to inherent instabilities that exist in traffic networks. Fortunately, sufficient amounts of adaptive drivers who re-route to avoid congestion have been proven to help eliminate the instability of MFDs. Unfortunately, drivers cannot re-route themselves adaptively all the time as routing choices are controlled by multiple factors, and the presence of adaptive drivers is not something that traffic engineers can control. Since MFDs have shown promise in the design and control of urban networks, it is important to seek another strategy to mitigate or eliminate the instability of MFDs. Furthermore, it is necessary to develop a framework to account for the uncertain phenomena that emerges on the macroscopic, network-wide level to address these unavoidable stochastic behaviors.

This first half of this work investigates another strategy to eliminate inherent network instabilities and produce more reliable MFDs that is reliable and controllable from an engineering perspective—the use of adaptive traffic signals. A family of adaptive signal control strategies is examined on two abstractions of an idealized grid network using an interactive simulation and analytical model. The results suggest that adaptive traffic signals should provide a stabilizing influence that provides more well-defined MFDs. Adaptive signal control also both increases average flows and decreases the likelihood of gridlock when the network is moderately congested. The benefits achieved at these moderately congested states increase with the level of signal adaptivity. However, when the network is extremely congested, vehicle movements become more

constrained by downstream congestion and queue spillbacks than by traffic signals, and adaptive traffic signals appear to have little to no effect on the network or MFD. When a network is extremely congested, other strategies should be used to mitigate the instability, like adaptively routing drivers. Therefore, without sufficient amounts of adaptive drivers, the instability of MFDs could be somewhat controlled, but it cannot be eliminated completely. This is results in more reliable MFDs until the network enters heavily congested states.

The second half of this work uses stochastic differential equations (SDEs) to depict the evolutionary dynamics of urban network while accounting for unavoidable uncertain phenomena. General analytical solutions of SDEs only exist for linear functions. Unfortunately, most MFDs observed from simulation and empirical data follow non-linear functions. Even the most simplified theoretical model is piecewise linear with breakpoints that cannot be readily accommodated by the linear SDE approach. To overcome this limitation, the SDE well-known solutions are used to develop an approximate solution method that relies on the discretization of the continuous state space. This process is memoryless and results in the development of a computationally efficient Markov Chain (MC) framework. The MC model is also supported by a well-developed theory which facilitates the estimation of future states or steady state equilibrium conditions in a network that explicitly accounts for MFD uncertainty. Due to the fact that current formalization of Markov Chains is restricted with a countable state space, some assumptions which redefine the traffic state and stochastic dynamic process need to be set for the MC model application in dynamic traffic analysis. These assumptions could be sabotaged by inappropriate parameter selections, producing excessive errors in analytical solutions. Therefore, a parametric study is performed here to illustrate how to select two key parameters, i.e. bin size and time interval to optimize the MC models and minimize errors.

The major advantage of MC models is its wide flexibility, which has been demonstrated by showing how this method could well handle a wide variety of variables. A family of numerical

tests are designed to include instability of MFD model, stochastic traffic demand, different city layouts and different forms of MFDs in the scenarios under static metering strategies. The results suggest that analytical solutions derived from MC models could accurately predict the future traffic state at any moment. Furthermore, the theoretical analysis also illustrates that Markov chains could easily model dynamic traffic control based on traffic state and pre-determined time-varying strategies by adjusting the transition matrix. Overall, the developed MC models are promising in the dynamic analysis of complicated urban network control under uncertainty for which simpler algebraic solutions do not exist.

Table of Contents

List of Figures

List of Tables

Acknowledgement

I would like to thank my advisor, Dr. Vikash Gayah, for giving me the opportunity to pursuit my Ph.D. degree at Penn State. Without his guidance and support, it would not have been possible for me to complete my study.

I also want to thank Dr. Eric Donnell, Dr. Ilgin Guler and Dr. Minghui Zhu for their sincere advice and insightful comments while writing this research project. I owe special thanks to Dr. Minghui Zhu for providing invaluable knowledge about Markov chain approximation method.

I am also grateful for the support from my friends, Andrew Nagle, Zhengyao Yu, Anthony Deprator and many others who stood by me in good and bad. I wish them well in all their endeavors now and hereafter.

Finally, I would like to thank my parents who provide me with unconditional support and love throughout this long journey. I am forever grateful.

Chapter 1

Introduction

Urban traffic congestion is a significant problem in the United States, particularly densely populated areas with over 1 million inhabitants. The costs of congestion in the US are severe: in 2011, congestion accounted for more than 5.5 billion hours of wasted productivity and 2.9 billion gallons of wasted fuel with a total estimated cost to society of $121 billion (Schrank et al., 2012). Furthermore, the worsening congestion has been a real threat to quality of life in terms of how it affects people's daily schedules. In cities with more than 3 million inhabitants, for instance, the average commuter suffers about 6 hours of congested road conditions during an average weekday (Schrank et al., 2012).

Building additional transportation infrastructure is not an optimal way to mitigate this congestion problem not only due to the high monetary costs, but because additional infrastructure construction also generates new travel demand (the induced demand phenomenon). This means that adding new roadway capacity will cause more people to travel, and this new travel demand might still exceed the expanded capacities of the renewed traffic network. Furthermore, building sufficient traffic infrastructure for peak-hour travel demands is also a waste of resources during the off-peak hours when it goes unused. Instead, an alternative to capacity expansion is the implementation of various traffic control strategies to alleviate the imbalance between increasing traffic demands and inability of the transportation infrastructure to support drivers. Such strategies, combined with advanced traveler information systems, communications and sensors, enable drivers to make better-informed decisions and allow existing transportation networks to be used more efficiently.

Traffic control strategies operate on a variety of scales within urban transportation networks. At the local level are strategies designed to improve traffic conditions at specific locations—e.g., a specific intersection or an arterial road. These strategies require a tremendous amount of detailed data to facilitate their design and to estimate their performance. Such data are possible to collect at individual locations through manual or automatic sources (such as in-pavement sensors and detectors). Reliable traffic models to describe traffic dynamics at the local scale have also been developed and refined for quite some time, including detailed car following models, the LWR theory of kinematic waves, and cell transmission models (Gipps, 1981; Lighthill and Whitham, 1955; Richards, 1956; Daganzo, 1994).

These aforementioned models can also be used to describe detailed traffic dynamics throughout an entire network to test large-scale traffic control strategies (e.g., a comprehensive coordinated signal timing plan that impacts multiple intersections). The network model is typically constructed by modeling each of its individual components. However, the process of collecting enough data to use each detailed model to assess the network-wide impacts of large-scale control strategies generally makes these approaches infeasible. Those approaches are also very time- and resource-intensive since very large urban networks are often made up of thousands or millions of individual agents, links and nodes, and each of these must be modeled individually. Furthermore, those approaches make it more difficult to see big picture trends and insights that are often lost among the complex analysis of detailed data.

Further complicating this is the fact that congestion is not a local phenomenon: queue spillovers from nearby junctions means that traffic in one region is impacted by traffic conditions in neighboring regions. In other words, congestion is actually a large-scale phenomenon, especially recurring congestion that occurs during busy peak hours. Thus, a network-level perspective is required when designing and implementing large-scale traffic control strategies. Such a perspective of studying large-scale traffic control strategies is particularly useful and has shown promise in

several urban settings across the world. Examples of large-scale traffic control include pricing entire busy downtown regions (known more commonly as congestion pricing), which has recently been implemented in London, Stockholm and Milan (de Palma and Lindensy, 2011), car entry restrictions in congested areas, for instance, in Singapore and Beijing (Zhang et al., 2003), and zonal-based adaptive signal control, such has been applied in Zurich (Benesty and Huang, 2003).

To alleviate this concern, recent work has developed aggregated traffic models that can be used to directly study and model large-scale network dynamics, and these models are useful to both develop and test various network-wide traffic control strategies. This model is commonly known as the Macroscopic Fundamental Diagram (MFD), and provides a relationship between average network density and flow on an urban traffic network. The former metric describes the accumulation of vehicles in the network and is analogous to how busy the network currently is, while the latter describes the productivity of the network. The model is able to describe in a physically realistic fashion all traffic states that may arise in a urban traffic network, from nearly empty networks that are not productive due to lack of demand to highly congested networks that are not productive because they are too busy. As will be described in the next chapter, recent work has shown that when a reliable and reproducible MFD model of an urban network exists, it can be used to feasibly describe traffic dynamics in a way that is sufficient to analyze various traffic management schemes, such as pricing strategies or perimeter flow control between distinct neighborhoods or regions. However, the existence of a reproducible and well-defined MFD is not universally expected. According to field data, these relationships are highly stochastic such that multiple levels of network productivity (or average flow) are observed for a given accumulation (or average density). Furthermore, simulated and empirical MFDs exhibit more complex phenomena such as multivaluedness and hysteresis behavior. Overall, these phenomena serve to decrease the applicability of MFDs for the development and analysis of large-scale traffic control strategies.

Besides the endogenous instability that exists within MFDs, other types of uncertainty also arise during the application of traffic control schemes that further limit the applicability of using MFDs to analyze large-scale control policies. Such random behavior might be caused by limitations of our ITS technologies and the randomness of individual driver behavior. Take the traffic signal control on the perimeter of urban network as an example. Existing research has used the existence of a well-defined MFD to develop perimeter flow control plans that limit the rate at which vehicles enter congested urban centers to ensure that overall network productivity is always maximized. In theory, it should then be fairly easy to carefully control traffic signals on the perimeter of the network in an attempt to achieve the desired vehicle entry rates. However, the time-varying nature of travel demand and the presence of aggressive and timid drivers would cause actual entries into the network to randomly fluctuate around the desired values. These factors could result in unexpected overall behavior and less efficient control strategies if this stochastic behavior is not explicitly accounted for in any modeling representation.

Thus, the purpose of this dissertation work is 1) to examine in more detail some of the features that might impact the uncertainty that is endogenous to the MFD and 2) to develop a probabilistic modeling framework that can explicitly incorporate this uncertainty (and other types of uncertainty) into large-scale network representations of urban traffic networks. The former will help unveil conditions under which well-defined MFDs might be expected in practice, which would allow transportation engineers the ability to model networks using this novel approach. The latter will then provide engineers with the tools necessary to adequately develop traffic control strategies that account for inherent uncertainties that arise on a macroscopic scale, which will hopefully lead to more efficient and effective large-scale traffic control strategies.

Chapter 2

Literature Review

This literature review is divided into two parts. The first part describes efforts at modeling large-scale urban transportation networks using aggregated or macroscopic modeling techniques. This includes historical efforts and more recent research on the Macroscopic Fundamental Diagram. The second part describes various traffic control strategies that have been developed using the Macroscopic Fundamental Diagram framework.

2.1 Aggregated Modeling of Urban Traffic Networks

Aggregate traffic models seek to describe the average behavior of vehicles when aggregated across an entire spatial region, usually a small urban network or neighborhood/region within an urban network. Effective aggregated models, if they exist, could be used to unveil inherent principles of the aggregated network behavior, without the distortions that typically arise due to the complexity of detailed models at the local level. This section will describe historical efforts to unveil such aggregate relationships and new advances that led to the development of the Macroscopic Fundamental Diagram.

2.1.1 Older Models

The first works on traffic flow descriptions began in the 1930s by Greenshields (1935). Greenshields postulated a linear relationship between average traffic speed and density at the local level—a point on a single link. Combining this assumption and the well-known relationship that

vehicle flow is equal to the product of average vehicle speed and density, Greenshields developed the first unimodal relationship between traffic flow and density on links. This relationship is depicted in Figure 2-1, and is now commonly known as the Fundamental Diagram (FD). The Fundamental Diagram clearly distinguishes between free flow traffic (the increasing portion of the curve) and congested traffic (the decreasing portion of the curve). However, the Greenshields' model and similar disaggregate link- or vehicle-based modeling efforts, such as car following models (Gipps, 1981) and cellular automation models (Daganzo, 1994), were derived from data measured at a single link and not sufficient to describe the entire complexity of traffic dynamics on urban road networks.

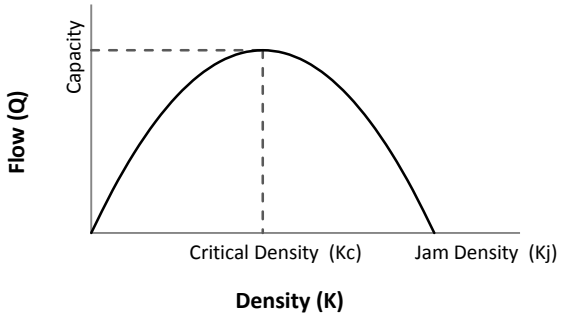


Figure 2-1. Greenshields' Model.

The notion of relationships between aggregated traffic metrics measured at the network-level was first proposed in the 1960s. During this decade, many researchers proposed various models to describe how vehicles behave on a network-wide level. Most efforts collected limited empirical data at multiple points across a network and tried to fit statistical models to describe the observed relationships (Smeed and Wardrop, 1964; Smeed, 1967; Thomson, 1967; Wardrop, 1968). For example, Smeed and Wardrop (1964) proposed a traffic model for central London, which described the observed relationship between average traffic flow and average speed using a

decreasing linear function between flow and the cube of average speed. Smeed (1967) developed another macroscopic model to describe maximum flow of vehicles that are able to enter the central area of a city as a function of multiple aggregate parameters, including network design, traffic control and traffic demand distribution, the area of the town, the fraction of the area devoted to roads and the capacity per urban street width. However, the model did not include the case when the demand exceeds capacity and thus could not describe congested (or oversaturated) network conditions. Thomson (1967) used empirical data collected over 14 years from central London to propose a decreasing linear relationship between average traffic flow and average speed. This research found that the linear relationships gradually changed each year, which was attributed to the geometric and intersection traffic control improvements throughout the network. Wardrop (1968) incorporated the average street width and average signal spacing into the function of average speed and flow, which also showed a monotonically decreasing, but not exactly liner relationship. Overall, the models described within this paragraph all explicitly found a monotonically decreasing relationship between traffic flow and speed. However, this functional form is unable to describe congested conditions that arise with both low flow and low speeds. Thus, these models were severely limited as they were not physically realistic for the full range of traffic conditions that could be expected in a network.

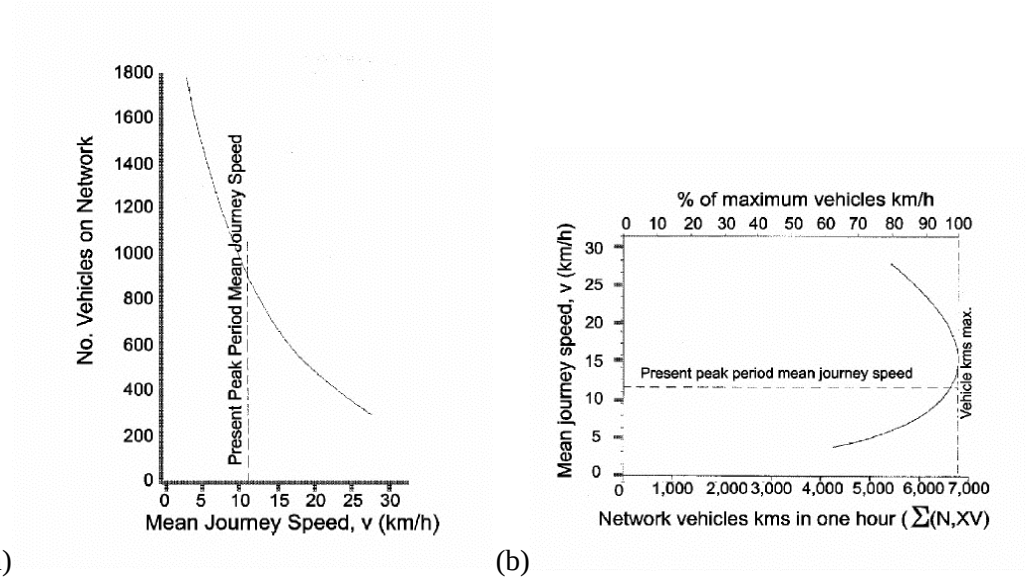
Godfrey (1969) was the first to propose a unimodal relationship between average vehicle speed and average vehicle flow. This study was groundbreaking in that it described traffic conditions measured across an entire network using the same basic shape as the Fundamental Diagram that described traffic conditions on a single link. The unimodal relationship is more in line with expectations and more physically realistic than previous efforts, as it is able to describe free-flow (low flow, high speed), capacity (high flow, moderate speed) and congested conditions (low flow, low speed) using the same relationship. However, sufficient traffic data were not available to fully confirm that such a relationship would actually exist on a network.

Following this proposal, several studies tried to develop models of network-level traffic performance that could describe the full range of traffic conditions that may arise. These models include other traffic variables in addition to the traditional variables like flow, density and speed. For example, Zahavi (1972) defined a parameter α as a measure to describe the relationship of traffic flow, density and speed, which worked as an indicator for traffic performance and network characteristics. Another effort, the two-fluid model (Herman and Ardekani, 1984), describes traffic performance in the urban network by keeping track of the fraction of vehicles running or stopped at any one point in time, and relates these fractions to other parameters in the network. Based on the two-fluid model, the aggregated travel behavior on the town traffic described with the average streets speed and the fraction of the delay time could be estimated by a function of the vehicle concentration. A number of field studies have been carried out to test the two-fluid model using data from Austin, Houston and other cities (Herman and Ardekani, 1984; Ardekani and Herman, 1987; Ardekani et al., 1985). Those two models were an improvement over the earliest efforts in that they were able to describe both free-flowing and congested traffic states. However, they were not without faults. For one, these models relied on some fairly restrictive modeling assumptions that were never verified using empirical data. More importantly, these models were static and were only able to describe expected traffic conditions based on some static input parameters. For this reason, they could not be used to model the dynamics of traffic in a large network over a period of time, such as a peak demand period.

2.1.2 MFD

Concept of MFD

As previously mentioned, Godfrey (1969) seemed to be the first to propose the concept of the Macroscopic Fundamental Diagram (MFD) as a unimodal relationship between average vehicle flow and speed within a network. The original relationships between average speed and total vehicle number (which is proportional to density) and between average speed and the total vehicle miles traveled in one hour (proportional to flow) that were first found by Godfrey are shown in Figure 2-2. Although the relationship between average flow and density, which describes the MFD, is not directly depicted, these analogous relationships suggest that a critical certain density exists that results in the highest average flow. For densities below this value, flow increases with density, and for densities greater than this value, flow decreases with density. Figure 2-2 also indicates that these relationships are well-defined from the free flow traffic state to almost gridlock, although this was never verified by Godfrey using empirical data. Simulation-based studies by other researchers were later performed that observed similar patterns at the network-level (Mahmassani et al., 1984, 1987; Olszewski et al., 1995).



(a) (b)

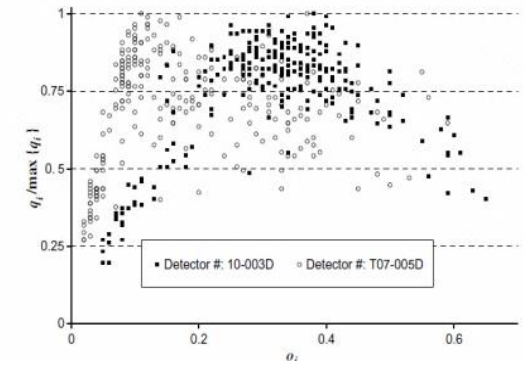
Figure 2-2. (a) Relationship between Average Speed and Number of Vehicles on Town Center Network; (b) Relationship between Average Speed and Total Vehicle Mileage on Network. (Godfrey, 1969)

The concept of the Macroscopic Fundamental Diagram (MFD, or network-level Fundamental Diagram) was recently reintroduced by Daganzo (2005; 2007) after almost two decades in which network-wide traffic relationships were mostly ignored in the literature. Daganzo (2007) showed that the existence of an MFD implies a similar relationship between the rate vehicles exit a network (or complete their trips) and the accumulation (or number of vehicles) in a network, as long as the average length of trips in the network remains more or less constant. This latter relationship has come to be known as the Network Exit Function, or NEF. The existence of such network-level relationships means that the state of a network can be described by simply monitoring aggregate vehicular accumulation at the neighborhood level. Furthermore, the NEF provides a dynamic model to predict how congestion will evolve over time in a network in a computationally efficient way. Such knowledge can then be used to inform the development and design control strategies to mitigate this congestion. However, the existence of a well-defined MFD

is the prerequisite for the application of these modeling techniques and the development of network-wide control strategies.

Daganzo and Geroliminis (2008) theoretically proved the existence of a well-defined MFD under the assumption that network congestion is homogenously distributed and travel demand is slow-varying. Simultaneously, Helbing (2009) also derived analytical theories of a well-defined MFD under various traffic states with a utilization-based approach. These studies related the MFD to inherent characteristics of a network (e.g., signal timing, total network length, number of traffic signals) and suggested that MFDs should be fairly stable as long as the network geometry does not change significant. Geroliminis and Daganzo (2007) first unveiled the existence of a fairly well-defined MFD using simulation of the San Francisco, CA urban traffic network. More importantly, Geroliminis and Daganzo (2008) were the first to empirically prove the existence of the MFD using combined experimental data from floating vehicle probes and 500 urban fixed detectors in the city of Yokohama, Japan. An example of their empirical relationship is depicted in Figure 2-3. Notice that the combined flow-occupancy diagram (the latter being proportional to density) created using local data from two separate detectors over the course of an entire weekday (Figure 2-3a) shows great disorder, especially when flows are maximal ($o_i \cong 0.3$). This general disorder persists on the link level when examining local behaviors. However, this disorder disappears when flow-density relationships are constructed with the aggregate data obtained from all of the detectors in the network, shown in Figure 2-3b. What is especially interesting is that the data observed in Figure 2-3b is obtained from very distinct periods (denoted as A1…D2), including both morning and evening rush hours and weekdays and weekends. Thus, this work showed that traffic data might be more organized when studied at the aggregate (network-wide) level and indicates that the urban MFD may not vary with time or demand, but instead exists as a reproducible phenomenon that is a property of the network itself.

(a)



(b)

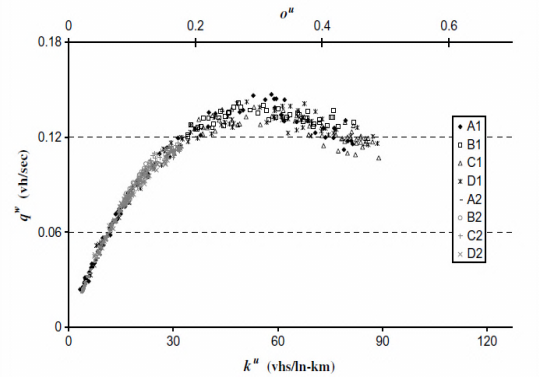


Figure 2-3. Loop detector data: (a) flow vs. occupancy for two single detectors across a day; (b) average flow vs. average occupancy from all the detectors across two different days. (Geroliminis and Daganzo, 2008)

Well-defined Urban MFD

While the work of Geroliminis and Daganzo (2008) is promising, experimental data for different cities reveal that not all MFDs are as well behaved as the one observed in Yokohama. Specifically, instability phenomena arise in the urban MFD, which limits the applicability of the MFD model in practice. Buisson and Ladier (2009) collected the traffic data from 483 loop detectors in the urban city of Toulouse, France, and the resulting MFD exhibited a clear clockwise hysteresis-like shape under the time-varying travel demand, shown in Figure 2-4. To provide an explanation of this type of behavior, Daganzo et al. (2011) theoretically analyzed urban network behavior and demonstrated that an endogenous instability exists even in idealized networks with perfectly homogeneous layout and time-independent, spatial uniform travel patterns. Gayah and

Daganzo (2011a) further explored the endogenous instability under the time-varying travel demand and reasonably explained the existence of the clockwise hysteresis phenomenon. This work suggests that such hysteresis patterns are not unlikely and arise due to the aforementioned instabilities. Fortunately, these works also showed that the presence of adaptive drivers that route themselves to avoid localized pockets of congestion could mitigate this instability and reduce the likelihood of hysteresis loops from arising. This provides an explanation of well-defined MFDs in the field (like Yokohama and the Toulouse on June 6, 2008) as drivers navigate adaptively in real life. In the case of the hysteresis of Toulouse, further examination of the data reveals a large event on the day of June 13, 2008 that specifically caused the clockwise loop to form and this too is consistent with the theoretical work proposed.

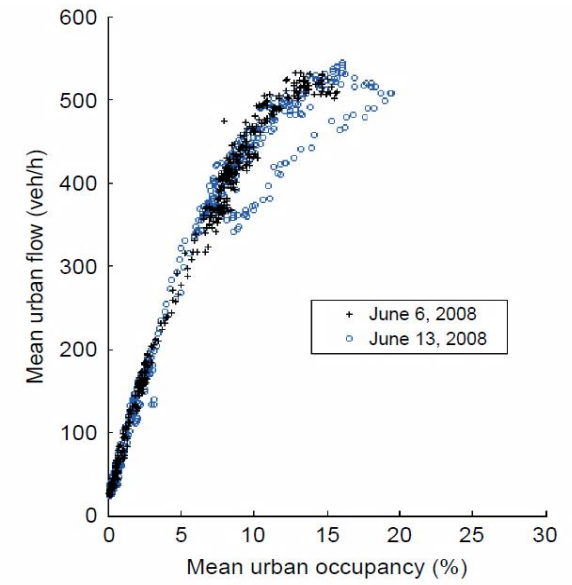


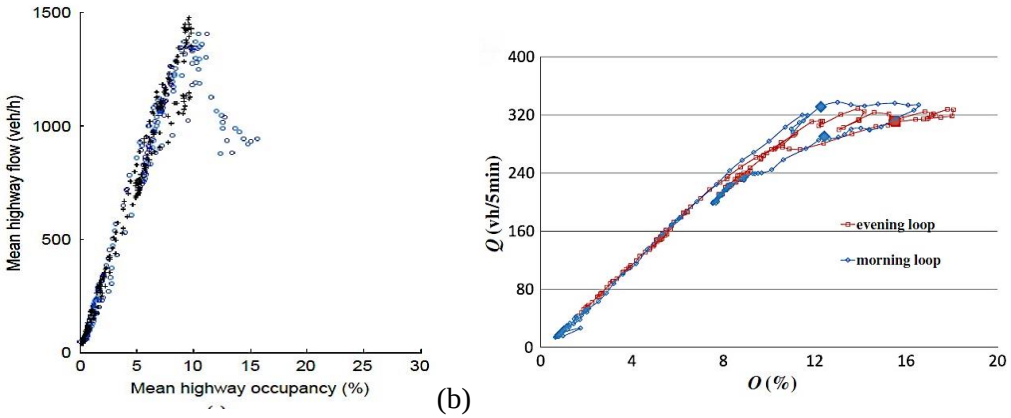
Figure 2-4. Urban MFD in Toulouse, France. (Buisson and Ladier, 2009)

Additional work has shown that the variation of vehicle densities on individual links cannot be avoided in the stable equilibrium traffic state and remains an impediment for the existence of well-defined urban MFDs. Mazloumian et al. (2010) studied the effects of the spatial variability of

density at the network level and found that only sufficiently small density variations provide robust and well-defined macroscopic functional relationships. However, strict homogeneity of traffic density is not necessary for the existence of a reliable urban MFD (Geroliminis and Sun 2011b). Instead, well-defined MFDs would arise as long as congestion patterns remain reproducible from day-to-day. Other factors, such as different degrees of variability in link lengths, signal characteristics for different city topologies and signal structures, have been evaluated by Geroliminis and Burak (2012). They also pointed out that average network capacity can be significantly smaller than predicted from previous theory because of the correlations developed through the different values of signal offsets. Therefore, efficient traffic control for the urban network could be utilized to find the optimum maximum traffic flow.

Freeway MFD

Compared with the arterials and local streets in an urban network, a freeway network shows different topological and control characteristics. Therefore, the network-wide MFD of freeways might demonstrate different properties from the urban MFD. Several field surveys have shown that the scatter data from freeways always follows a hysteretic pattern. Buisson and Ladier (2009) collected data measures with 45 loop detectors along a freeway area in the city of Toulouse, France, to build a MFD with the hysteresis shape, shown in Figure 2-5a. And the empirical data from a larger network collected by Geroliminis and Sun (2011a) with 600 loop detectors from Minneapolis freeways also demonstrates similar scatter and hysteresis phenomena in the network-wide MFD, shown in Figure 2-5b. In general, these empirical data failed to provide enough evidence to demonstrate the existence of a well-defined and reproducible MFD for a freeway network.



(a) (b)

Figure 2-5. (a) Freeway MFD in Toulouse, France (Buisson and Ladier, 2009); (b) Freeway MFD in Minneapolis. (Geroliminis and Sun, 2011a)

Cassidy et al. (2011) demonstrated that a well-defined freeway MFD should exist only under the stringent conditions that all the lanes of a freeway system are in similar uncongested or congested traffic states. However, these conditions are not likely to exist for a large freeway network and thus a well-defined freeway MFD cannot be universally expected. Daganzo (2011) utilized an idealized KW (kinematic wave) and CTM (cell transmission model) to demonstrate that freeway traffic is just unstable under congested states, which should make hysteresis loops expected in freeway MFDs. More importantly, this research pointed out that driver adaption does not play as large of a role on freeway networks as in urban networks because the freeways are long and opportunities to adapt are so few. Recently, Saberi and Mahmassani (2012) used data from Portland, Oregon, and further analyzed the impact of the spatial and temporal distribution of congestion on the formation and characteristics of hysteresis patterns. They found that the maximum network average flow is not a constant value but varies across different days. With a larger scale of data from Chicago (Illinois), Portland (Oregon), and Irvine (California), Saberi and Mahmassani (2013) further confirmed and characterized the inconsistent hysteresis and capacity drop phenomena in freeway networks.

These studies provide theoretical and empirical support that reproducible freeway MFDs should not be universally expected, especially when the freeway network covers a large spatial area and contains longer link lengths. But for urban networks, the presence of adaptive drivers could efficiently mitigate the endogenous instability and guarantee the existence of the well-defined and reproducible urban MFDs (Daganzo and Gayah, 2011; Gayah and Daganzo, 2011a).

Adaptive Signals

The sufficiency of adaptive drivers, who could spontaneously route their paths according to current congestion distribution, cannot always be guaranteed, especially during the irregular disturbances of traffic states or under the lack of access of sufficient real-time information. Therefore, other traffic control strategies should be considered to stabilize the consistent performance of urban network. One potential solution is adaptive traffic signal control that can force drivers away from congested areas and towards less congested areas (or allow drivers to escape congested areas and keep them from leaving less congested areas). Adaptive signal control with the ability to optimize allocation of the right of way according to real-time traffic information has been applied in many countries since the early 1970's (Schroeder, 1989). The two most widely utilized systems are SCATS (Sydney Coordinated Adaptive Traffic System) and SCOOT (Split Cycle and Offset Optimization Technique). Since 1990, FHWA has recognized the benefits and several programs to develope adaptive signal systems have been implemented (Ghamann et al., 2004).

deGier (2011) utilized a stochastic cellular automation model to compare the effects of adaptive signals and fixed time signals at intersections. The simulation results showed that adaptive signals could reduce total travel times and limit variation in individual travel times. Expanding this work from a single intersection to a network-level, Zhang et al. (2013) also adopted the same cell

automation model to study the MFD of an idealized urban network under different traffic control systems applied across the network. The authors found that the choice of traffic signal systems significantly impacts the shape of urban MFD; see Figure 2-6. However, even with adaptive signals, the urban network still has a tendency towards gridlock when the average traffic density of the urban network is less than the maximum physically possible density of a single link. This demonstrates that adaptive signals help to mitigate the instabilities in urban networks that cause poorly defined MFDs, but that this ability is limited. Unfortunately, the work was also deficient in that the authors studied just a few adaptive signal control schemes and were not able to generalize the impacts of adaptive signals to other network types or conditions.

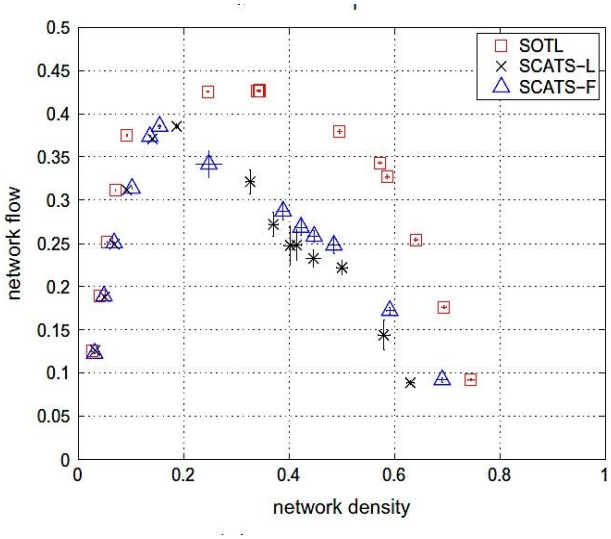


Figure 2-6. Stationary MFDs under different traffic control strategies[1]. (Zhang et al., 2013)

[1] SCATS-L is a model of SCATS with linking and adaptive cycle lengths; SCATS-F is a version of SCATS-L, with no signal linking, described in Zhang et al. (2013) and SOTL is self-organizing traffic lights, described in de Gier et al. (2011).

2.2 Traffic Control with MFD

To mitigate the congestion caused by high traffic demands in urban cities, and alleviate the corresponding negative effect, such as excessive delays, increased environmental impacts and reduced safety, various urban traffic control strategies have been proposed and applied in practice (Dinopoulou et al., 2006). The effective application and analysis of traffic control strategies require appropriate tools to optimize network-level control strategies and to study their impacts on traffic congestion patterns. The MFD model provides a reliable network-level description with the ability of the dynamic network analyses, and thus it works well as an efficient tool for large-scale traffic network modeling. Extensive research has been performed with the urban MFD models to both develop and study traffic control strategies at the urban network level.

2.2.1 Perimeter Flow Control (Metering/Gating)

Perimeter flow control (otherwise known as metering or gating) is a practical traffic control strategy to limit vehicle access to critical facilities. This strategy is frequently applied on freeways (e.g., ramp metering) and in some urban areas (e.g., Singapore). The basic principle behind metering is to control the input flow to some protected areas by holding vehicles in areas that are less critical. In other words, metering keeps vehicles held in locations in which they cause the least damage to traffic. Applied in urban networks, this involves holding queued vehicles outside of the congested urban network where they do not contribute to the existing congestion problems.

Metering was firstly implemented on freeways to manage the rate of automobiles entering the freeway on-ramps in the 1960s. Since then metering has been deployed in many urban areas, such as Los Angeles, San Diego, Sacramento, etc. Aided by advanced Intelligent Transportation System (ITS) technologies, the metering strategy has been utilized as an efficient and practical tool

against over-saturation by controlling the travel demand (Wood et al., 2002; Bretherton et al., 2003; Luk and Green, 2010). Gal-Tzur et al. (1993) proposed a signal timing design with the goal of metering traffic entering congested urban areas. However, the method was only designed for small networks with only one critical intersection. Therefore, it failed to solve the oversaturation on the large-scale urban network.

Fortunately, the MFD urban network model makes it possible to analyze the metering strategy on the network-level. Daganzo (2007) first tested the perimeter strategy with the corresponding MFD model of the homogeneous urban network. Based on the MFD, the efficiency of the urban network could be significantly improved by gently controlling the input flow, which further helps the network avoid its natural tendency towards gridlock. However, although the basic idea was proposed in this work, a feasible implementation strategy that might be used in practice was not provided.

Haddad et al. (2012) and Geroliminis et al. (2013) proposed a model predictive control (MPC) approach to design a metering strategy for two-region networks (note: two-region networks imply that the urban network can be subdivided into two unique regions, each with their own unique MFD). But the application of the MPC approach requires sufficient real-time traffic data, which is a serious challenge for the implementation of control in practice. Li et al. (2012) proposed an algorithm to determine fixed-timing signal timing to hold some proportion of traffic outside, albeit without adaption to the real-time conditions.

Keyvan-Ekbatani et al. (2012a and 2012b) proposed a real-time feedback-based gating methodology based on the urban MFD model. A simulation based on the real network of Chania, Greece, also demonstrated that feedback-based metering could effectively improve the mobility of urban traffic with the full information of the urban traffic. However, full access to information is feasible for the traffic simulation, which is not the case for the application. Ortigosa et al. (2014) evalated the data requirement for a possible implememtation of a MFD control scheme in an urban

area. This study showed that full data are not necessary for the effective application of the control strategies. Keyvan-Ekbatani et al. (2013a) demonstrated that feedback-based gating is also efficient to alleviate congestion when using reduced MFDs that rely on just a subset of data that might actually be available for real-time control. Later, the strategies were expanded with multiple boundaries to cope with different level of congestion (Keyvan-Ekbatani et al., 2013b). Recently, Keyvan-Ekbatani et al. (2014) points out that feedback-based control could be applied further upstream of the protected area with better urban mobility, allowing more flexibility in the ability to implement gating control. In the case of multi-region cities with multiple congestion centers, control policies are more complicated and less understood. Haddad and Geroliminis (2012) analyzed stability of urban traffic under the feedback perimeter metering control in the two urban regions. And Aboudolas and Geroliminis (2013) extended the strategy of perimeter metering to the urban network with multiple regions.

One drawback in the literature is that most of the previous studies were carried out with the assumption that a well-defined MFD exists with minimal scatter and no hysteresis patterns. However, discussed in the preceding sections, real MFDs tend to exhibit stochastic behavior and are not always well-defined. Thus, additional work needs to be performed to explicitly account for the uncertainty in traffic behavior that arises on the macroscopic level.

2.2.2 Road Congestion Pricing

Road pricing involves directly charging fees for the use of road facilities. Although road pricing could generate the revenue for infrastructure financing, the control strategy could also be applied as a transportation demand management tool to alleviate traffic congestion by changing traveler behavior. This practice is more commonly known as congestion pricing. Congestion pricing has been applied to some cities and road areas, such as the electronic road pricing in

Singapore, the London congestion charge, the Stockholm congestion tax, the Milan Area C, and high-occupancy toll lanes in the United States (Litman, 2006; Eliasson and Mattsson, 2006; Rotaris et al., 2010; de Palma and Lindensy., 2011).

Researchers have long explored the economic models to predict traveler behavior based on the trip costs. Yang and Huang (2005) gave a comprehensive literature summary of pricing models, which could be classified into marginal-cost pricing models and bottleneck models. The marginal-cost model was proposed by Pigou (1912) and developed by Vickrey (1963) and others researchers (Marchand, 1968; Yang and Huang, 1998; Small and Yan 2001; Verhoef, 2002; Maruyama and Sumalee, 2007). However, the model could only be applied to analyze the stationary traffic state. To overcome this, Vickrey (1969) proposed a basic bottleneck model for the dynamic case that deals with time-dependent equilibrium distribution of arrivals at a bottleneck (Smich, 1984; Daganzo, 1985). The Vickrey Bottleneck Model has been used to mainly evaluate toll policies (Arnott et al., 1990; Daganzo and Garcia, 2000; Laih, 1994). However, these works only pertain to the case when congestion occurs at a single link with a constant output rate, while the capacity of the freeway and urban networks would vary according to the different traffic states.

For the pricing control application of a freeway network, Lago and Daganzo (2007) analyzed the morning commute problem for some freeway networks, considering the effects of the queue spillovers and merging interactions to the change of the capacity of freeway network. To deal with complexity of the urban network, the MFD model was later utilized to depict the traffic dynamics in the entire urban network. Geroliminis and Levinson (2009) combined MFD models and the Vickery Bottleneck Model to estimate the corresponding travel demands in the network with only cars. Gonzales and Daganzo (2012) extended the research in the network combined with cars and public transit. Zheng et al. (2012) developed and tested a dynamic cordon-based congestion pricing scheme, in which tolls are controlled by an MFD model. In all cases, the pricing scheme created using the MFD could effectively save travel time without extra congestion in the

neighborhood. However, like the gating case before it, previous studies assume a well-defined MFD exists with minimal scatter and no hysteresis patterns, which is not always realistic.

2.2.3 Network Design & Routing

Both perimeter metering and congestion pricing improve the mobility of urban traffic by limiting the excessive accumulation of vehicles in some areas, and keeping the urban traffic at an optimal state with the maximum output rate—the capacity of the urban network. Other strategies, such as the network design and routing, could improve the urban traffic by expanding the capacity of the network itself. Gayah and Daganzo (2012) analyzed the effects of one-way and two-way operation on the capacity of the urban network traffic. They found that two-way streets generally provide larger capacities than one-way streets if trips are short, and that they can always provide higher capacities if left turns are banned at intersections. Ortigosa and Menendez (2014) have shown analytically that it is possible to remove some streets from a city without worsening traffic capacity excessively. Recently, Knoop et al. (2014) analyzed how the network structure influences the shape of the urban MFD. It is possible that adding an arterial could actually decrease the maximum production if not well designed.

The studies on routing strategies also have been spread on the aggregate level. Wang et al. (2006) and Wang (2008) evaluated various routing strategies using the aggregate variables including the network capacity and the average travel time. Recently, Knoop (2012) tested the routing strategies based on the detailed information of all links and on the accumulation data of sub-networks. And the traffic control based on the MFD has been demonstrated to be the same effective routing strategies as the routing strategies on all links to reduce bottleneck occurrence.

The network design and routing both shows significant impact to the formation of MFD models, and should be carefully considered to optimize the urban traffic operations. However, how

those strategies would change the stochastic characteristics of urban MFDs still remains and should be further studied.

Chapter 3

Research Objectives

According to the research literature documented above, the urban MFD model only works as an efficient and reliable tool to evaluate the performance of urban traffic if an MFD is well-defined—that is, describes the flows observed at all traffic states without significant scatter or hysteresis behavior. However, because of the endogenous instability that exists within urban networks (Daganzo et al., 2011; Gayah and Daganzo, 2011a), well-defined MFDs should not be universally expected. Fortunately, extensive research has shown that sufficient amounts of adaptive drivers (that route to avoid congestion) could spread congestion more evenly in a network and guarantee that the MFD serves as a more reliable traffic analysis tool. However, drivers cannot route themselves adaptively all the time as routing choices are determined by multiple factors and there might exist unexpected traffic disturbances, non-redundant routing options or lack of information to make appropriate routing decisions. Furthermore, the presence of adaptive drivers is not necessarily something that traffic engineers can control.

Another strategy that is more reliable and controllable from an engineering perspective, and which could complement presence of adaptive drivers in urban network, is the presence of adaptive traffic signals. Adaptive signals have the potential to help drivers avoid congestion or escape congested areas more quickly, which mimics the impacts of adaptive driver routing, by changing signal timing according to local conditions around the signal. Previous research has suggested that this has the potential to provide more stable and well-defined MFDs although a comprehensive analysis has never been performed. In light of this, the first research objective is to

thoroughly study how to use adaptive traffic signals to stabilize the urban MFD, and compare its impacts with adaptive drivers to understand the limitations of two different control strategies.

Even if the combined application of adaptive signals and adaptive drivers could provide a well-defined and reliable urban MFD, it is impossible to achieve a perfect one-to-one relationship between average traffic flow and density. Some scatter or uncertainty will always exist where the network exhibits multiple possible traffic flows for the same average density; see the "well-defined" Yokohama MFD in Figure 2-3b for which flows still deviated by about 10% of the average value in some cases for a given density. To summarize, the inherent stochasticity of an urban MFD could be mitigated, but cannot be avoided altogether. Besides the inherent stochasticity of the MFD itself, another stochastic factor in traffic management comes from the traffic control strategies themselves. It is impossible to get drivers to behave exactly as planned as each driver behaves differently. Take the application of perimeter metering as an example. Traffic signals can be used to control the desired inflow into the urban network. However, the presence of an inattentive driver can cause this value to drop below the intended amount while aggressive drivers might enter the network at a higher rate. Unfortunately, the existing approaches that study network-wide control using the MFD all rely on the presence of a well-defined (i.e., not uncertain) MFD curve. Therefore, the second research objective is develop a probabilistic framework to account for the uncertain behavior that emerges on the macroscopic, network-wide level. This framework can be used to study how existing control strategies operate under uncertainty and can also be used as a tool to help develop traffic control strategies that maximize traffic mobility within a network while considering the stochasticity in the urban MFD and traffic control strategy application.

Those two questions could help us gain a more thorough understanding of how aggregate-level traffic uncertainty can be mitigated, and how it can be incorporated into modeling and design of control. The research about adaptive signals and adaptive drivers could guide us to stabilize the urban traffic network and provide a more reliable MFD. To deal with the stochasticity that we

simply cannot avoid, the probabilistic framework will help us accommodate this in existing modeling frameworks. This can then be used to develop strategies that explicitly account for this stochasticity that arises on a macroscopic level to guarantee more robust and reliable urban traffic control.

Chapter 4

MFD Model with Adaptive Signals Control

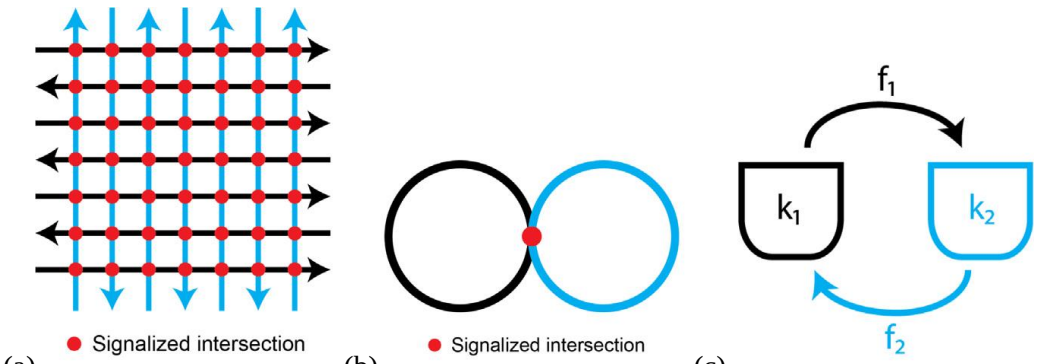
The section describes research performed to accomplish the first research objective: determining the impacts of adaptive signal control on the stability of network traffic and the existence of well-defined MFDs. Multiple abstractions of an idealized network are used to explore the effects of locally adaptive signal control on the nature of the urban MFD model. These abstract networks are also used to directly compare the abilities of adaptive signals and adaptive drivers to mitigate instabilities that arise in urban traffic networks.

4.1 Idealized Urban Network Description

This section describes the idealized network that will be used to test the impacts of adaptive signal control, and two abstractions of this network that are used to simplify the analysis.

4.1.1 Idealized Grid Network

The idealized urban network considered here is a homogeneous square grid constructed of alternating one-way links; see Figure 4-1a. At each intersection, a vehicle approaching the intersection has only two potential movements: either go through or turn onto the cross street.



(a) (b) (c)

Figure 4-1. (a) Homogeneous grid network; (b) two-ring abstraction; and (c) two-bin abstraction.

Traffic on each link is assumed to follow the same fundamental diagram, $Q(k)$. For simplicity, a triangular fundamental diagram is assumed with free flow speed, v, backward wave speed, w, capacity, q_m and the jam density, $k_j = q_m(1/v + 1/w)$, shown in Figure 4-2. According to the triangular fundamental diagram, the output flow at each link i could be described by the following piece-wise linear function:

$$q_i = f(k_i) = min\{ v * k_i, w(k_j - k_i)\} \quad \textbf{(4-1)}$$

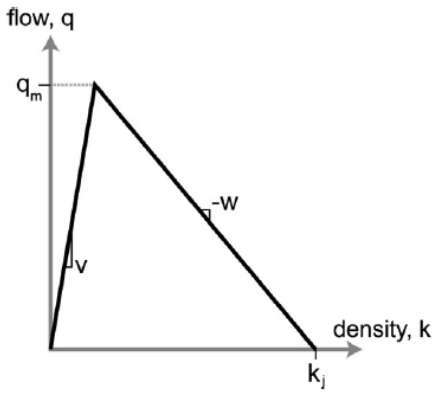


Figure 4-2. Fundamental Diagram of each single link.

Vehicles are assumed to travel in the network without specific destinations. Instead, routing is entirely determined by turning probabilities assigned at each intersection. A proportion of the drivers, α, are assumed to be adaptive and actively seek to avoid congestion. The remaining $1-\alpha$ fraction of the drivers are assumed to turn randomly at each intersection with some fixed probability P_T, independent of traffic conditions downstream. These vehicles are assumed to be following a predetermined path since the routing decision is determined independent of the traffic state, as if made ahead of time. The α fraction of adaptive drivers make routing decisions that are dependent on existing traffic conditions. We adopt a simple rule to emulate this behavior: if a random turning maneuver would have made this adaptive driver turn onto a less congested link than the alternative through link, then the turn maneuver is made. If the random turning maneuver requires a turn onto a more congested link, the vehicle would instead make a through maneuver to avoid the traffic congestion that would have otherwise been experienced.

Since there are no destinations in the network, vehicles will travel indefinitely until they reach the downstream-most end of a particular street. Once a vehicle exits the network at the downstream end, another is simultaneously inserted at the upstream-most end of that same street. In this way, the network maintains periodic boundary conditions and the number of vehicles remains fixed.

All intersections are controlled by traffic signals with fixed cycle length, C, and zero offset between adjacent signals. We consider a family of adaptive traffic signal control strategies that operates as follows. A fraction $1-\gamma$ of the cycle length is split evenly between the two competing directions, providing a minimum green time of $(1-\gamma)*C/2$ to each. The remaining γ of the cycle is allocated dynamically every cycle based on real-time traffic conditions. For simplicity, we assume that this green time is divided between the two directions proportional to vehicle densities on the upstream links. The parameter $\gamma \in [0,1]$ represents the level of signal adaptivity; see Figure 4-3. The extreme value $\gamma = 0$ represents traffic signals that have fixed timings with C/2 allocated

to each direction. At the other extreme, the value $\gamma = 1$ represents fully adaptive signals in which all of the green time is allocated proportional to upstream densities. Of course, more realistic adaptive control schemes consider other factors, such as dynamically changing the cycle length or signal offsets. However, for analytical completeness we choose to consider a simpler strategy that can be studied exhaustively, as will be done through simulations and analytical models here.

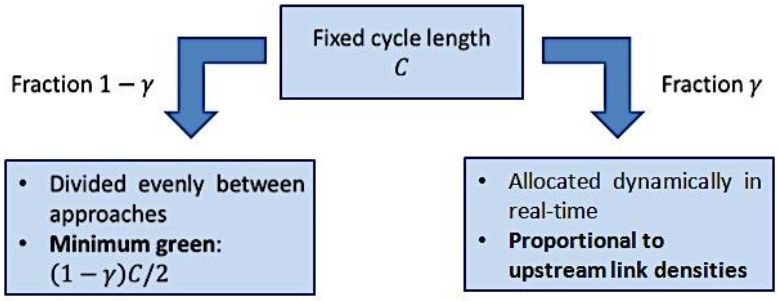


Figure 4-3. Basic principle of the adaptive signals control.

Once a vehicle reaches the downstream-most end of a street, another is simultaneously reinserted into the network at the upstream-most end of that street. In this way, the number of vehicles in the network remains fixed, and the network maintains periodic boundary conditions. In other words, it is as if the network stretched infinitely long in all directions. Although the urban network described here is simple and relies on some highly idealized assumptions, it is not feasible to obtain MFD model directly by analyzing traffic behavior on each of the individual links due to the large number of links and stochastic vehicle behavior at the intersections. Micro-simulations of this idealized network might be used, but these tools are highly stochastic and tend to cloud findings such that generalizable trends are often mistaken for stochastic noise (and vice versa); see Daganzo et al. (2012) for more details. To alleviate these issues, two physically equivalent abstractions of this network are proposed that help unveil insights on the impact of adaptive signal control on congested networks.

4.1.2 Two-Ring Abstraction

The idealized grid network described here is physically equivalent to a system of two interconnected ring roads under some simplifying assumptions. These are described here.

Assume that vehicles are initially distributed across links in the grid network such that the distribution of vehicles across all east-west streets is identical (and the same for all north-south streets). Moreover, assume that the random turning probabilities P_T and proportion of adaptive drivers α are constant and equal on each link at the beginning of any analysis period. Lastly, assume that the signal adaptivity parameter γ at each intersection is the same at all locations and constant over time. Under these assumptions, the density profile on all east-west links (or north-south links) will always be identical at any moment since all east-west links (or north-south links) share the initial states and follows the exactly same rules that describes their dynamics over time. Thus, the behavior of all east-west links (similarly north-south links) can be described by examining the behavior of just one.

In this way, the grid network can be abstracted into the much simpler "two-ring" network depicted in Figure 4-1b. This system is much simpler to study, as we only need to consider the movement of vehicles along two links and the behavior of the single intersection that connects them. The analysis of this two-ring model will be performed through the use of a cellular automation model that was created specifically to study the impact of the adaptive signal control scheme proposed.

4.1.3 Two-Bin Abstraction

Although the two-ring model is much simpler than the idealized grid network, it must still be analyzed using simulation techniques due to the stochastic nature of vehicle behavior—

specifically, the turning behavior at the intersection—and the presence of the traffic signal. Existing analytical descriptions of the two-ring model are only able to describe dynamics when the junction in the model is not signalized (Jin et al., 2013). Thus, to get the analytical insights on the behavior of this system, the two-ring model is abstracted further.

Assume now that vehicles within each ring are always uniformly distributed across its length, and that turning vehicles are able to push their way into the receiving ring if the latter is not completely jammed. Under these conditions, the state of each ring is completely described by the number (or density) of vehicles on it at any moment. The two-ring network can be modeled as a simple system of just two "bins" of vehicles that interact through the turning flows between them, as shown in Figure 4-1c. This two-bin network abstraction was previously used by Daganzo et al. (2011) to show that bifurcations and instabilities exist in congested urban networks, which result in a tendency towards gridlock.

4.2 Two-Bin Analysis

4.2.1 Analytical Description of the Two-Bin Model

Consider now the abstract two-bin model depicted in Figure 4-1c. In this model, $i \in \{1, 2\}$ is used as an index for each bin. The density and flow of vehicles within each bin is denoted k_i and q_i, respectively. The average density, k_T, and average flow, q_T, for the idealized homogeneous network is an unweighted average of the individual densities and flows within each bin (Daganzo et al., 2011):

$$k_T = \frac{k_1 + k_2}{2} \qquad \textbf{(4-2)}$$

$$q_T = \frac{q_1+q_2}{2} \quad \textbf{(4-3)}$$

A function $Q_i(k_i)$ is used to describe the fundamental diagram, which relates the density and flow within each bin. However, the triangular fundamental diagram for the idealized unsignalized network shown in Figure 4-4a cannot be used here, because traffic signals at the intersections disrupt flow. Wu et al. (2011) used the empirical data to unveil that the relationship between average flow and density on signalized arterials generally follows a trapezoidal shape, and the height of this trapezoid is proportional to the amount of green time provided to the direction of interest (Daganzo and Geroliminis, 2008).

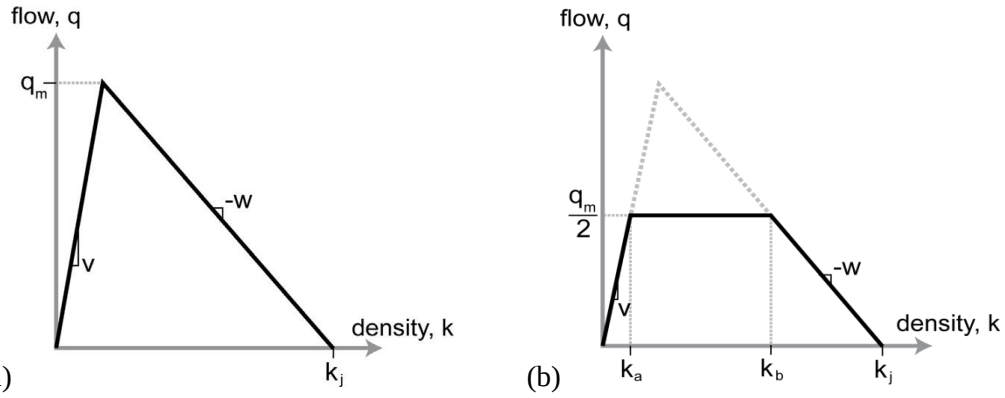


Figure 4-4 (a) Fundamental diagram of each bin without signal control; (b) Fundamental diagram of each bin with the fixed signal control.

We now develop a rule to determine $Q_i(k_i)$ under the adaptive signal control scheme considered here. For any γ, the green time ratio provided to bin i, g_i, described by the control logic depicted in Figure 4-3 is:

$$g_i = (1-\gamma)\frac{1}{2} + \gamma\frac{k_i}{2k_T}. \quad \textbf{(4-4)}$$

The maximum flow in each bin must be limited by the maximum flow through the intersection, which is equal to $g_i \cdot q_m$. Although the traffic signal may impact the increasing and decreasing parts of the fundamental diagram, as well as the maximum (capacity) flow (see Daganzo and Geroliminis, 2008), we assume here that these portions of the curve are linear and have constant slopes v and $-w$, respectively, to simplify our analysis. Since the changes of those parts are relatively minor compared with capacity flow, the overall results are not significantly affected by this assumption[2]. This results in the following function to relate flow and density in each bin under adaptive signal control:

$$q_i \doteq Q_i(k_i) = min\{\, vk_i,\ g_i q_m,\ w(k_j - k_i)\}. \qquad \textbf{(4-5)}$$

The first term in Equation 4-5 represents free flow conditions, the second represents capacity or maximum flow, and the last represents congestion. When the traffic signals are fixed, i.e. $\gamma = 0$, $g_i = 0.5$, and the MFD is equal to the fundamental diagram illustrated in Figure 4-4b.

To study the dynamics of the two-bin system, we also need to define the flow of vehicles between the bins. Based on the routing rules previously described, at any moment the fraction of vehicles passing the intersection that turn out of bin i, denoted p_i, is:

$$p_i = \begin{cases} P_T & \text{for } k_i > k_T \\ (1-\alpha)\cdot P_T & \text{for } k_i \le k_T \end{cases} \qquad \textbf{(4-6)}$$

The flow turning of bin i, denoted f_i, is the product of the flow in bin i and the fraction of vehicles turning:

[2] Numerical tests have confirmed this to be the case. The qualitative results presented here are not significantly impacted by this assumption.

$$f_i \doteq F_i(k_i) = p_i q_i. \quad \textbf{(4-7)}$$

This dynamic system will only be in equilibrium when the turning flows out of each bin are the same; i.e., when:

$$f_1 = f_2 \quad \textbf{(4-8)}$$

Two types of equilibrium states may exist: stable or unstable. Unstable equilibria are short-lived and will not be reproducible over the long run with minor perturbation. But stable equilibria are those that are robust, and could be used to describe the reproducible aggregate behavior of the network. Thus, it is important to identify these stable equilibria. The character of equilibrium states can be determined by examining how the system evolves after a minor perturbation according to the research of Daganzo et al. (2011). These minor perturbations will shrink with time at stable equilibrium, and grow with time at unstable equilibrium. Mathematically the equilibrium is stable if and only if:

$$dF_1(k_1)/dk_1 + dF_2(k_2)/dk_2 > 0. \quad \textbf{(4-9)}$$

The critical conditions of Equations 4-8 and 4-9 will be used to determine the set and character of all equilibrium states when adaptive signal control and adaptive drivers are applied.

4.2.2 Stability of Adaptive Signal Control Only

To illustrate the benefits of adaptive signals, we first analyze the urban state and identify the equilibrium states when the two-bin network operates with fixed traffic signals; i.e., when γ= 0. In this case, the green ratio, $g_1 = g_2 = 0.5$, as previously described. The flow-density relationship for each of the subnetwork is also constant and is shown in Figure 4-4b.

The complete set of equilibrium states that satisfy the conditions of Equation 4-8 in the fixed traffic signal control case is presented on a phase diagram in Figure 4-5. This diagram visually depicts the range of possible states that the system can exist in, defined by the pair (k_1, k_2). On the phase diagram, thin dotted grey lines on the Figure 4-5 represent iso-density contours that represent a fixed total number of vehicles (or total density of vehicles) that exist in the system. Because of the boundary conditions for the grid network, the number of vehicles is always kept the same once initially chosen. Thus, the traffic state would only evolve along lines parallel to the iso-idensity contours until it finally reaches an equilibrium state. The system should tend to remain in that equilibrium state until a random disturbance causes it to evolve again.

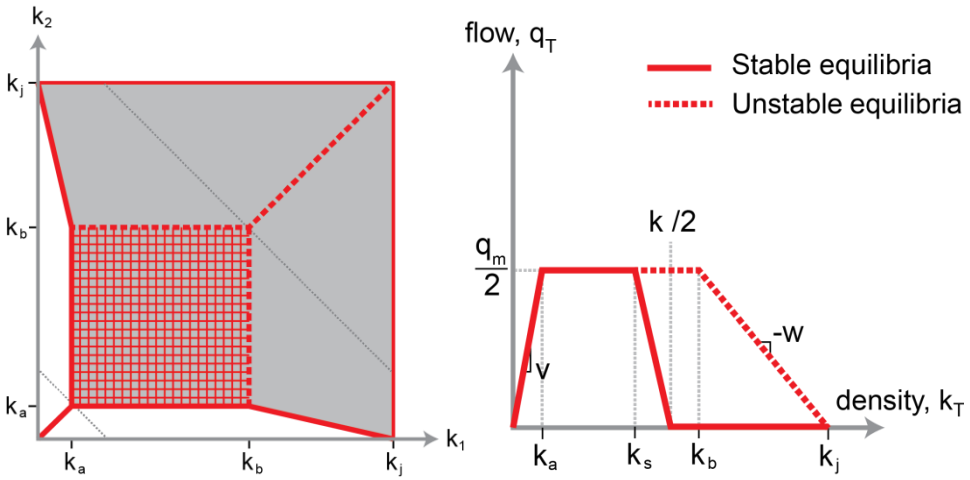


Figure 4-5. Phase diagram and aggregate flow-density relationship when signals have fixed timings ($\gamma = 0$).

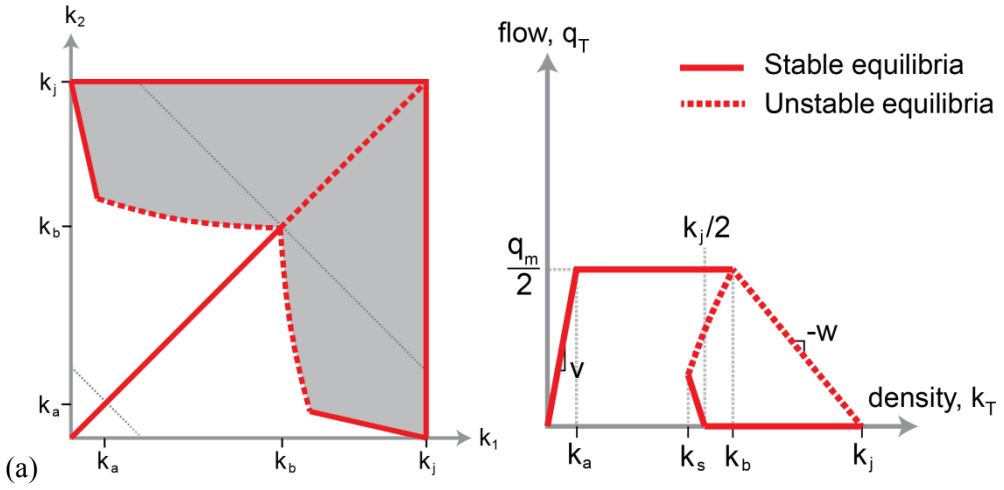
In the phase diagram, the dark solid lines denote the stable equilibrium states, while the dark dotted lines denote the unstable equilibrium states. The hatched area represents a third type of equilibrium state that only exists in signalized networks with fixed signal timings. We call these states quasi-stable states because minor perturbations will not grow or shrink with time. Instead, after a perturbation the system will remain at the perturbed state until another change occurs.

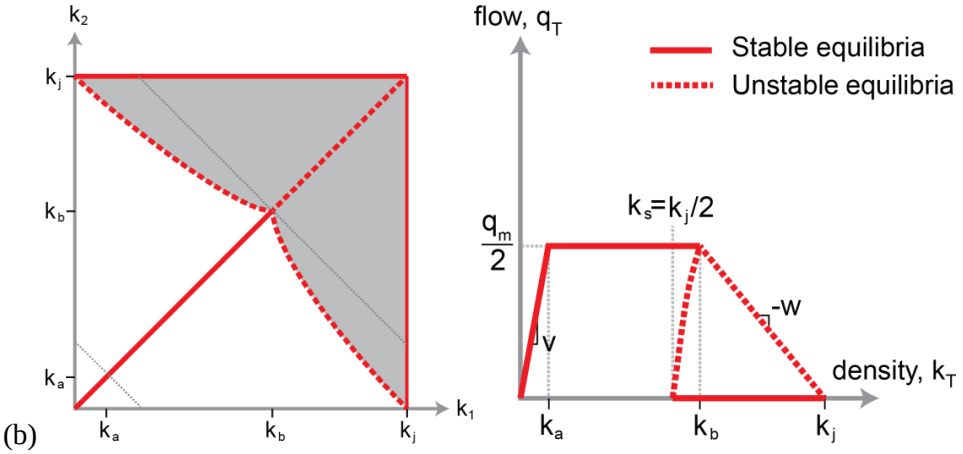
Notice from Figure 4-5 that an even distribution of vehicles (i.e., the same density or number of vehicles in each of the two bins) is only sustainable when there are very few vehicles in the network ($k_T < k_a$). As the density increases, traffic in the two-bin system naturally tends towards inhomogeneous vehicle distributions in which one bin has more vehicles than the other. The range of initial states for which the network will tend towards further imbalance is shaded in Figure 4-5. The area of this shaded region provides an indication of the ability of the network to maintain balance: larger regions indicate that congestion is more likely to distribute itself unevenly. Another index of stability is the density at which unstable equilibrium states emerge, k_S; see Figure 4-5. This is sometimes referred to as the bifurcation density because for densities greater than this value, multiple equilibrium states start to arise. For fixed signals, $k_S = (k_a + k_b)/2$. The flow-density relationship of the entire two-bin system associated with the equilibrium states (i.e., the two-bin system's MFD) is also presented in Figure 4-5. What is very interesting to note is that the uneven vehicle distributions at higher densities result in lower flows than would be expected if vehicles were evenly distributed across the two bins. These uneven distributions can even result in zero flow at densities lower than jam density, which is also known as gridlock.

We now examine how the set of equilibrium states changes when adaptive signal control is applied in the network and analyze how this influences the aggregate behavior of the network (i.e., MFD itself). Figure 4-6a and 4-6b presents the phase diagrams and aggregate flow-density relationships for partially adaptive signals ($0 < \gamma < 1$) and fully adaptive signals ($\gamma = 1$), respectively. Compared with the Figure 4-5, the changes in phase diagrams after the adaptive

signals are applied only occur in the region between the two iso-density contours shown in Figure 4-5 and Figure 4-6. Correspondingly, changes in the aggregate flow-density relationship would be observed in the density range $[k_a, k_b]$.

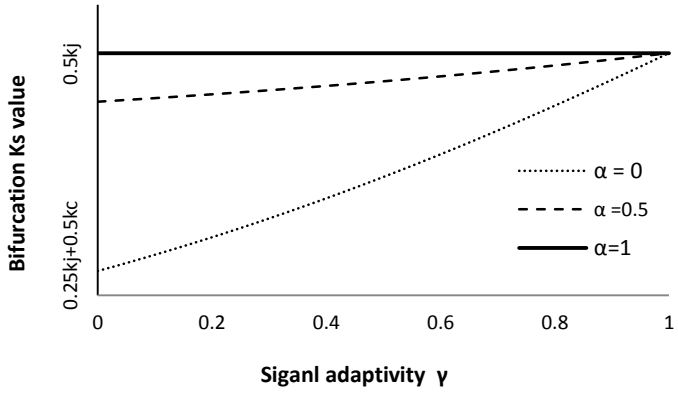
When adaptive signals are applied, the quasi-stable states that existed when $\gamma = 0$ disappear. Additionally, a balanced system with traffic evenly distributed across the two bins is maintained for a wider range of densities, $[0,\ k_b]$. This is compared to the range of densities that results in a balanced system with fixed traffic signals, $[0,\ k_a]$, with $k_a < k_b$. Stability of the network also increases with γ. E.g., as γ increases, the bifurcation density k_S increases and the set of states that lead to further imbalance (i.e., the shaded region on the phase diagrams) shrinks, which are both clear in Figure 4-7.This stronger push towards more balanced states results in higher average network flows for a given density as exhibited by the flow-density relationships presented in Figure 4-6. However, the improvement in the MFD plot is limited and occurs only in the range $k_T \in [k_a, k_b]$.



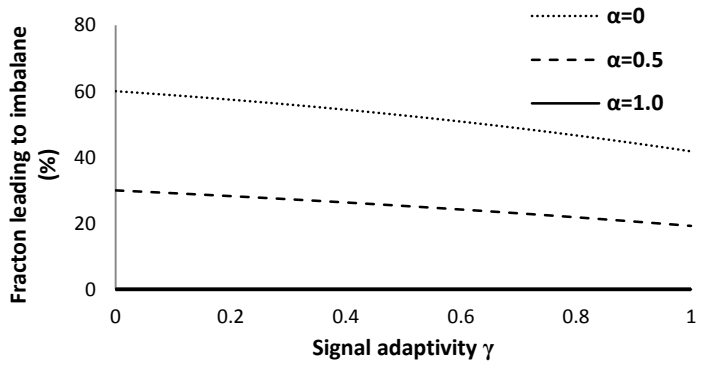


(b)

Figure 4-6. Phase diagram and aggregate flow-density relationship for: (a) partially adaptive signals ($0 < \gamma < 1$); and (b) fully adaptive signals ($\gamma = 1$).



(a)



(b)

Figure 4-7. Measures of stability in the two-bin network: (a) bifurcation density; and (b) fraction of states leading to network imbalance.

Notice that the set of equilibrium states for $k_T > k_b$ is invariant to γ. Thus, the network will still tend to unevenness for high densities regardless of the adaptive signal scheme used, even when the signals are as adaptive as possible. This occurs because when both bins are sufficiently congested, the circulating flows in each are governed by the congested branch of each bin's MFD (i.e., the decreasing portion of $q(k_i)$), and this was assumed to be more or less independent of the green proportion g_i.

4.2.3 Stability of Adaptive Drivers Routing Only

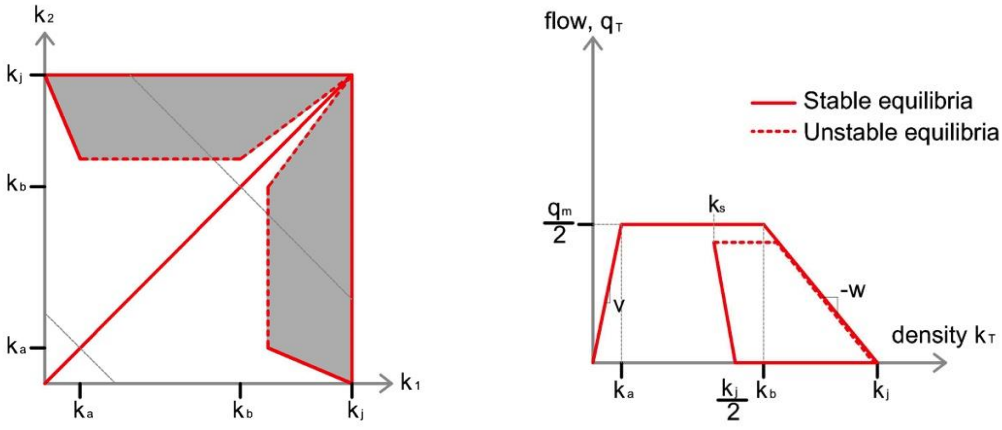
Compared with the adaptive signal control strategies that affect driver behavior by adjusting signal timings, adaptive drivers could spontaneously avoid the additional congestion and might show different impacts. We now apply the two-bin model to study how adaptive drivers might change equilibrium behavior in the network (when traffic signals are fixed).

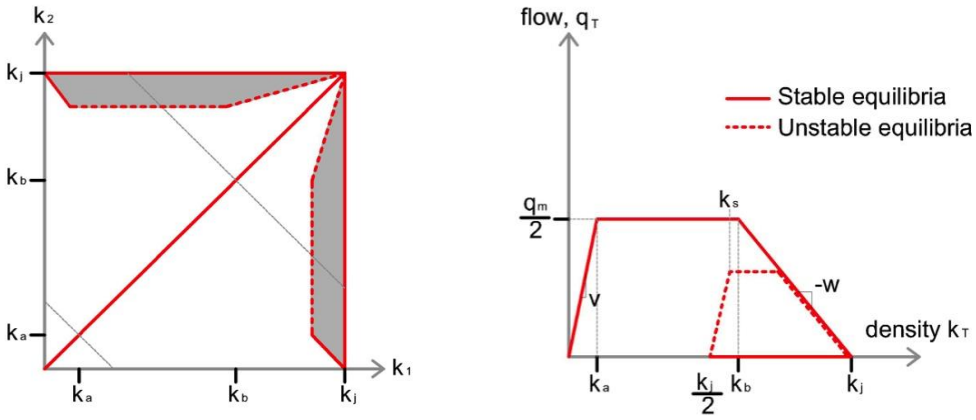
Partially adaptive drivers (i.e., $\propto \in (0, 1]$) are considered in the two-bin model to get the phase diagram and aggregate flow-density relationship shown in Figure 4-9. Similar to the effects of adaptive signals, the quasi-stable states shown in the phase diagrams with the fixed signal timing in Figure 4-5 disappear and are replaced with the balanced traffic states, see Figure 4-8. The adaptive drivers, similar as adaptive signals, are able to adjust the traffic states in the quasi-stable states to reach the evenly distributed traffic density and maintain the more stable urban traffic. For example, under fixed signal control, the traffic state in the quasi-stable areas might change due to minor perturbation, and cumulative minor perturbation might lead the traffic state leaving the quasi-stable areas and fall into the unstable states. With the adaptive drivers or adaptive signals, the same traffic state would be pushed near the stable equilibrium and only large enough disturbances fail to lead the urban traffic to leave the stable states. And those large disturbances occur with low probability. Moreover, the stability of the urban network increase with the level of the adaptive

drivers. Notice in Figure 4-8, the phase diagram with a high level of adaptive drivers demonstrates the small shaded areas, and the MFD with the high value of k_s. And these two indexes of traffic instability with different level of adaptive drivers ($\alpha = 0, \alpha = 0.5, \alpha = 1$) are also listed in Figure 4-7.

Different from the adaptive signals, the stable equilibrium traffic states under the adaptive drivers appear in the higher traffic density, which means that adaptive drivers still work at higher traffic density. Adaptive signals could not affect driver behavior by changing green times under the high average density, because those behaviors are limited by the traffic congestion instead of green times. While under the same high density, adaptive drivers also could mitigate the uneven distribution of vehicles. It explains why the stable equilibrium is possible under the high densities.

(a)





(b)

Figure 4-8. Phase diagram and aggregate flow-density relationship for low level of adaptive drivers ($\alpha < \alpha_c$) and high level of adaptive drivers ($\alpha > \alpha_c$) under fixed signal control strategy ($\gamma = 0$).

Assume that all the drivers are adaptive. Then unstable equilibrium states disappear and initial traffic state would reach to the strictly evenly distributed traffic states with the highest traffic flow under this certain network traffic density, shown in Figure 4-7(the fraction of imbalance area is 0 under $\alpha = 1$). Unfortunately, totally adaptive drivers are not possible in the real traffic network. And the effects of adaptive drivers, which are derived in the two-bin model, should be tested in more complex networks.

4.2.4 Stability with Adaptive Signals and Drivers

Adaptive signals could adjust the traffic states into the stable equilibrium states in a small range of network density $[k_a, k_b]$. For adaptive drivers, the set of equilibrium states changes throughout the entire congested region of the phase diagram. When adaptive signal control and adaptive drivers are combined, the benefits of each strategy are amplified. The set of stable equilibrium states would go throughout the entire traffic density under adaptive drivers and would expand the area under the adaptive signals.

However, the additional benefits obtained by combining the two strategies are only experienced with in the range $k_a < k_T < k_b$. The magnitude of the benefits to the bifurcation density and fraction of states leading to imbalance are also presented in Figure 4-7. Notice that the metrics improve significantly more as α increases than as γ increased. This is further evidence that suggests the adaptive driving is more beneficial than adaptive signal control at providing stability, although both strategies are beneficial.

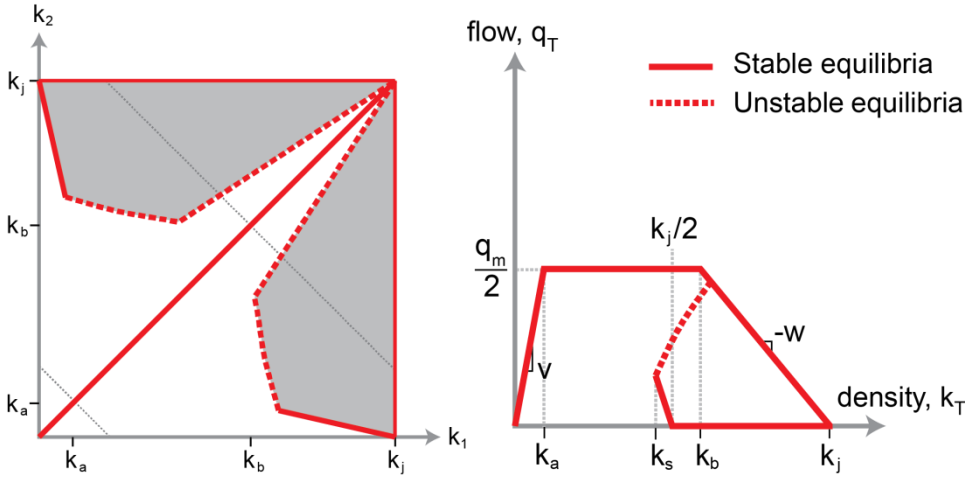


Figure 4-9. Phase diagram and aggregate flow-density relationship for partially adaptive signals ($0 < \gamma < 1$) and partially adaptive drivers ($0 < \alpha < 1$).

4.3 Two-Ring Simulation

4.3.1 Two-Ring Simulation Model

The two-bin model results suggest that adaptive signal control provides a stabilizing influence on network traffic, but only for a narrow range of densities representing moderate congestion levels. It also suggests that adaptive driver routing is more robust than adaptive signal control since it provides a larger benefit and does so for a wider range of densities. However, this

analytical model does not consider the impacts of vehicular dynamics along the ring and potential queue spillbacks at intersections. To verify that the same insights and behaviors hold when these impacts are included, simulations of the two-ring network abstraction are now used.

A two-ring network similar to the one proposed here was previously examined using an interactive simulation in Gayah and Daganzo (2011b). In this simulation, traffic on all links is modeled using a cellular automata model consistent with kinematic wave theory assuming a triangular fundamental diagram (Nagel and Schreckenberg, 1992; Daganzo, 2006). The simulation uses the following parameters: $v = 60\ mi/hr$, $w = -15\ mi/hr$, $q_m = 1800\ veh/hr$ and $k_j = 150\ veh/mi$ (equivalent to 60 vehicles per 0.4 mi ring). As the simulation runs, 1-min averages of network flow and density are calculated using the generalized definitions of Edie (1963).

This previous simulation considers only traffic signals with fixed timings. We have expanded this simulation to include the adaptive signal control strategy. The parameter γ is included as an interactive variable, labeled proportion of cycle adaptively allocated. The updated version of the two-ring simulation is available online at the following address: *http://www.engr.psu.edu/gayah/two_ring_adapt/two_ring_adapt.html*, shown in Figure 4-10. The reader is encouraged to interact with the simulation to verify the results that are presented in the remainder of this section. First, the behavior without adaptive signals is discussed as a base case, then the behavior with adaptive signals and the combination of adaptive signals and drivers is examined.

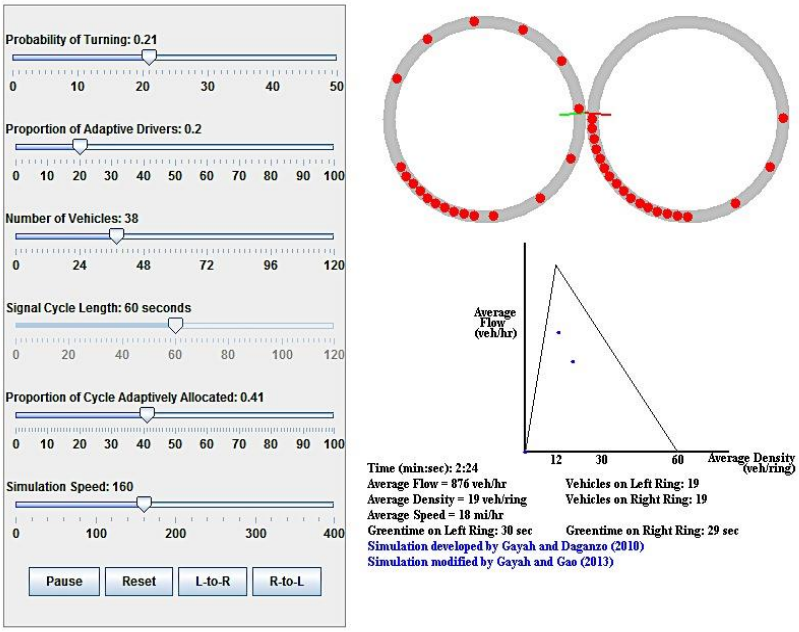


Figure 4-10. Online two-ring simulation with multiply input parameters.

4.3.2 Stability without Adaptive Signals Control

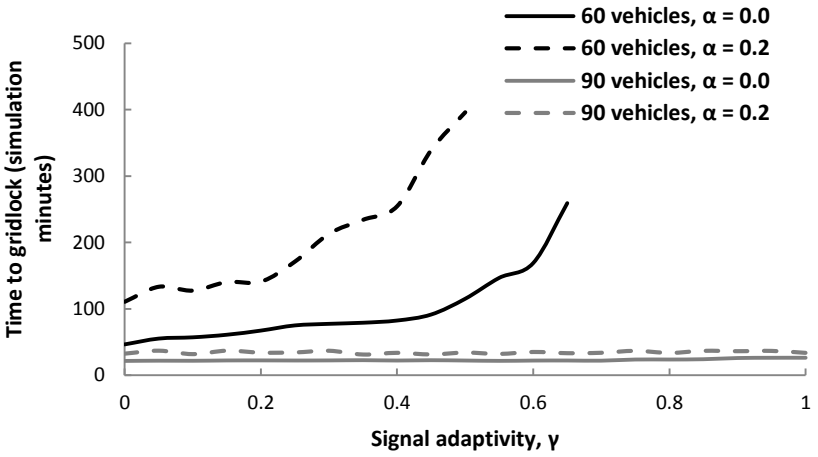
Simulations of the original two-ring network with fixed signal timing confirm that instability exists in the congested network that causes vehicles to distribute themselves unevenly across space. If enough vehicles are present to completely fill one of the rings, the network will eventually become so unbalanced that it will gridlock (defined as when the average flow in the network reaches 0 veh/hr. To verify this, try the following steps with the online simulation: reset the simulation; set the probability of turning to 0.2, number of vehicles to 60 (enough to completely fill one ring so that $k_T = 0.5\ k_j$) and proportion of cycle adaptive allocated to 0.0; and, start the simulation. The network quickly moves towards an unbalanced state where one ring has more vehicles than the other. Eventually, the rings become so unbalanced that one ring completely fills with vehicles while the other remains empty. At this point, the average flow becomes zero representing that the system is in complete gridlock.

4.3.3 Stability with Adaptive Signals Control Only

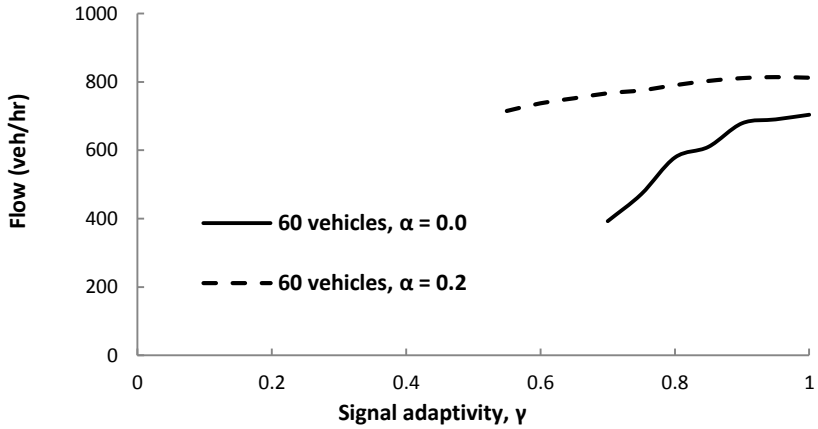
When the network is moderately congested (e.g., there are only enough vehicles to fill just one ring), the two-bin model suggests that adaptive signal control will help mitigate this innate instability. This should result in a reduced tendency towards gridlock, more homogeneous vehicle distributions and higher sustained flows in the network. To verify that this occurs, try the following: reset the simulation; set the probability of turning to 0.2, the number of vehicles to 60 and proportion of cycle adaptively allocated to 1.0; and, start the simulation. In this case, vehicles tend to remain fairly evenly distributed across the two rings. When stochastic fluctuations cause the system to become unbalanced, providing more green time to the more congested ring allows vehicles to exit this ring at a higher rate than at which they enter. This helps to move the system back towards a balanced state and to avoid gridlock.

The ability of the signals to mitigate the instability is related to the adaptivity parameter, γ. If signals are not very adaptive, the instability can dominate and gridlock may still occur. This can be observed by repeating the previous experiment but setting the proportion of cycle adaptively allocated to 0.2. However, even when lower signal adaptivity is unable to prevent gridlock in the moderately congested network, adaptive signal control still provides a benefit by delaying its occurrence. This can be seen in Figure 4-11a, which plots the average time required to reach gridlock over many simulation runs for different values of γ along with the 95% confidence interval of the mean. When 60 vehicles are in the network, the time to gridlock increases with signal adaptivity. Gridlock is mostly avoided after the level of signal adaptivity increases beyond a critical value (around $\gamma = 0.65$ for this battery of tests). For $\gamma > 0.65$, increasing γ provides an additional benefit by maintaining higher average flows; see Figure 4-11b, which presents the average flow measured over a period of 300 min across many simulation runs for the range of γ at which the network does not gridlock. Adaptive signal control can sustain a flow as high as 850 veh/hr when

60 vehicles are in the network, compared to a long run flow of zero that would occur when adaptive signals are not implemented.



(a)



(b)

Figure 4-11. (a) Average times to gridlock in the two-ring simulation; (b) average flow measured over 600 simulation minutes when the network does not gridlock.

The two-bin model also suggests that adaptive signal control will not provide any significant stabilizing influence when the network is heavily congested. To verify this, try the following: reset the simulation; set the probability of turning to 0.2, the number of vehicles to 90, and the proportion of cycle adaptively allocated to 1.0; and, start the simulation. In this scenario,

gridlock still occurs even though the entire cycle is dynamically allocated to the two competing approaches. Further examination of the green time provided per cycle (displayed in the bottom right-hand side of the simulation interface) shows that signals timings are not as flexible when the network is very congested. The presence of many vehicles in the less congested ring requires that some minimum amount of green time always be allocated to this ring, reducing the green time benefits experienced by the more congested ring. For example, in this case the less congested ring requires at least 20 s of green time even when the other is completely filled. The simulation also shows that flows become constrained by queue spillbacks at the intersection when the network is very congested. The reader could verify that gridlock still occurs with fully adaptive signals even when the number of vehicles is as low as 70, although gridlock might take a long time to occur. At these high levels of congestion, the time to gridlock is also insensitive to signal adaptivity; see Figure 4-11a. This verifies that adaptive signal control provides no stabilizing influence when the network is very congested.

The network-wide flow-density relationships of the two-ring system confirm this behavior. As illustrated in Figure 4-12a, larger values of γ are associated with MFDs that exhibit higher average flows for densities between 50 and 90 veh/mi. This demonstrates the additional stability provided by the adaptive signals in this region. However, for densities greater than 90 veh/mi, the adaptive signals have no impact on the MFD, as expected from the previous analytical results.

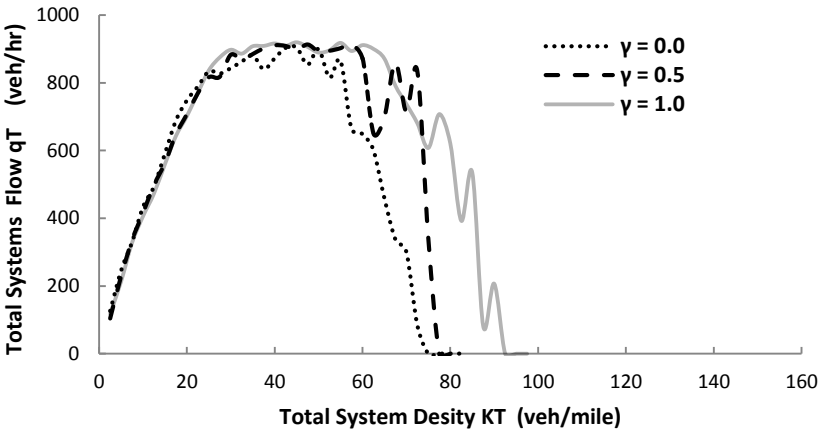
4.3.4 Stability with Adaptive Drivers Routing Only

As suggested by the analytical model, adaptive drivers routing is inherently different than adaptive signal control because adaptive drivers provides a stabilizing influence for the entire range of congested states. The reader can verify that a very high proportion of adaptive drivers (e.g., $\alpha = 0.8$) can prevent gridlock even when the network has 90 vehicles and is very congested.

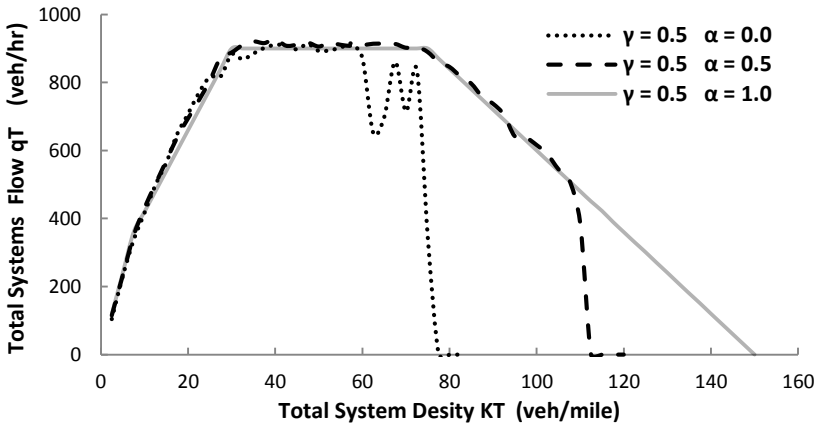
For a lower proportion of adaptive drivers (e.g., $\alpha = 0.2$), gridlock cannot be prevented. However, for cases with low levels of adaptive driving, additional stability is still provided, as the network is able to delay gridlock for a longer period of time and sustain higher average flows as illustrated in Figure 4-11. Thus, the simulation confirms that adaptive driver routing can provide stability even in cases where adaptive signal control cannot.

4.3.5 Stability Test with Adaptive Signals and Drivers

The combination of adaptive signals and adaptive drivers provides additional stability when the network is moderately congested. Once again, this behavior relates to behavior of the MFD. Figure 4-12 presents the MFDs obtained from the two-ring simulation when both signals and drivers are adaptive. The average traffic flow under moderately congested traffic increases with the existence of adaptive driver routing. Notice that the positive flows are maintained for a much wider range of densities than the case of adaptive signals alone, which is consistent with the results shown in Figure 4-11b.



(a)



(b)

Figure 4-12. network-wide flow-density relationship obtained from many iterations of the two-ring simulation when: (a) signals are adaptive; and, (b) signal are adaptive and drivers are adaptive.

4.4 Grid Network Simulation

Analysis of the two-bin and two-ring network abstractions reveals insights as to how adaptive signal control and adaptive drivers might influence the MFDs of urban traffic networks. However, the two-bin and two-ring network abstractions are built under some assumptions. To confirm the same patterns and behaviors also persist in more realistic network conditions, micro-simulations of the grid network with adaptive signals control and adaptive drivers were performed in the AIMSUN micro-simulation environment.

4.4.1 Idealized Grid Network Description

The 6 by 6 grid network with periodic boundary conditions, which is described in section 4.1.1, was created in the micro-simulation, shown in Figure 4-1a. Each link between the intersections is one-way with 800ft length.

A fixed cycle length of 60 s was selected with zero offset between adjacent signals. Since the adaptive signal control strategy proposed here was not native to the AIMSUN environment, it

was manually coded into the software using the Application Programming Interface (API). The traffic data on each link is collected and updated at every cycle length. And the green timing ratio for two competing directions at each intersection is function with the traffic density of two competing directions and the signal adaptivity γ, according to the Equation 4-4.

To simplify the simulation, we assume the initial vehicle distribution is homogenous: at initial time of the simulation, the same vehicle number would queue at each intersection. Then, the adaptive traffic signals are activated.

The simulation also utilized the periodic boundary condition described in section 4.1.1: when one vehicle exists at the end-most link, there is another vehicle entering from the up-most link spontaneously. Therefore, the density (vehicle number of the urban network) remained the same with possible multiple values of traffic flow.

4.4.2 Stability in Idealized Network

The resulting flow-density relationships for a long period of time and over many simulation runs are presented in Figure 4-13 for multiple values of γ. The MFDs shown in Figure 4-13 confirm the same general patterns as predicted from the simplified models: namely, that adaptive signal control improves network performance for a small range of densities representing moderate congestion. The range of densities for which large positive flows can be sustained increases with the level of adaptivity, as expected. However, the adaptive signals have no impact when the network is heavily congested (i.e., for densities greater than 150 veh/mi).

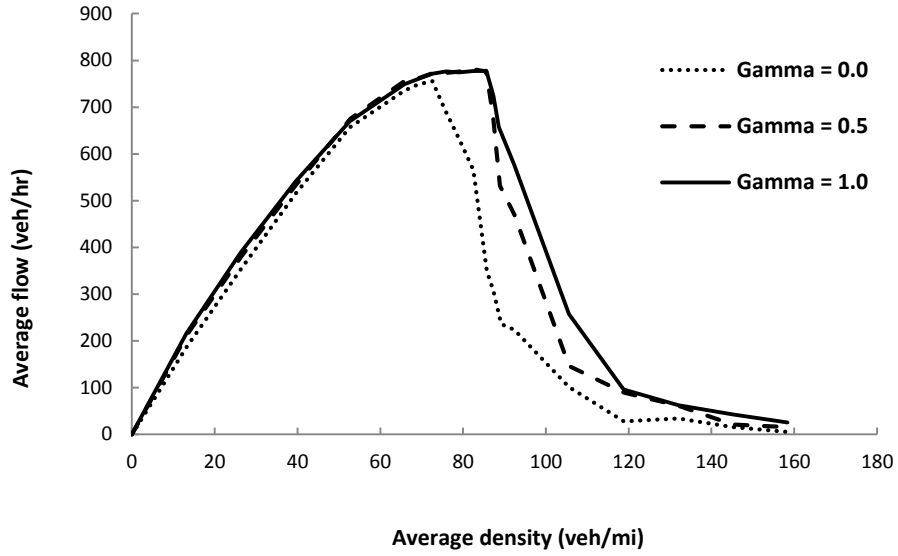


Figure 4-13. MFDs obtained from micro-simulation of a grid network with adaptive signals.

Another benefit brought with adaptive signals is longer jam times: see Figure 4-14. With the higher densities, it costs the urban network short times to turn gridlock. Fortunately, the short jam times of the urban network could be mitigated with some level of signal adaptively. Therefore, when we could extend the urban network jam time over the rush hours, the severe congestion could be avoided or we could reduce the probability of the gridlock.

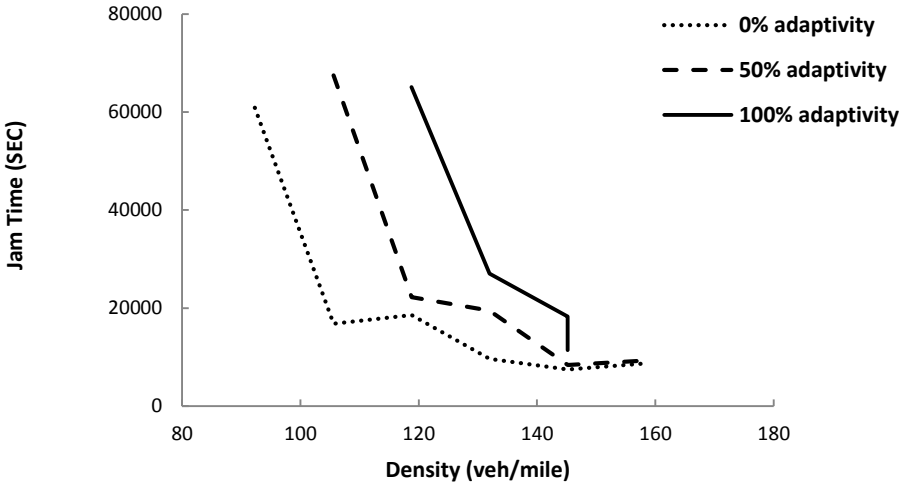


Figure 4-14. Jam Time obtained from micro-simulation of a grid network with adaptive signals.

4.4.3 Stability with a More Realistic Network

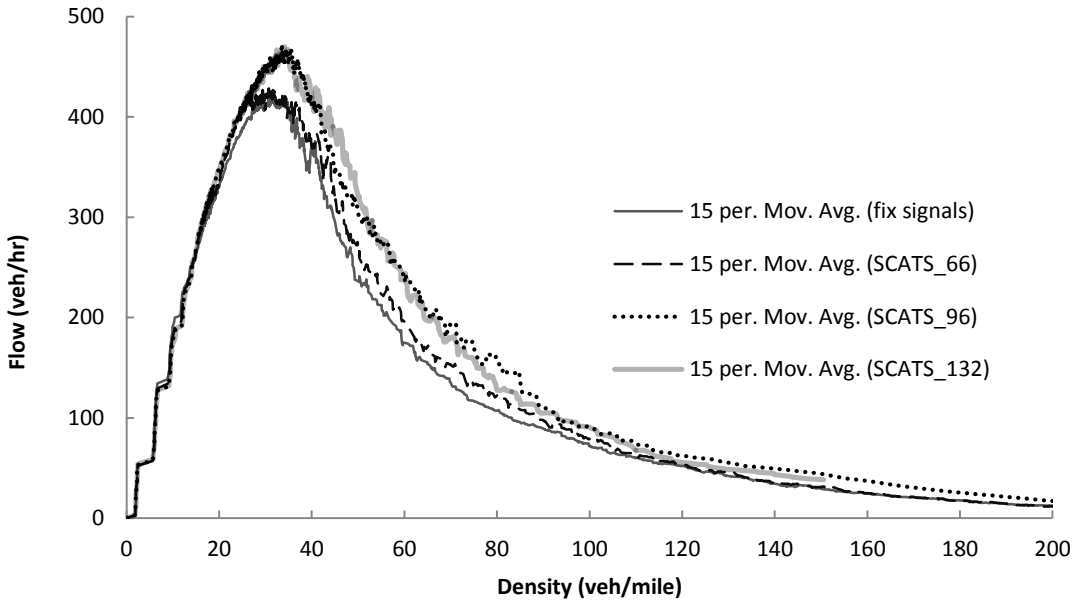


Figure 4-15. MFDs obtained from micro-simulation of a grid network when employing the SCATS algorithm.

The following simulations are performed on a more realistic network that relaxes the assumption of periodic boundary conditions. Two other realistic extensions are also considered: (1) vehicle entries and exits; and, (2) a more realistic adaptive traffic signal control scheme.

A 10 by 10 grid network was created with 800ft links and origins and destinations at all intersections and the upstream/downstream end of all entry/exit links. A uniform traffic demand pattern was assumed in which demands at all origins were the same and all destinations were assumed to be equally likely.

Furthermore, signals were assumed to operate using a modified version of the Sydney Coordinated Adaptive Traffic System (SCATS) algorithm at each of the traffic signals within the

grid network.[3] This algorithm works by adjusting both the green time allocated to each competing intersection approach and the total cycle length based on data obtained from detectors on the upstream approaches. The maximum cycle length is used as a measure of the signal adaptivity: higher maximum cycle lengths provide signal timings that are more flexible.

The MFDs for different maximum cycle lengths in the SCATS algorithm and for a network simulated with fixed signal timings are displayed in Figure 4-15. Notice that the same pattern observed in Figure 4-13 is repeated in Figure 4-15: the range of densities for which positive flows are observed increases as the maximum cycle length increases. And, as shown in the previous sections, the more flexible schemes help to mitigate congestion inhomogeneity and provide higher average network flows.

Some realistic micro-simulations of adaptive driver routing have already been explored in other studies; see Mahmassani and Saberi (2013) and Saberi et al. (2014) for more details. Therefore, the micro-simulation tests here focus only on the impacts of adaptive signal control.

[3] This algorithm is the SCATS-F algorithm as described in Zhang et al. (2013).

Chapter 5

Illustration of New Traffic Dynamics with Uncertainty

In this section, the evolutionary dynamics of urban traffic network is studies and analyzed while accounting for unavoidable uncertain phenomena that might exist. A metering control strategy is applied to an idealized scenario to illustrate some changes in network behavior that might occur when aggregate-level properties are uncertain.

5.1 Control Strategy Selection

Metering strategies maximize the production of the urban network by limiting the rate at which vehicles are allowed to access the network. For those strategies, various alternative algorithms have been developed to determine or update the metering rate. The metering strategies proposed in the literature review are dynamic feedback strategies that constantly modify the metering rate based on knowledge of both the MFD/NEF and real-time traffic conditions on the network. However, in practice, such real-time information might not always be readily available. There are significant challenges with collecting, storing and aggregating traffic data, typically obtained from detectors, in real-time that must be addressed before these dynamic metering strategies may be implemented. Although other types of dynamic feedback control have been implemented in many practical settings (e.g., feedback-based ramp metering control as in Papageorgiou et al., 2010), these typically rely on a relatively limited amount of locally obtained traffic data. Expanding this to network-wide traffic state estimations represents a potentially significant hurdle that must be overcome before dynamic metering can be implemented. To alleviate this concern, Keyvan-Ekbatani et al. (2013a) has explored dynamic metering control based on reduced MFDs that rely on information from just a subset of detectors within the network.

However, significant uncertainty may exist with reduced MFDs (Ortigosa et al., 2014) and detectors provide errors in MFD estimations (Leclercq et al., 2014). Mobile GPS data is another source that might be used to overcome these challenges; however, estimations made using data from probe vehicles will be associated with significant uncertainty until the penetration rate of probe vehicles is high enough (Gayah and Dixit, 2013; Nagle and Gayah, 2014; Leclercq et al., 2014).

To get around these significant barriers to implementing dynamic metering in practice, it might be more practical to first consider a static metering strategy in which the rate vehicles are allowed to access the network is fixed and set a priori. Such a strategy was tested in a micro-simulation network and shown to be effective at reducing congestion and improving network performance (Geroliminis and Daganzo, 2007). As can be expected, the static strategy would never be as effective as a dynamic feedback strategy (as the latter offers much more flexibility than the former); however, a static strategy would be easier to implement and might serve as a stepping stone for the eventual implementation of dynamic metering in the field. Unfortunately, when compared to their dynamic counterparts, static metering strategies have received much less attention in the literature.

In light of this, a static metering strategy is selected as a more appropriate choice to model the large-scale uncertainty, which also includes the uncertainty that exists in the implementation of static metering. This uncertainty accounts for the fact that actual vehicle entries into the network are likely to exhibit stochastic fluctuations around the controlled rate due to randomness in driver behavior.

5.2 Problem Scenario of Single-Region Network

An idealized scenario is built to illustrate some changes in network behavior that might occur when aggregate-level properties are uncertain. This scenario provides some of the notation that will be used in the remainder of this paper as well as motivates the creation of a probabilistic framework to explicitly model this uncertainty.

A downtown network that can be represented by a single region is considered here. In this representation, traffic is assumed to be fairly homogeneously distributed throughout the network such that the average density on all links within the network is more or less the same. All demand is assumed to originate outside the downtown network and is destined for inside the network, as would be typical for a morning peak for vehicles entering a central business district.

Traffic within the downtown region is assumed to obey a well-defined MFD; i.e., a well-defined function relates average flow in the network with the average density of vehicles on all links. MFD can be used to derive the NEF under the assumption that the average trip length within the urban region is invariant with time. Under this assumption, NEF is simply a rescaled version of the MFD. We denote the NEF by $F(n)$ [veh/hr] to relate the rate trips which are completed and leave the network, Q [veh/hr] with the total accumulated number of vehicles inside the city n [veh].

The simplest NEF is triangular in shape with the properties shown in Figure 5-1a. The triangular NEF adopts two strictly linear functions to represent two distinct states: 1) free flow traffic state in which exiting traffic flow increases with vehicle number, and 2) congested state in which large vehicle accumulation limits the speed and traffic flow starts to decrease. Such triangular NEFs are representative of deterministic NEFs that can be estimated using analytical methods. Empirical traffic data (Geroliminis and Daganzo, 2008 and Buisson and Ladier, 2009) suggests that empirical traffic models have a non-linear form, see Figure 5-1b.

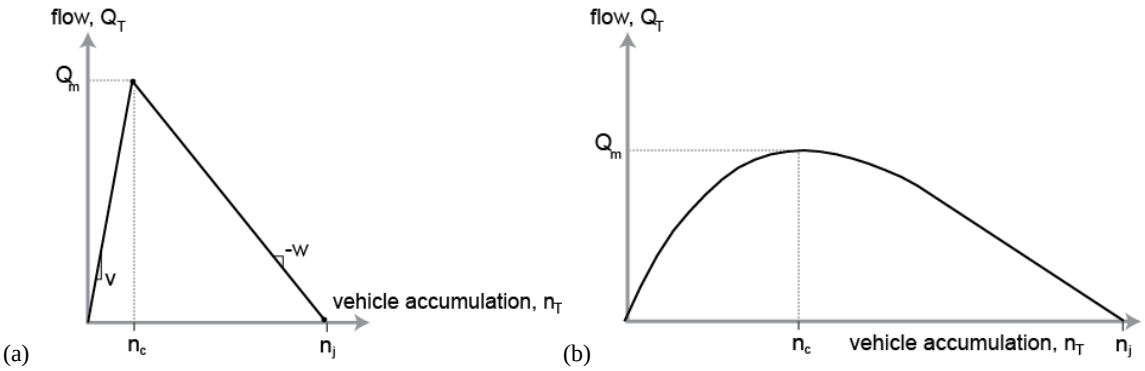


Figure 5-1. NEF models: (a) triangular NEF and (b) curved NEF model.

Let n_t denote the number of vehicles currently in the network at any time t. At the beginning of the peak period, the network is initially empty (n_0 = 0) and experiences some large unknown demand, A_t [veh/hr], for $t < T$, where T denotes the duration of the peak demand period. The demand is expected to exceed the capacity of the network (i.e., $A_t > Q_m$) during this time. In this scenario with large entering traffic flow, if vehicles are allowed to enter the network freely, congestion will occur, reducing the overall network efficiency and potentially leading to gridlock.

To prevent this congestion, and thus improve network efficiency, a metering flow strategy should be implemented in the downtown area, which could be achieved by pricing the drivers and signal timing design. A static perimeter metering strategy is implemented to restrict vehicle entry rate into the network to some value E. Thus, when vehicles are queued along the perimeter of the network, drivers will enter the network at this average rate, E to avoid the server congestion.

Stochastic driver behavior at each of the perimeter signals might cause actual entries measured at any t to fluctuate around E. For example, entry rate could exceed E if vehicles are aggressive and squeeze into the downtown network. In contrast, timid or inattentive drivers might have large headways causing the vehicle entry rate to fall below E. Therefore, the actual entry rate at any time t will be random with mean E and some standard deviation, G_0[veh/hr] as long as

perimeter queues exist. Different from the noise phenomena of the MFD/NEF model, the stochastic part is independent of traffic state. Furthermore, we assume that the stochastic fluctuations are independent across time. Under these conditions, a schematic representation of this downtown network system is provided in Figure 5-2.

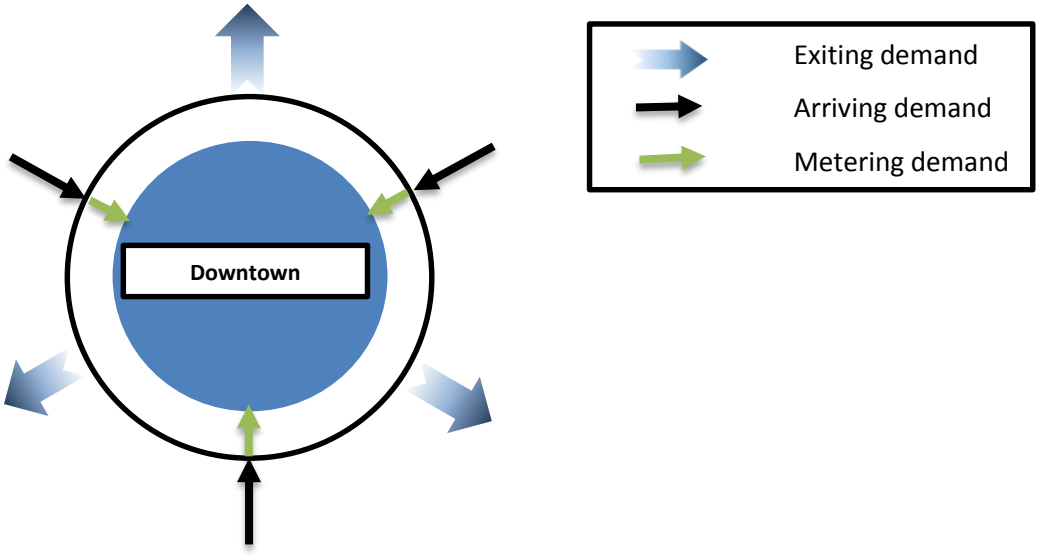


Figure 5-2. Visual representation of problem scenario in the single region.

The functional forms of the NEF presented here do not account for the uncertainty within the functions. However, uncertainty exists in NEF model as multiple values of exiting flow are typically observed for the same value of accumulation. In general, the deviation of those measures increases with traffic density, which means that the accuracy of the MFD reduces as the density exceeds the critical density. Even worse, some clockwise loop hysteresis might appear around the

capacity, see Figure 2-4. Those instable phenomena call into question MFD/NEF diagrams as an efficient tool.

Fortunately, some control strategies could improve the traffic model. Sufficient adaptive drivers and optimized adaptive signals can avoid the hysteresis loop and control the uncertainty of MFD model in minimum range. Here, we introduce a small range of uncertainty in the problem scenario. To emphasize in that the phenomena that uncertainty would increase with density, the deviation of traffic flow is assumed to be proportional to the vehicle number. Under those assumptions, the triangular and nonlinear NEF are assumed as Gaussian distribution with mean value of $f(n)$ and standard deviation $g(n)$, which are described in Equation 5-1 and 5-2, respectively.

$$F(n) = \begin{cases} f(n) = min\{vn, w(k_j - n)\} \\ + \\ g(n) = G_f n w_t \end{cases} \quad \textbf{(5-1)}$$

where w_t is a Gaussian white noise with 0 mean and 1 standard deviation. Therefore, the exiting flow according to the urban NEF model is a Gaussian distribution with mean value of $min\{vn, w(k_j - n)\}$ [veh/hr] and standard deviation of $G_f n$[veh/hr].

$$F(n) = \begin{cases} f(n) = \begin{cases} 2.28 \times 10^{-8} n^3 - 8.62 \times 10^{-4} n^2 + 9.58n, & 0 \le x < 14{,}000 \\ 27{,}731 - 1.38655(n - 14{,}000), & x \ge 14{,}000 \end{cases} \\ + \\ g(n) = G_f n w_t \end{cases} \quad \textbf{(5-2)}$$

where the capacity is $33{,}168$ [veh/hr] and jam density $k_j = 34{,}000$[veh].

The curved NEF is built with empirical data of Yokohama, Japan (Geroliminis and Daganzo 2008), which is described using Equation 5-2 and Figure 5-1b. The two NEF models would be utilized in the dynamic analysis and also be used to seek a probabilistic framework with accounting for the uncertainty in the next section.

5.3 Analysis under No Uncertainty ($G_0 = 0;\ G_f = 0$)

To evaluate the efficiency of the control strategies on the urban network, it is necessary to conduct the dynamic analysis and predict how the urban traffic responds. For simplicity, we start with the scenario without uncertainty in vehicle entries and in the NEF model ($G_0 = 0; G_f = 0$), which is relatively straightforward. The dynamic process for urban city with a single downtown is described by Equation 5-3.

$$\begin{cases} \frac{dn_t}{dt} = E - f(n_t) \\ n_0 = 0 \end{cases} \quad \textbf{(5-3)}$$

where n_t is vehicle number on the urban traffic network at the time point t, and E is the restricted entering flow a, and $f(n_t)$ is the function of NEF model, which depicts the relationship between existing flow and vehicle number.

The analytical solution of the dynamic differential equation is available with different forms of NEF functions, such as the triangular or polynomial functions shown in Figure 5-1. This solution provides the estimated traffic state at each moment. For the triangular NEF model with two different linear parts, the time point t_c when traffic state is in the critical vehicle accumulation is determined, and then the appropriate branch of NEF function could be used in the calculation. The estimated traffic state could be derived according to Equation 5-4 and 5-5.

When triangular MFD model is used and traffic state is in free flow state:

$$n_t = \frac{E}{v}[1 - e^{-vt}] \quad \textbf{(5-4)}$$

When E is large enough $E > Q_m$ and traffic state reaches Q_m at the time point t_c:

$$n_t = (n_j - \frac{E}{w}) + \left(n_c - n_j + \frac{E}{w}\right) e^{w(t-t_c)} \quad \textbf{(5-5)}$$

The curved NEF function is described with a cubic polynomial equation using the form $f(n) = a \cdot n + b \cdot n^2 + c \cdot n^3$. The dynamic process is described using Equation 5-6.

$$\frac{dn_t}{dt} = E - a \cdot n_t - b \cdot {n_t}^2 - c \cdot {n_t}^3 \quad \textbf{(5-6)}$$

This is the Abel equation of the first kind, which was known as an unsolvable equation. Panayotounakos and Zarmpoutis (2011) discovered an analytic method to solve the above equation generally. And Rostami (2015) presented an exact solution for the Abel Equation. It refers to a mathematical theory with large amounts of equations, which are not discussed in detail.

Intuitively, the optimal static metering strategy will seek to maintain the accumulation associated with the maximum exit flow (for the NEF assumed here, n_c, which is associated with maximum flow, Q_m) without allowing the network to enter the congested regime. If the network ever enters the congested regime it will continue to get more congested during the peak period when a static metering scheme is employed, as illustrated by the arrow on the right-hand side of Figure 5-3.

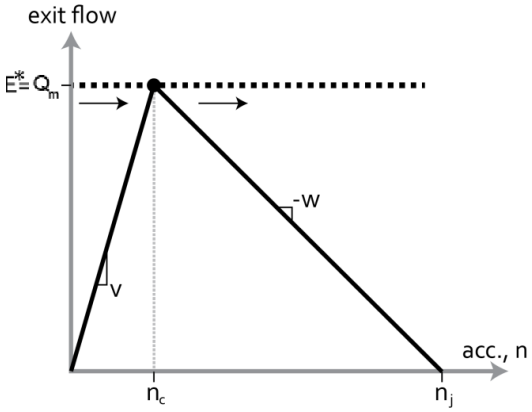


Figure 5-3. Network behavior with static metering strategy.

The maximum exit flow Q_m is desired because it is associated with the lower average vehicle delays and the highest network productivity. Thus, the network will initially start empty and tend towards a limiting state of n_c at $t \to \infty$ as illustrated by the arrow on the left-hand side of Figure 5-3. To ensure that the network never enters the congested regime (i.e., $n_t \leq n_c \ \forall t$), the parameter setting $E \leq Q_m$ should be guaranteed . Since the highest exit flow is desired, the most appropriate metering rate would be $E^* = Q_m$ to both avoid congestion and maximize efficiency.

5.4 Analysis with Uncertainty ($\boldsymbol{G_0 \neq 0}$; $\boldsymbol{G_f = 0}$)

Unfortunately, this optimal static metering strategy is not practical when stochastic phenomena exist in the dynamic process of the urban traffic. We first assume that there is uncertainty in vehicle entries, i.e. $G_0 \neq 0$. When the critical accumulation is near n_c, a large fluctuation in entering flow may cause the network to move into a congested state. Once congested, $f(n) < E^*$ and accumulation will tend to grow with time as the network tends towards gridlock; this is illustrated by the arrow on the right-hand side of Figure 5-3. *(Note that when $n_t = n_c$, random disturbances that cause the entry rate to be less than the maximum exit rate are easily*

handled and the network will tend back towards the limiting state of n_c.) Therefore, in practical situations where there is uncertainty in vehicle entries, a value of $E < Q_m$ should be selected if congestion is to be avoided.

5.4.1 Analytical Solution using Stochastic Differential Equation for Linear NEFs

When considering uncertainty in the input flow E, the stochastic evolution of vehicle accumulation over time could be described with the stochastic differential equation (SDE) using Equation 5-7:

$$\frac{dn_t}{dt} = E + G_0 w_t - f(n_t) \quad \textbf{(5-7)}$$

The general analytical solution for these SDEs do not exist. However, analytical solutions are well-known for SDEs that are linear. This would occur when the NEF $f(n_t)$ is a linear function and would not hold using the common polynomial functions nor triangular functions. The dynamic control process cannot be studied analytically under the simplest assumption of a triangular NEF, which consists of two different linear parts, since it is not a truly linear function. If the time point that the network switches from the free flow to the congested branch of the NEF, t_c, was known, then the linear SDE solutions could be brought to bear on the problem. However, this time point that decides which linear part of the triangular function is used in the analysis becomes a random variability when uncertainty is added to the problem and thus a simple analytical solution does not exist.

As a simplification, we can assume that the network always operates within the free-flow regime (which is typically desired). In this case:

$$\frac{dn_t}{dt} = E - vn_t + G_0 w_t \quad \textbf{(5-8)}$$

Linear stochastic differential equations in this form have been well-studied in the literature (e.g., Gelb, 1974). Analytical solutions exist for the special case in which w_t is treated as white Gaussian noise (i.e., w_t is randomly drawn at every instant from a normal distribution with mean 0 and standard deviation 1). The assumption of Gaussian noise may not be physically realistic since it implies that errors in entry rate are unbounded and can be unreasonably small or large. In practice, we would expect these errors to be bounded by some general physical constraints (e.g., a minimum and maximum number of vehicles entering the network during the signal green period). Therefore, we make this assumption to facilitate an analytical solution but will compare the behavior under the assumption of bounded errors using simulations in the following analysis.

Using the analytical solution provided by Gelb (1974), the expected accumulation at any time, μ_t, with the initial condition $n_0 = 0$ is:

$$\mu_t = \frac{E}{v}[1 - e^{-vt}] \quad \textbf{(5-9)}$$

Note that the evolution of the expected accumulation in the stochastic case is the same as the evolution of the accumulation in the deterministic case described with Equation 5-4. Therefore, uncertain vehicle entries will not change the evolution of expected accumulation (i.e., μ_t independent of G_0) as long the network remains within the free-flow regime. The network still tends toward a long-run accumulation of $\mu = E/v$ as $t \to \infty$.

What changes when $G_0 \neq 0$ is that there is some uncertainty associated with actual accumulations with respect to the expected accumulation. The solution of the stochastic differential

equation in Equation 5-10 shows that the actual accumulation will follow a normal distribution with a mean value described by Equation 5-9 and variance of:

$$\sigma_t^2 = \frac{{G_0}^2}{2v}[1 - e^{-2vt}] \quad \textbf{(5-10)}$$

This variance is initially zero (since the network starts out empty), increases with time, and is bounded by $\sigma^2 = {G_0}^2/2v$. The maximum variance is described by only two parameters: the magnitude of uncertainty in entry flows (G_0) and the slope of the free-flow branch of the NEF (v). Overall, the variance of actual accumulations increases as the entry flow uncertainty grows and as the slope of the free flow branch flattens.

The limiting values μ and σ provide the distribution of the accumulations that should be expected in the network by the end of the long peak period for a given E. Since these limiting values are upper bounds, these values can be used to estimate the largest possible accumulations that can be expected in the network at any point during the peak period. These values are only valid if the network remains in the free flow regime during the entirety of the peak period, which holds if the probability that the network accumulation ever exceeds n_c is extremely small (<0.001). An example is illustrated in Figure 5-4a, which presents a magnified portion of the NEF. Here the limiting distribution is shown by the thin black curve centered on an accumulation μ. Since the probability distribution does not extend past n_c, the network is not likely to become congested and the static metering rate selected is reasonable.

Notice if the distribution obtained by assuming the network is governed by the free flow branch of the NEF exceeds n_c by a significant amount, as illustrated in Figure 5-4b, then the network is likely to enter the congested regime at some point during the peak period. In this case, the distribution of potential accumulations provided by μ and σ will no longer hold. Instead, the

lower exit flows experienced in the congested regime would help push the network to more congested states, eventually leading to complete gridlock. Even if actual fluctuations in entry flows cause the network to return to the free flow state, the limiting distribution suggests that the network will eventually return to the congested regime until it eventually gridlocks.

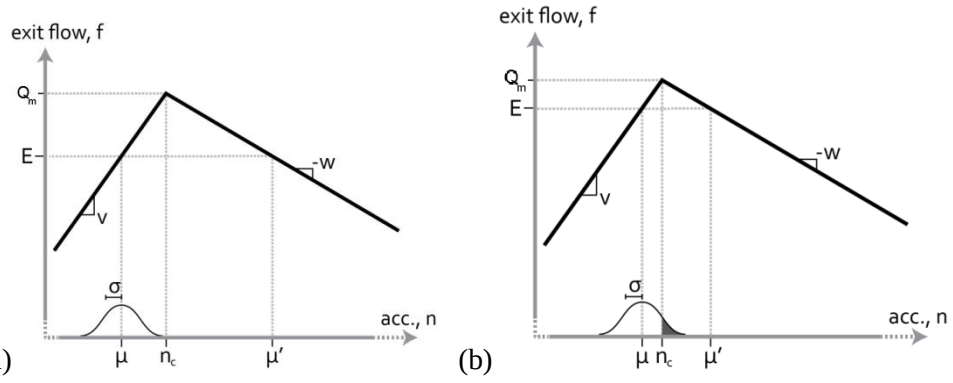


Figure 5-4. two examples of set metering rates: (a) a case where the network is not expected to enter the congested state; and, (b) a state in which the network is expected to become congested.

From this, and knowing that the highest entry rate is desired to maximize network efficiency, the optimal static metering rate to avoid congestion is:

$$E^* = v\left(n_c - Z\sqrt{\frac{{G_0}^2}{2v}}\right) \qquad \textbf{(5-11)}$$

where Z is the standard normal value associated with the desired probability of congestion.

A value of $Z = 2.58$ will result in a 1% chance of congestion. All values of $E > E^*$ will result in a larger probability that the network will become congested during the peak period.

5.4.2 Numerical Simulation Tests

A discrete-time simulation is used to emulate the behavior of this single-region downtown network when a static metering scheme is employed for comparison with the analytical formulae developed. In this simulation, the accumulation of vehicles in the downtown is updated at discrete time intervals of Δ_t [hr] << T. A value of $\Delta_t = 1$ min is selected, which might represent the length of the fixed-time signal cycle on the periphery of the network at which the static metering scheme is applied. The accumulation of vehicles is updated according to the discretized version of Equation 5-7:

$$n[t + \Delta t] = n[t] + E\Delta t - f(n[t])\Delta t + G_0{}'w[t]\Delta t \qquad \textbf{(5-12)}$$

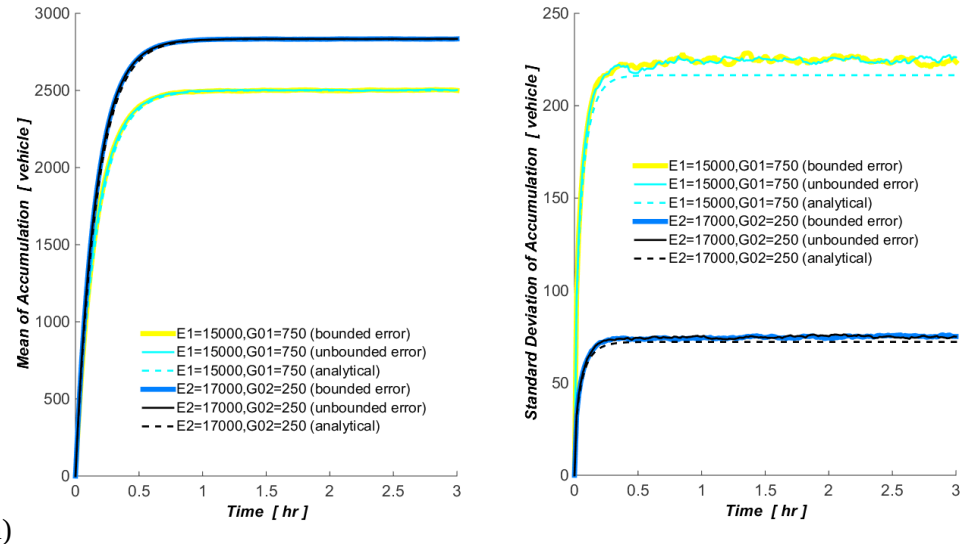
where the [t] notation represents the corresponding values in Equation 5-7 measured at the end of each discrete time interval. The primary difference between Equations 5-7 and 5-12 is that $G_0{}' \neq G_0$ since the former is measured in units of veh/time interval while the latter is veh/hr. Since errors in individual time intervals are assumed to be independent, the two values are related by $G_0 = G_0{}'\sqrt{\Delta t}$. Another difference is that the error terms, $w[t]$, are no longer restricted to unbounded white Gaussian noise. Instead, $w[t]$ can be drawn from any distribution with mean value 0 and variance.

We now use the simulation to verify that the behavior of the system is the same as that described by Equations 5-9 and 5-10 when the network always remains uncongested. In all of these simulations, the following parameters are assumed to describe the NEF of the downtown network: $v = 6$, $Q_m = 18{,}000$, $n_c = 3{,}000$, $w = 1.5$, $n_j = 15{,}000$ and $G_f = 0$. We first compare the results predicted from Equations 5-9 and 5-10 and the simulation results when the

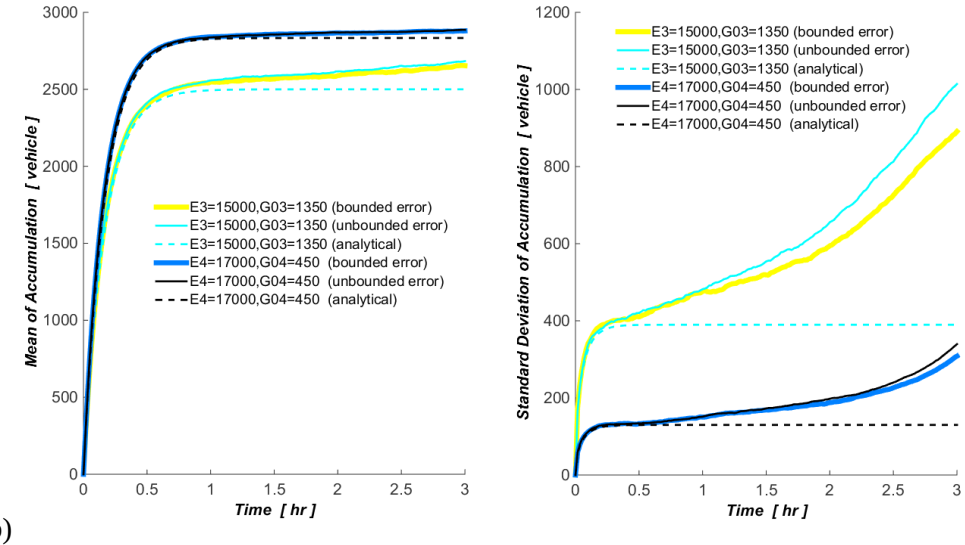
network never becomes congested. Then, we use the simulation to explore cases in which the network enters the congested regime when the static metering scheme is applied.

We first consider two representative pairs of E and G_0 values for which the probability of the network accumulation exceeding the critical accumulation obtained using the limiting distribution is small (around 1% in both cases): $E_1 = 15{,}000$ and $G_{01} = 750$, and $E_2 = 17{,}000$ and $G_{02} = 250$. (Note the values of E were obtained using Equation 5-11.) Since the network is unlikely to get congested for these two scenarios, Equations 5-9 and 5-10 should accurately describe the evolution of accumulation over time. These analytically obtained values for mean accumulation and standard deviation of accumulation are plotted as the dotted lines in Figure 5-5a.

The same values obtained from 10,000 simulation runs of each scenario are also plotted as solid lines in these figures. The different line thicknesses for the simulation results represent two types of errors simulated: the thin lines represent unbounded errors, while the thick lines represent bounded errors. For the unbounded errors, the error terms $w[t]$ were assumed to be drawn from a standard normal distribution, which faithfully replicates the white Gaussian noise used to derive the analytical equations. In the bounded case, $w[t]$ was drawn from a uniform distribution between $-\sqrt{3}$ and $\sqrt{3}$. (The bounds of the uniform distribution are selected so that the variance of $w[t]$ = 1.)



(a)



(b)

Figure 5-5. Evolution of network accumulation and variance of network accumulation under static metering scheme: (a) when network is expected to remain uncongested ($E_1 = 15{,}000$and $G_{01} = 750$, and $E_2 = 17{,}000$ and $G_{02} = 250$); and, (b) when network is not-expected to remained uncongested ($E_3 = 15000$, $G_{03} = 1350$ and $E_4 = 17000$, $G_{04} = 450$).

Both the mean accumulation and variance of accumulation predicted by the analytical equations are consistent with the simulation when unbounded errors that represent white Gaussian noise are used. This is expected since the analytical solutions rely on the assumption of Gaussian

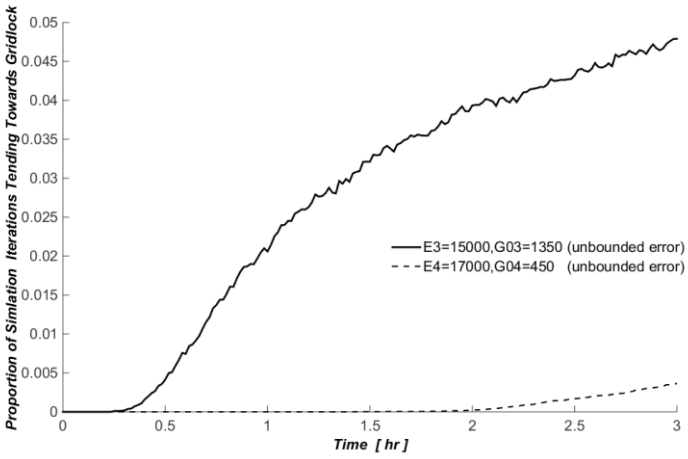
noise terms. The analytical results also accurately describe the evolution of the system in the more realistic case of bounded errors. This suggests that the analytical solution should be fairly robust and valid in a variety of more realistic situations. One reason for this is that even if the errors during short time periods are bounded (e.g., errors during a single signal cycle), the central limit suggests that the sum of the errors over a longer time period (e.g., an hour) can be approximated quite well by a normal random variable. Therefore, the assumption of unbounded Gaussian noise terms, while unrealistic physically, is okay from a modeling perspective. For the remainder of the simulation results in this paper, it is assumed that the noise terms are Gaussian distribution for simplicity and consistency with the analytical derivations.

We also consider two representative pairs of E and G_0 values for which there is significant probability that the network accumulation will exceed the critical accumulation when using the limiting distribution (about 10% in both cases): $E_3 = 15000$, $G_{03} = 1350$ and $E_4 = 17000$, $G_{04} = 450$. The analytically derived mean accumulation and variance of accumulation are plotted on Figure 5-5b as dotted lines while the simulation results for both types of error terms are also included as the solid lines. The simulations verify that the analytical equations do not perform well if the network is expected to become congested. In fact, the mean accumulation from the simulations exhibits significant and constant drift away from the analytically predicted means. This drift is even more noticeable in the plot of standard deviation of accumulation. Closer inspection reveals that this occurs because a significant portion of the 10,000 simulation iterations become congested and stay in the congested regime.

Therefore, we find that the analytical method works very well when the network is not expected to enter the congested regime and fails to accurately predict network performance when there is a significant probability of entering a congested state. Selecting $E = E^*$ should then ensure that the network does not enter a congested state that might not be well-defined and from which it might tend towards gridlock and not be able to recover.

5.4.3 Setting Optimal Metering Rate

In practice, selecting $E = E^*$ might be extremely conservative: a relatively small value of E must be chosen to avoid any significant risk of the network entering the congested regime and these will be associated with relatively low values of network productivity (measured by the average exit rate). Allowing vehicles to enter the network at a greater rate might result in higher network productivity for some time, but this will be associated with a greater change of the network entering the congested regime. Even if the congested branch of the MFD is well-defined, if the network becomes too congested it may still be susceptible to gridlock. This presents a trade-off between network productivity and the potential for catastrophic outcomes (i.e., gridlock).



(a)

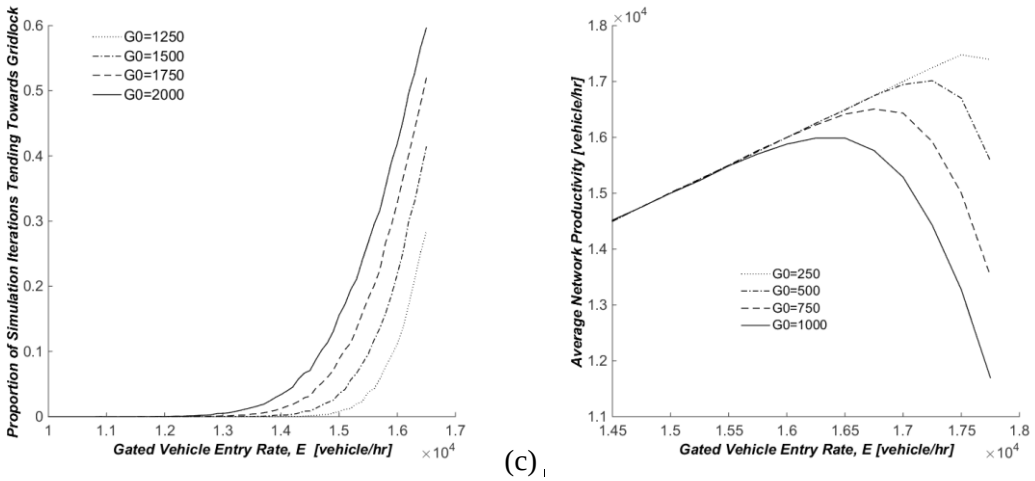


Figure 5-6. (a) Relationship between gridlock tendency and length of peak period; (b) trade-off between gated entry rate and gridlock tendency; and, (c) average network productivity vs. gated entry.

The simulation previously described is used to examine this trade-off in more detail. We first examine how a network's susceptibility to gridlock is a function of the length of the peak period. In this battery of tests, 10,000 simulation iterations are performed for the combinations of E and G_0 presented previously presented in Figure 5-5, which are known to have a limiting distribution that leads to congestion. Here, the susceptibility of the network to gridlock is qualified as the fraction of simulation iterations for which the network accumulation exceeds the congested accumulation associated with an exit flow equal to the current average entry flow, E. This critical value for a given inflow E is depicted visually in Figure 5-4 as the value μ'. When the network accumulation exceeds this value, the average exit flow will be less than the average entry flow and the network will experience a strong push towards a gridlocked state in which $n = n_j$. Figure 5-6a presents the proportion of simulation iterations that exhibit a tendency towards gridlock as a function of the length of the simulated peak period. As shown in the plot, this proportion generally increases with the length of the peak period. Thus, for peak periods that are very short, implementing $E = E^*$ might be overly restrictive as vehicles could enter at a larger rate with no

fear of congestion spiraling out of control. For larger peak periods, the probability that the network accumulation will exceed the critical value and exhibit a tendency towards gridlock is relatively small. For example, in the set of tests performed here, there is only a 5% and 0.5% chance that the network will be sufficiently congested as to be in danger of gridlock for the tests with E_3 and E_4, respectively. This suggests that the trade-off between average productivity and susceptibility to gridlock is a function of the length of the peak period considered.

This trade-off is presented more clearly in Figure 5-6b, which illustrates the relationship between the metered entry rate, uncertainty in entry rate and the proportion of simulation iterations that exhibit a tendency towards gridlock for a fixed peak period of 3 hours. As expected, the risk of the network becoming gridlocked increases with the metered entry rate and the level of uncertainty associated with vehicle entries. For a given G_0, the risk is relatively flat at lower values of E and increases rapidly at higher values of E. This suggests that large increases in metering rate near E^* result in a small risk of gridlock, but small increases can be disastrous when $E > E^*$. The entry rate in Figure 5-6b can be used as a surrogate for average network productivity when $E > E^*$. However, it does not fully capture the loss of productivity achieved when the network becomes congested and enters the gridlock process. To account for this, Figure 5-6c plots the average vehicle exit flow from the downtown network after the three hour period for various values of metered entry rate, E, and uncertainty in entry rate, G_0. The results show that average or expected productivity increases proportionally to entry rate at low E values. As E increases and the network begins to get congested, productivity increases at decreasing rates and eventually starts to decline. For any given G_0, there is a critical E exists that is associated with peak expected productivity. This value might the best balance between productivity and susceptibility to gridlock since this measure incorporates the ill-effects of congestion at the end of the peak period.

Although the dynamic analysis of urban control could provide insight on how to optimize a static metering strategy with the disturbance of the uncertain factor in the metering strategy to balance between productivity and susceptibility to gridlock, this method could not be widely applied in general traffic control strategies due to some limitations. First, the method is designed for the urban network that could be represented with a triangular NEF model. Unfortunately, the triangular NEF is an idealized model to simplify the dynamic analysis, which does not fit into the empirical model built with traffic data. Besides, the dynamic analysis only considers the uncertain factor in the metering flow, i.e. $G_0 \neq 0$, which is independent of urban traffic state. An uncertainty factor in the NEF model ($G_f = 0$) is ignored which is dependent on traffic state and certainly complicates the dynamic analysis. Furthermore, the analysis only focuses the traffic evolutionary dynamic in the free-flow regime, but does not account for the dynamic analysis in the congested regime. Therefore, to comprehensively depict the dynamic process, another method with wider application and flexibility is proposed.

Chapter 6

Markov Chain (MC) Model for a Single Region

6.1 MC Model Application

To overcome the limitations of SDE theory, a Markov Chain (MC) model is developed to describe network-wide traffic dynamics under uncertainty. A Markov Chain describes a random process that undergoes transitions from one state to another on a state space. It must possess a property that is usually characterized as "memorylessness": the probability distribution of the next state depends only on the current state and not on the sequence of events that preceded it. This specific kind of "memorylessness" is called the Markov property. The dynamic process of urban traffic control strategy possesses the Markov property of "memorylessness": vehicle accumulations at the next moment depends only on the current vehicle accumulation given some input flow and an NEF function. Therefore, the stochastic network-wide traffic dynamics should be able to be described by a MC model defined by three specific elements: a finite state space, transition matrix and the initial state.

In the previous section, the traffic state—i.e., network accumulation—is a continuous variable. Here, we discretize the traffic state into n "bins" with each of size Δn and indexed by i, see Figure 6-1a. In this way, we could build a finite state space ranging from bin 1 to bin n. Each bin i will be represented and modeled using its mean accumulation value, $\overline{n_i}$; i.e., it will be assumed that the network is at accumulation $\overline{n_i}$ if it exists in bin i regardless of the actual accumulation. Time will also be discretized into identical intervals of length Δt. Similarly, nonlinear NEF traffic models are redefined and approximated as a set of piece-wise linear functions; see Figure 6-1a. When the traffic state is represented by a particular bin, traffic dynamics at that current moment

can be described using the corresponding linear segment for that bin. In this way, the dynamic process, which is initially described with a nonlinear stochastic differential equation in the continuous period of time, could be simplified as a set of linear stochastic different equations in discrete time intervals.

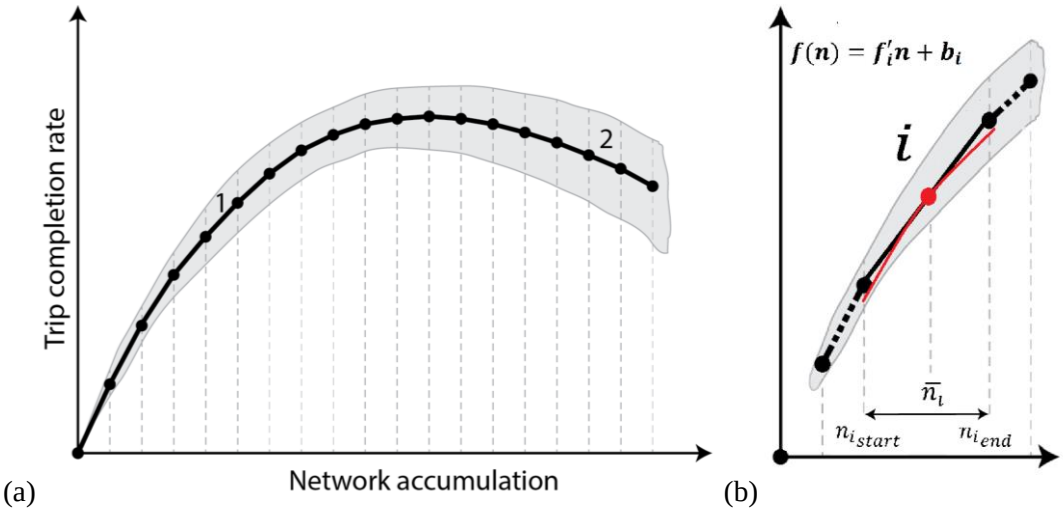


Figure 6-1. (a) Linearization of general NEF to affine function and (b) the linearized function based on one bin of NEF model.

Existing solutions for linear stochastic differential equations can then be appropriately used to study the dynamics and form the transition matrix of the MC model. This is done using the following process. Let us consider that the traffic state exists within bin i, which represents the range of accumulations $[n_{i_{start}}, n_{i_{end}})$ at some initial time t_0, as illustrated in Figure 6-1b. Here, we assume that the traffic state operates at the mean value of the bin $\overline{n_i} = 0.5(n_{i_{start}} + n_{i_{end}})$. The corresponding linear segment of the NEF used to describe traffic dynamics within this bin is $f(n) = f_i'n + b_i$, see Figure 6-1b. The dynamic process in a small time interval Δt is described by Equation 6-1. Using the solution to the linear SDE, we can obtain the distribution of traffic state in the next moment $t_1 = t_0 + \Delta t$; under the assumption of Gaussian error terms, the accumulations

expected at this future time period are normally distributed with mean and standard deviation values provided by Equations 6-2 and 6-3, respectively. Notice, Equations 6-2 and 6-3 are consistent with Equations 5-9 and 5-10, since Equations 6-2 and 6-3 are the results in a specific case when the intercept $b_i = 0$, the slope $f_i' = v$, and the initial state $\overline{n_\iota} = 0$.

$$\frac{dn_t}{dt} = E - (f_i' n_t + b_i) + G_0 w_t \quad \textbf{(6-1)}$$

$$\mu_{t_1} = \frac{E - b_i}{f_i'}\left[1 - e^{-f_i' \Delta t}\right] + e^{-f_i' \Delta t}\overline{n_\iota} \quad \textbf{(6-2)}$$

$$\sigma_{t_1}^2 = \frac{{G_0}^2}{2f_i'}\left[1 - e^{-2f_i' \Delta t}\right] \quad \textbf{(6-3)}$$

Given the future distribution of traffic state at time point t_1, we could estimate the probability that the network state exists within any of the n bins. We denote the probability from going from bin i to another bin j as a transition probability, $p_{ij}(\Delta t)$. This probability can be calculated using the following equation:

$$p_{ij}(\Delta t) = cdf_i\left(n_{j_{end}}\right) - cdf_i(n_{j_{start}}) \quad \textbf{(6-4)}$$

where cdf_i is the cumulative distribution function of traffic state at t_1 when the initial state is in bin i.

These transition probabilities can be calculated for the network state changing from each of the n starting bins to all n bins at the end of a time period Δt. The result is a complete transition matrix $P(\Delta t)$ that describes all transition possibilities for the network state; see Equation 6-5.

$$P(\Delta t) = \begin{bmatrix} p_{11}(\Delta t) & \cdots & p_{1n}(\Delta t) \\ \vdots & \ddots & \vdots \\ p_{n1}(\Delta t) & \cdots & p_{nn}(\Delta t) \end{bmatrix} \quad \textbf{(6-5)}$$

This transition matrix $P(\Delta t)$ is a key element of a Markov Chain that allows us to predict the traffic state distribution at any time in the future. Let the traffic state distribution be defined as a row vector N_t with n elements whose i^{th} entry represents the probability that the network accumulation lies within bin i. The distribution of traffic states at any tmie t in the future, N_t, can easily be calculated using the following equations if the initial traffic state vector N_0 and transition matrix $P(\Delta t)$ are known (and the latter is time-homogeneous; i.e., constant from time 0 to time t):

$$N_t = N_0 \cdot P(\Delta t)^{t/\Delta t} \quad \textbf{(6-6)}$$

$$N_t = [n_1(t), n_2(t), \dots, n_n(t)] \quad \textbf{(6-7)}$$

$$\sum_{i=1}^{n} n_i(t) = 1 \quad \textbf{(6-8)}$$

Thus, the MC framework is able to provide the future traffic state distribution for any time period in the future. The only needed information is the initial state distribution N_0 and the transition matrix P, which requires only knowledge of the NEF and vehicle entry rate. Note that a different value of P can be estimated for each time period Δt if network properties (i.e., NEF or the vehicle entry rate) change. Minor modifications would be needed for Equation 6-6 but the computational complexity would not increase significantly other than the calculation of these new transition matrices.

6.2 Insights from MC Model and Extensions

The Markov Chain could provide the analytical solutions of the estimated traffic state distribution over any long time period with very low calculation costs, since Markov Chains are general computationally efficient. This method could reduce the high cost and time required to get the numerical solutions of SDEs with Monte Carlo simulations.

The MC model also facilitates the study of steady-state properties of the network; i.e., the properties of network dynamics over an infinite period of time if conditions do not vary with time. If vehicle entry rate does not vary with time (or is constant over a significantly long period), the MC becomes time-homogeneous or time-independent and the transition matrix, $P(\Delta t)$, is constant; i.e., $p_{ij}(\Delta t) = (n_{t+\Delta t} = j | n_t = i)$ is independent of t. A row vector $\boldsymbol{\pi}$ is defined as the steady state (stationary distribution or equilibrium distribution), if $\boldsymbol{\pi}$ satisfies:

$$\boldsymbol{\pi} = \boldsymbol{\pi} \cdot P(\Delta t) \quad \textbf{(6-9)}$$

$$\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_n] \quad \textbf{(6-10)}$$

$$\sum_{i=1}^{n} \pi_i = 1 \quad \textbf{(6-11)}$$

According to the description of steady state, once the network is in the steady-state distribution, the network state does not vary with time and stay at the steady state. A MC model may own many steady-state distributions. Fortunately, all steady states of urban network could be derived by analysis of the transition matrix of Markov Chains, $P(\Delta t)$. By comparing the definition of the steady state $\boldsymbol{\pi}$ with Equation 6-9 and that of a left eigenvector $\boldsymbol{v}$ satisifying $\lambda \boldsymbol{v} = \boldsymbol{v} \cdot P(\Delta t)$, the two concepts are related that $\boldsymbol{\pi}$ is a left eigenvector of $P(\Delta t)$ with an eigenvalue $\lambda = 1$. Thus, by analyzing the left eigenvectors of $P(\Delta t)$, all the steady states can be identified.

Note that no assumption is made regarding the initial state distribution. There exists the possibility that different initial state would converge towards different stationary distributions. For instance, when the network is controlled with a relative low value of metering rate E, the traffic state distribution would evolve and stay at the stationary distribution at free-flow branch when the network is initially empty $n_0 = 0$, as illustrate in Figure 5-4a. If the network is initially congested, (i.e., when the initial state is near n_j), the traffic state will gradually approach to grid lock with accumulation n_j. If the network is implemented with a high value of E, as shown in Figure 5-3, the state would evolve toward the gridlock state regardless of the initial state.

The traffic state near n_j with almost zero trip complement represents gridlock, which is a disaster for traffic engineers. Therefore, the metering rate E should be carefully selected to guarantee the existence of steady state in the free-flow branch or near the critical near the critical accumulation n_j to avoid excessive delay. The steady state analysis allows us to derive the steady states of network, which could be utilized to optimize the control strategies and maximize the trip complement rate.

Additionally, MC model is an ideal analysis tool for modelling the time-dependent and complex dynamic control of the network due to its flexibility and extensibility. For instance, when a dynamic perimeter metering strategy is applied in an urban network, the metering flow is time-inhomogeneous according to the current traffic state. Markov chains could model this complicated scenario by building a time-homogenous transition matrix, whose elements in each row are calculated with a different value of E. Consider again Figure 6-1a. When traffic state is in bin 1, a relatively high value of E should be chosen to maximize network productivity. When the traffic state stays at bin 2, a low value of E is picked to ensure that the network moves away from the congested branch and back to the free-flow branch. For both of these cases, different values of E would have to be used in the derivation of the transition probabilities. However, since the E value

is state-dependent, they can easily be incorporated into the same matrix $P(\Delta t)$. Thus, in this way, the state-dependent entrance flow (E_i) does not add any difficulty of traffic dynamic modelling with MC model, as it requires only a simple extension:

$$\frac{dn_t}{dt} = E_i - (f_i' n_t + b_i) + G_0 w_t \quad \textbf{(6-11)}$$

Besides, that Markov Chain model can easily be updated to include time-dependent entrance flows. In this case, a unique transition matrix can be calculated for each time period, $P(\Delta t, t)$.

Also note that the previous analysis considered only one type of uncertainty: the uncertainty of metering flow E, which is a stable distribution with a constant standard deviation G_0. However, the MC framework is also able to consider the uncertainty in the exiting flow when the NEF model exhibits large amounts of stochastic behavior. The uncertainty in the traffic model, i.e., $g(n)$ in the Equation 5-1 and Equation 5-2, depends on traffic state and solutions to these types of SDEs are difficult to solve analytically. Fortunately, MC method could solve this problem by simply assigning a different level of uncertainty to each of the bins. The level of uncertainty chosen would have to appropriately accommodate the uncertainty due to vehicle entries as well as uncertainty in the NEF, but this is fairly trivial to do as shown below:

$$\frac{dn_t}{dt} = E - (f_i' n_t + b_i) + (G_0 + G_f \cdot n_t) w_t \quad \textbf{(6-12)}$$

$$\frac{dn_t}{dt} = E - (f_i' n_t + b_i) + (G_0 + G_f \cdot \overline{n_i}) w_t \quad \textbf{(6-13)}$$

where $\overline{n_i}$ is the mean value of bin i, $\overline{n_i} = 0.5(n_{i_{start}} + n_{i_{end}})$.

6.3 Numerical Simulation Tests

To test the accuracy of the MC model, this method is evaluated by comparing the predicted dynamic processes within an urban network with those obtained using a simple Monte Carlo simulation that represents the ground truth situation. Although the random effects are involved into the control process, Monte Carlo could obtain reliable numerical results based largely on repeated random sampling. In the discrete-time Monte Carlo simulation, a small time interval $\Delta t = 1$ [sec] is used for a 5-hr traffic simulation to guarantee the accuracy and reliability of the results.

As discussed in the scenario description, the NEF traffic model could take different forms: a simple triangular NEF model and a more complex nonlinear NEF model. Uncertainty also takes various forms. The large-scale stochastic phenomena might only exist in the metering flow with a well-defined NEF model, named Type I uncertainty, or it also occurred in the exiting flow (Type II) because the NEF traffic model is not well-defined. Furthermore, we have different types of traffic conditions that might arise: those that lead to complete gridlock and those that do not. We test the MC model under various combinations of those variables; a complete list is provided in Table 6-1.

Table 6-1. Scenarios Description with Different Variables

Scenario	NEF model		Uncertainty		Complete gridlock	
	triangular	nonlinear	Type I	Type I&II	No	Yes
1	√		√		√	
2	√		√			√
3	√			√	√	
4	√			√		√
5		√	√		√	
6		√	√			√
7		√		√	√	
8		√		√		√

6.3.1 Triangular NEF Model

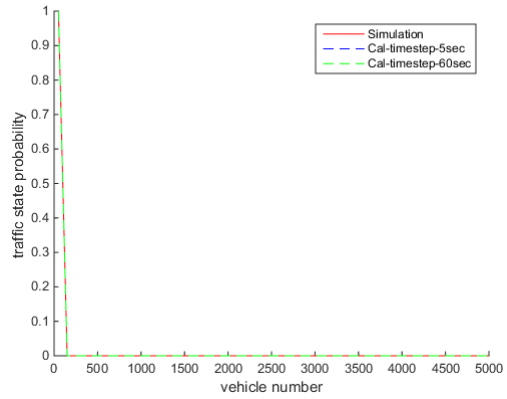
Type I Uncertainty

A single-region downtown during a 5-hr typical morning peak period with a perimeter metering strategy is considered, as shown in Figure 5-2. Traffic control could restrict the vehicle entry rate onto the network as a Gaussian distribution with mean value of E and standard deviation of G_0. Since only Type I uncertainty is included in the first scenario, the triangular NEF function is described with Equation 5-1 with $G_0 \neq 0$ and $G_f = 0$. In the application of MC model, we use $\Delta n = 100$ [veh] as the unit of traffic state given the maximum vehicle accumulation $n_j = 15{,}000$ [veh], and two values of time interval $\Delta t = 5$ [sec] and $\Delta t = 60$ [sec] to build the transition matrix.

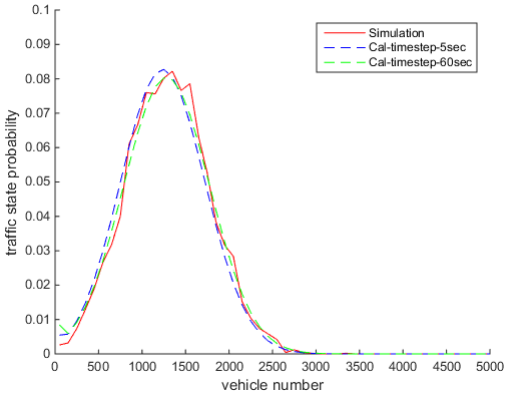
Different combinations of E and G_0 would result into different urban traffic state distributions. The MC model can help determine whether there exists the stationary distribution of urban traffic state around the capacity section or free-flow branch by analyzing the known transition matrix. If this stationary distribution exists, network traffic would stay at the distribution and completely avoid the traffic jam. Otherwise, there is a high opportunity to create the congestion, and the possibility increases over time. In that regard, it is necessary to draw a distinction between the non-jam cases and jam cases.

In the non-jam case, we choose the pair of of $E = 120{,}000$ [veh/hr] and $G_0 = 350{,}000$ [veh/sec] = 5,833.33 [veh/hr] to describe the entry rate under the perimeter metering control. The discrete-time Monte Carlo simulation could get the numerical solution of traffic state distribution at each time point by 50,000 simulation iterations. Compared with simulation results, the MC model could derive the analytical solution without the need of largely repeating simulation. The analytically derived traffic state distributions at different times ($t = 0, 1min, 5\ min\ and\ 5hr$) are

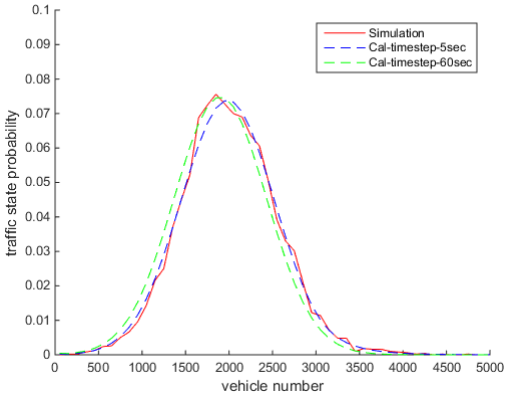
plotted on Figure 6-2, as dotted lines while the simulation results are also included as the solid lines.



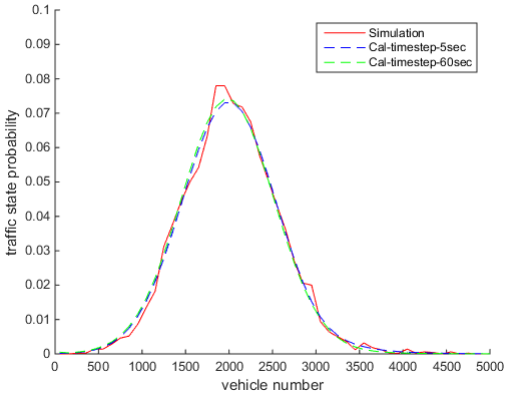
(a)



(b)



(c)



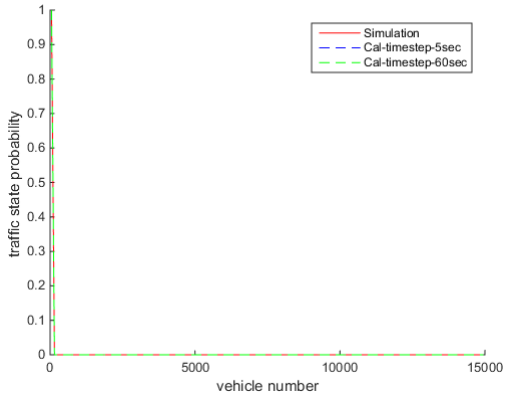
(d)

Figure 6-2. Traffic state distributions in scenario 1 with bin size =100 at the time (a) t = 0; (b) t = 1 min; (c) t = 5 min; and (d) t = 5hr.

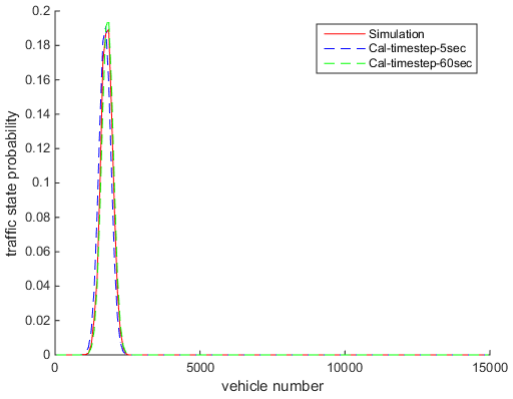
As shown, the urban network starts initially empty (i.e., $n_0 = 0$), and then gradually approaches a stationary distribution in the free-flow range [0 3,000], which it stays at until the simulation completes. Of course, the simulation results still show some stochastic fluctuations compared with MC model although simulations are repeated by 5,000 iterations to get a reliable numerical solutions; see in Figure 6-2d. In order to achieve an idealistic results, more simulation iterations are needed with high computational cost, and the necessary iteration number is unknown to us. However, Figure 6-2 shows that analytically predicted traffic states are consistent with the simulation with negligible error. The urban network reaches the stationary state in 5 min and stays at same state distribution during the rest of the peak hours. It demonstrates that MC model could fairly accurately describe the evolution of the system within consideration of stochastic phenomena in traffic control.

In the scenario with the expectation of jammed traffic, we use a different pair of $E = 170{,}000$ [veh/hr] and $G_0 = 140{,}000$ [veh/sec] = 2,333 [veh/hr] to build a Gaussian distribution of metering flow. In this scenario, vehicle state distribution changes significantly over time—from an empty network to a congested network—and approaches the traffic gridlock marked with the maximum vehicle accumulation $n_j = 15{,}000$[veh], see Figure 6-3. Compared with the simulation

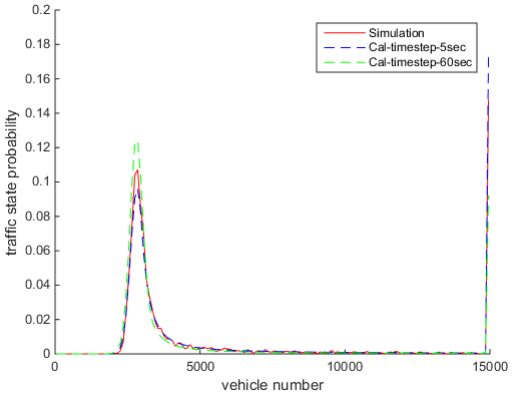
results, the analytical solutions of MC model could work well to reveal the dynamic process of urban control. Notice, two different time intervals are used here, $\Delta t = 5$[sec] and $\Delta t = 60$[sec], which generate two different expected distribution with small deviations. Therefore, the choice of an appropriate time interval might slightly affect the accuracy of analytical solutions, which will be further tested in other scenarios.



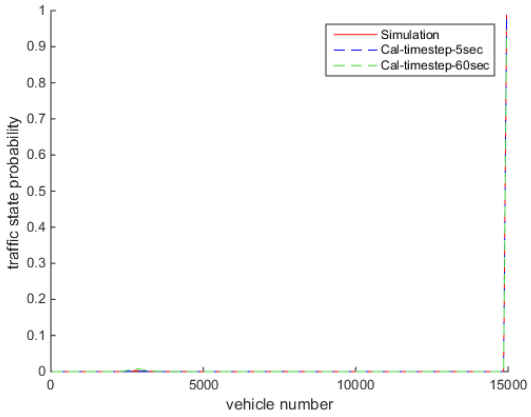
(a)



(b)



(c)



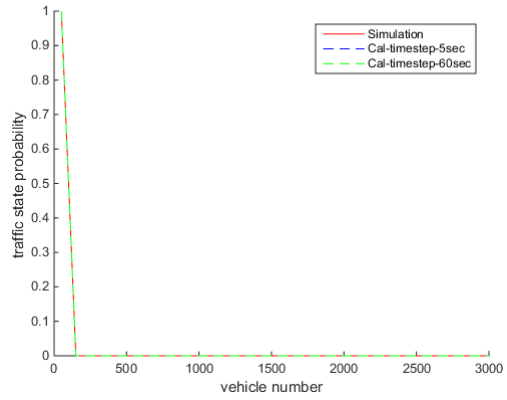
(d)

Figure 6-3. Traffic state distributions in scenario 2 with bin size =100 at the time (a) t = 0; (b) t = 1 min; (c) t = 0.5hr; and (d) t = 5hr.

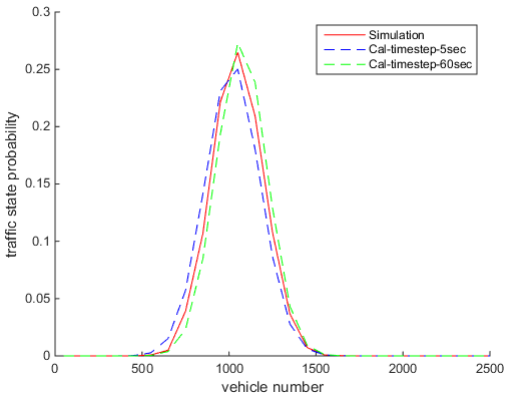
Type I & II Uncertainty

The Type II uncertainty in the NEF traffic model is added in these scenarios to account for the phenomena of a range of exiting traffic flows associated with the same accumulation. As discussed above, MC method could easily model the Type II uncertainty in the analysis by simply assigning a different level of uncertainty to each of the bins. Here, we set the value of $G_f = 0.333$[veh/hr] Therefore, the standard deviation of exiting flow is in the range $[0\ \ 5{,}000(G_f * n_j)]$ [veh/hr]. Type I uncertainty of entry rate is also considered here with $E = 100{,}000$ [veh/hr] and

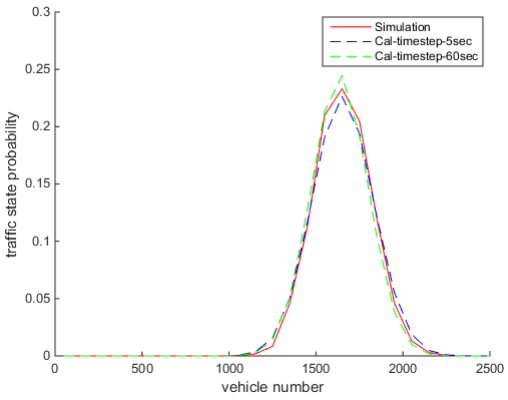
$G_0 = 1{,}666.67$ [veh/hr] in non-jam case and $E = 170{,}000$ [veh/hr] and $G_0 = 1{,}666.67$ [veh/hr] in the jam case. Therefore, Type I and Type II are in the same range and neither is assumed to dominate.



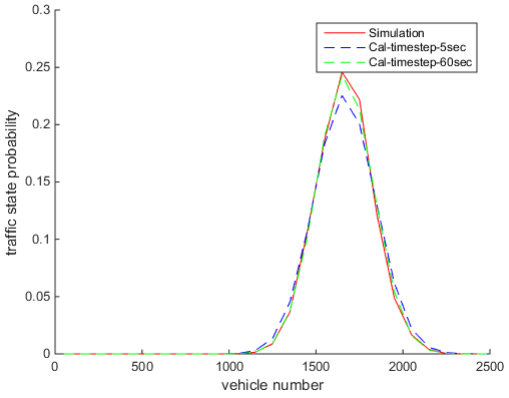
(a)



(b)

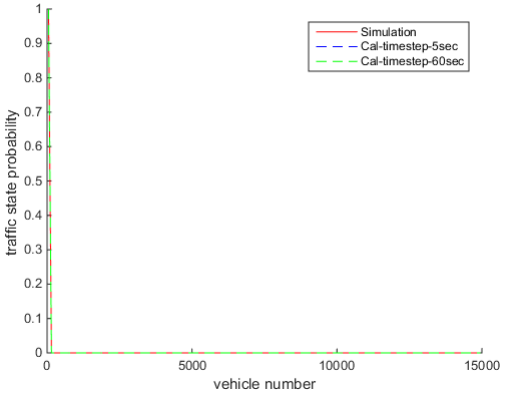


(c)

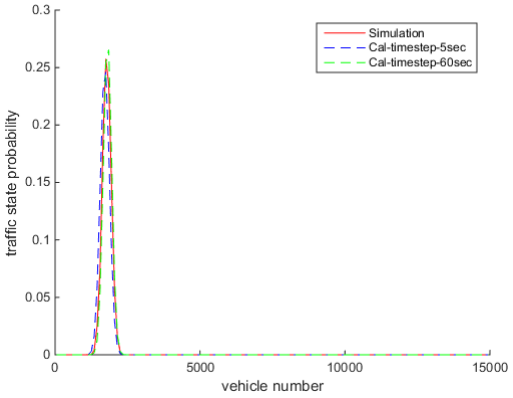


(d)

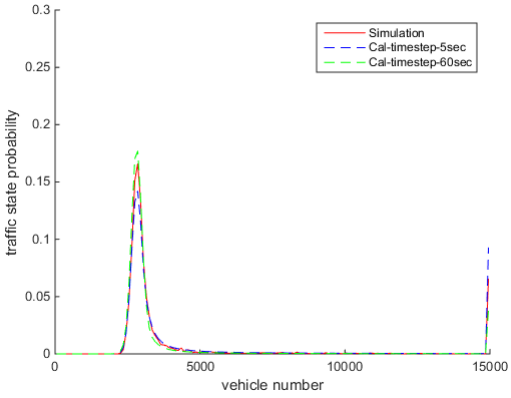
Figure 6-4. Traffic state distributions in scenario 3 with bin size =100 at the time (a) t = 0; (b) t = 1 min; (c) t = 5 min; and (d) t = 5hr.



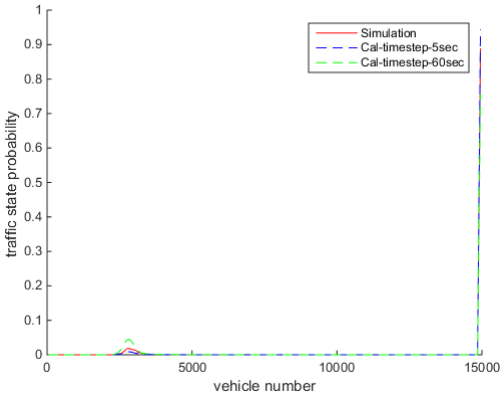
(a)



(b)



(c)



(d)

Figure 6-5. Traffic state distributions in scenario 4 with bin size =100 at the time (a) t = 0; (b) t = 1 min; (c) t = 0.5 hr and (d) t = 5hr.

Figures 6-4 and 6-5 reveal that the analytically predicted network state distributions in non-jam case and jam case obtained from the MC model are consistent with simulation results. This suggests that MC model with low calculation costs is fairly robust and valid for modelling the complicated dynamic process considering two different types of uncertainty.

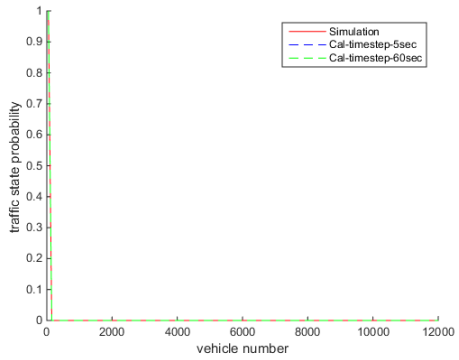
6.3.2 Nonlinear NEF Model

The triangular NEF traffic model consists of only two unique linear segments. However, the nonlinear model is modeled using a set of multiple piece-wise linear segments in the MC model application. In fact, a slight change in vehicle accumulation in the nonlinear model can change the presentation of the linearized function. Therefore, the conclusion that the MC model is an effective tool of dynamic analysis based on the triangular model does not directly apply to the nonlinear model without further verification. Similar numerical simulation tests are conducted in this section to verify the consistency between the MC models with the nonlinear NEF model. The nonlinear model used in the section is built on the empirical data from Yokohama, Japan (Geroliminis and Daganzo 2008), which is described using Equation 5-2.

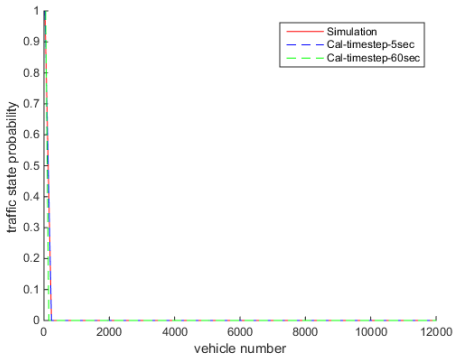
Type I Uncertainty

The numerical tests start with Type I uncertainty in the entry rate. Here, two representative pairs of parameters $E = 25{,}000$ [veh/hr], $G_0 = 4{,}166.67$ [veh/hr], and $E = 33{,}000$ [veh/hr], $G_0 = 4{,}166.67$ [veh/hr] are chosen in the non-jam case and in the jam case, respectively. Simulation iterations are increased to $10{,}000$ to get smoother simulated traffic distributions. The simulation results, however, still oscillate up and down as the simulation represents a random process, see Figure 6-6 and 6-7. Two different bin sizes $\Delta n = 100$ [veh] and $\Delta n = 340$ [veh] are used in the MC calculations along with the two time intervals $\Delta t = 5$ [sec] and $\Delta t = 60$ [sec]. In the non-jam case, the urban network would reach the stationary state at around 30 min, as illustrated in Figure 6-6. And for the jammed case, the MC model predicts that the increasing jam probability reaches the maximum value of 50% at the end of the peak hours.

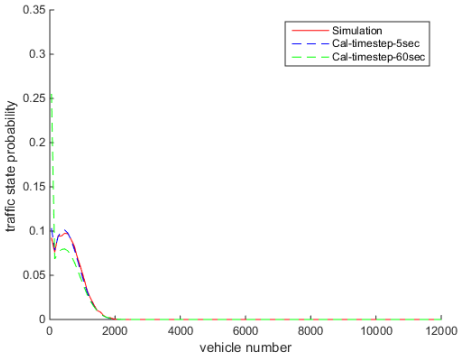
(a)


Simulation
Cal-timestep-5sec
Cal-timestep-60sec
traffic state probability
vehicle number

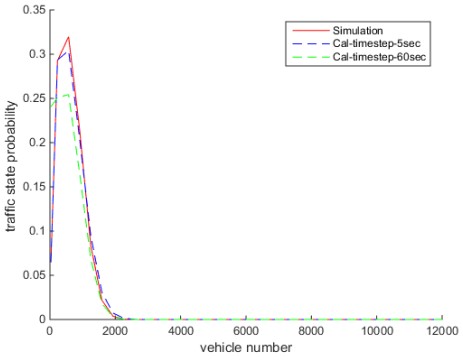
(f)


Simulation
Cal-timestep-5sec
Cal-timestep-60sec
traffic state probability
vehicle number

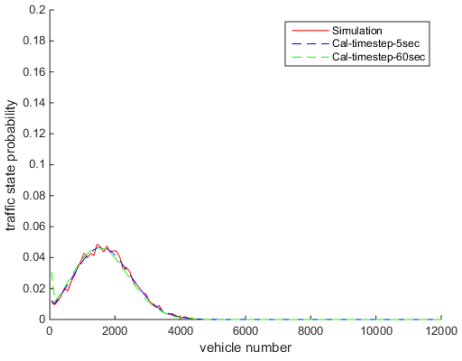
(b)


Simulation
Cal-timestep-5sec
Cal-timestep-60sec
traffic state probability
vehicle number

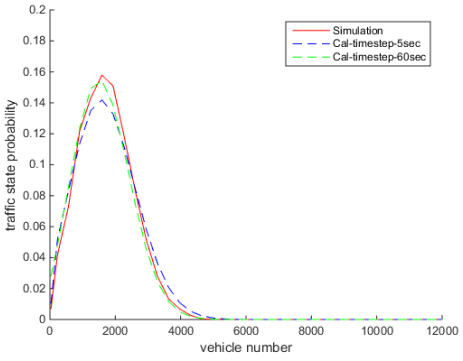
(g)


Simulation
Cal-timestep-5sec
Cal-timestep-60sec
traffic state probability
vehicle number

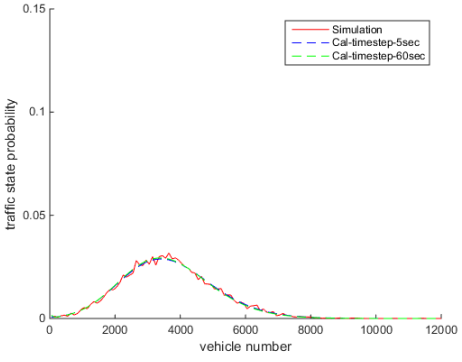
(c)


Simulation
Cal-timestep-5sec
Cal-timestep-60sec
traffic state probability
vehicle number

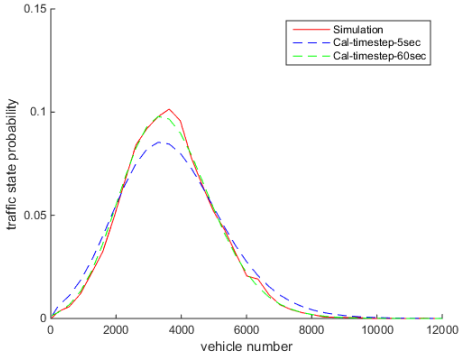
(h)

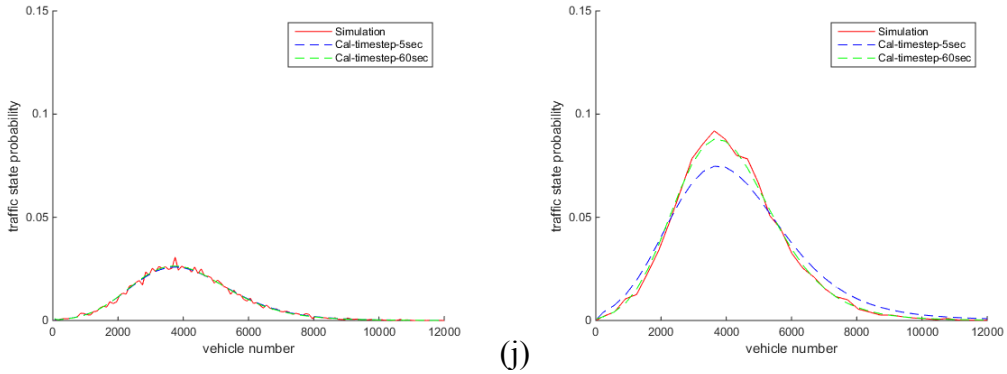

Simulation
Cal-timestep-5sec
Cal-timestep-60sec
traffic state probability
vehicle number

(d)


Simulation
Cal-timestep-5sec
Cal-timestep-60sec
traffic state probability
vehicle number

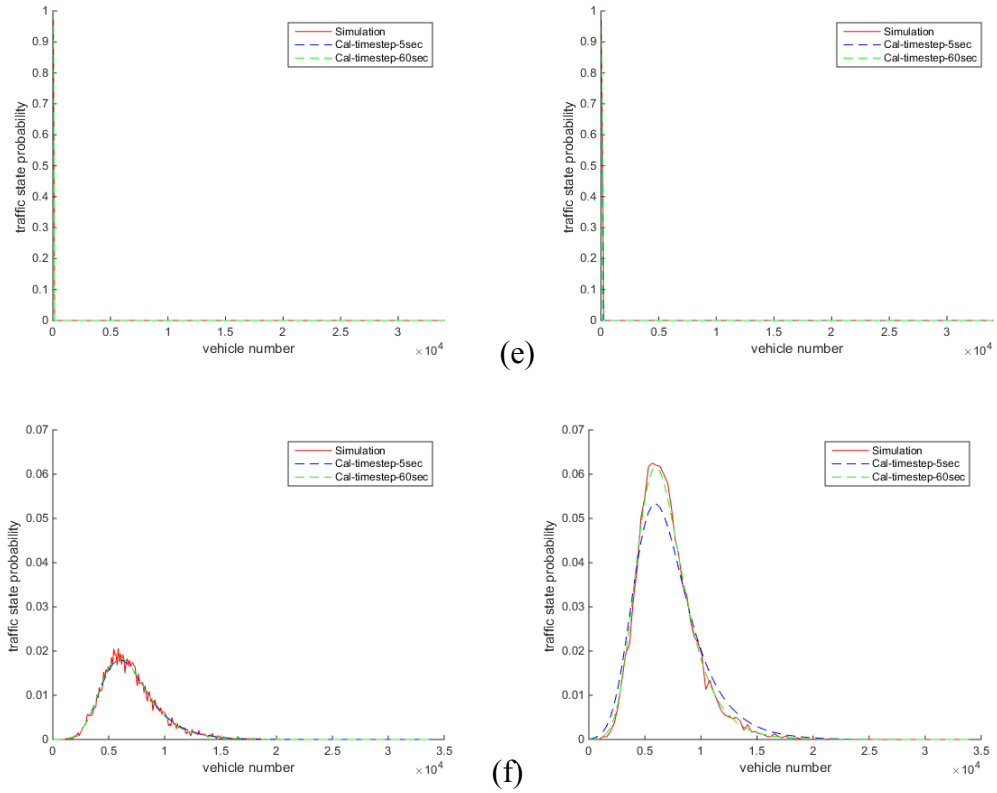
(i)


Simulation
Cal-timestep-5sec
Cal-timestep-60sec
traffic state probability
vehicle number



(e) (j)

Figure 6-6. Traffic state distributions in scenario 5 at the time (a) t = 0; (b) t = 1 min; (c) t = 5 min; (d) t = 30 min; (e) t = 5 hr with bin size =100 and (f) t = 0; (g) t = 1 min; (h) t = 5 min; (i) t = 30 min; (j) t = 5 hr with bin size =340.



(a) (e)

(b) (f)

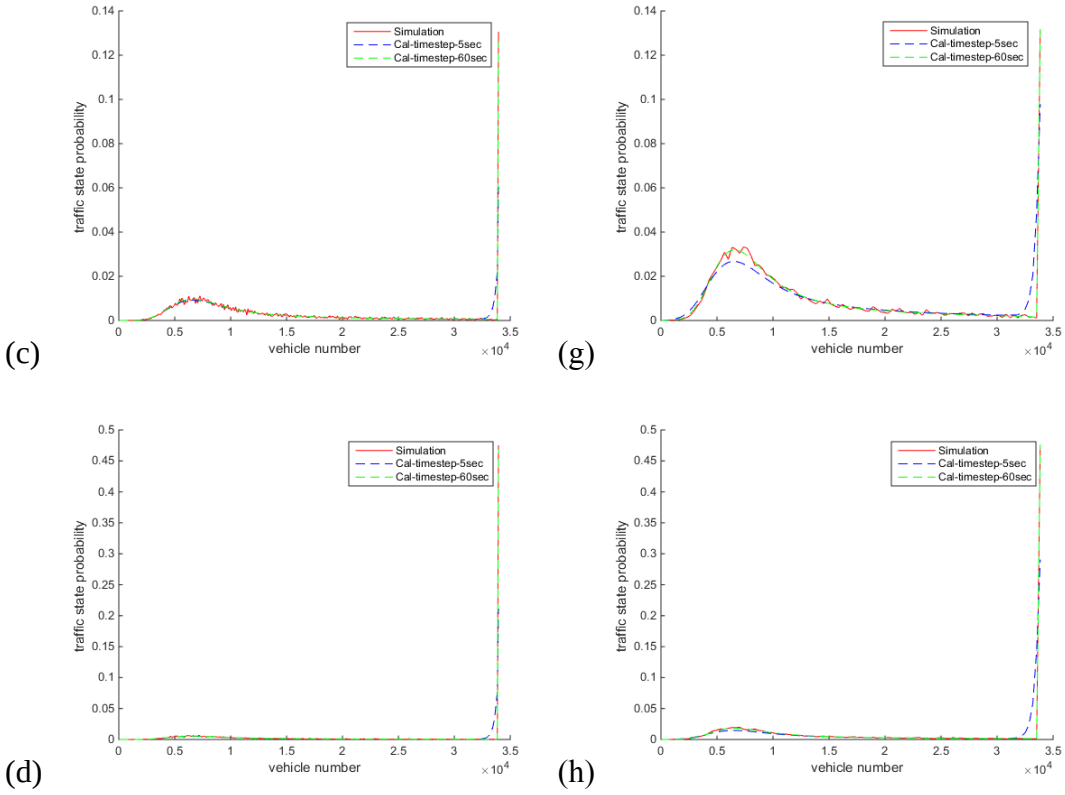


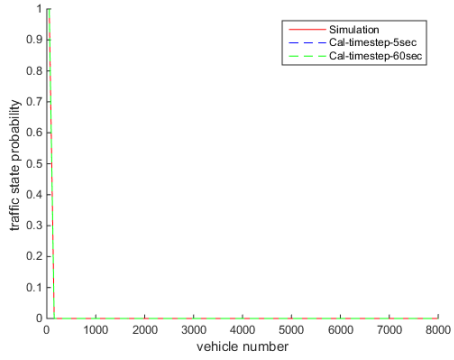
Figure 6-7. Traffic state distributions in scenario 6 at the time (a) t = 0; (b) t = 1hr; (c) t = 3hr; (d) t=5hr with bin size =100; (e) t = 0; (f) t = 1hr; (g) t=3hr; and (h) t = 5hr with bin size =340.

Different values of bin size would significantly influence the shape of traffic state distribution diagrams in Figure 6-6 and Figure 6-7. One obvious reason is that traffic state distribution is calculated in the unit of bin size, representing the probability of vehicle accumulation in a bin. But it doesn't affect the accuracy of traffic state distribution prediction. Interestingly, the analytical results under time interval $\Delta t = 5$ [sec] and bin size $\Delta n = 340$ [veh], i.e. dotted blue lines in Figure 6-6 (g)~(j) and Figure 6-7 (f)~(h), show small deviations from the simulated results. Fortunately, the errors between analytical solutions of MC model and numerical solutions by simulation could be mitigated when a large time interval $\Delta t = 60$ [veh] is adopted. These results suggest that setting a pair of appropriate time interval and bin size could improve the accuracy of MC model.

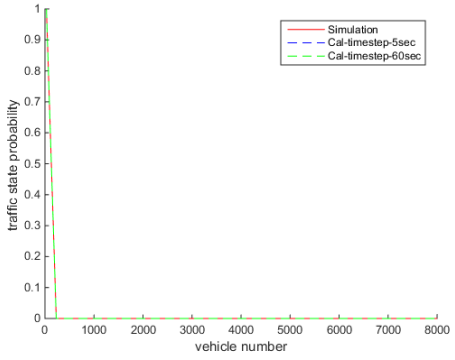
Type I & II Uncertainty

Then the simulation tests are extended to include Type II uncertainty in the exiting flow. Here, we set the value of $G_f =$ [veh/hr]. Therefore, the standard deviation of exiting flow is in $[0\ \ 11{,}333.33\ (G_f * n_j)]$ [veh/hr]. Notice, Type I uncertainty is also included as well. The pair of of $\mathrm{E} = 250{,}000$ [veh/hr] and $G_0 = 4{,}166.67$ [veh/hr] is applied in the non-jam case, while the parameters $\mathrm{E} = 330{,}000$ [veh/hr] and $G_0 = 1{,}666.67$ [veh/hr] would result in high jam probability, see in Figure 6-8 and Figure 6-9.

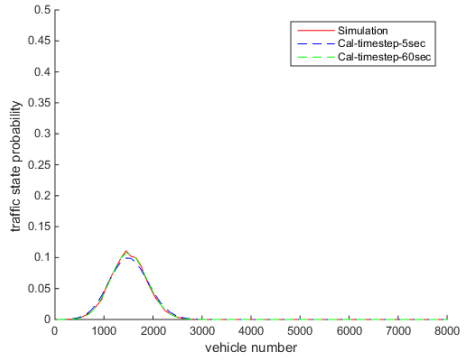
(a)



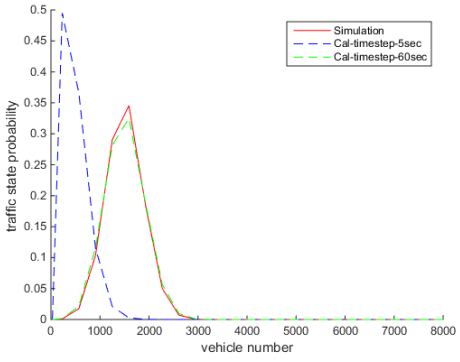
(e)



(b)



(f)



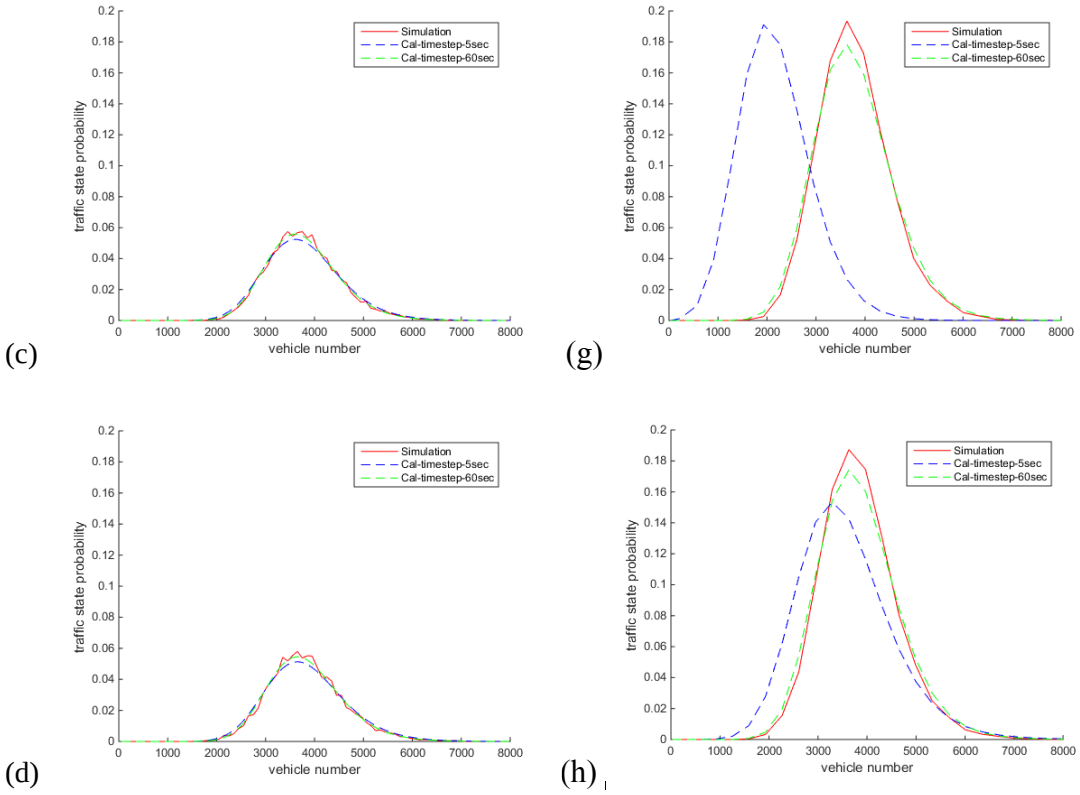
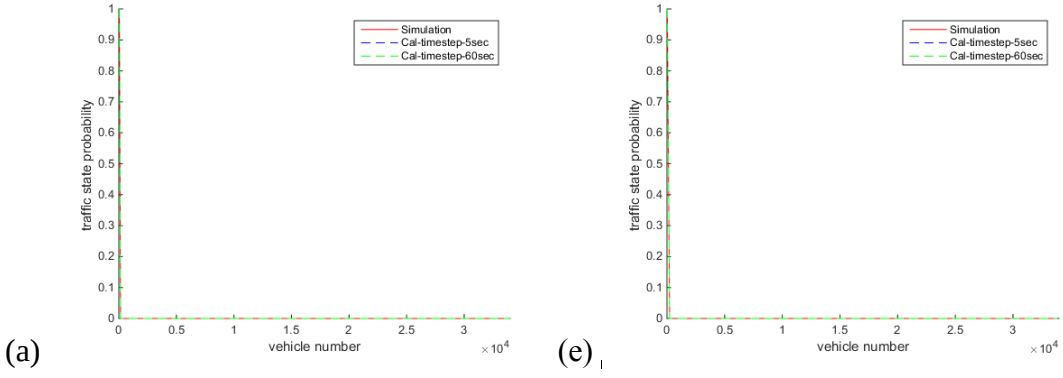


Figure 6-8. Traffic state distributions in scenario 7 at the time (a) t = 0; (b) t = 5min; (c) t = 1hr; (d) t = 5hr with bin size =100; (e) t = 0; (f) t = 5min; (g) t = 1hr and (h) t = 5hr with bin size =340.



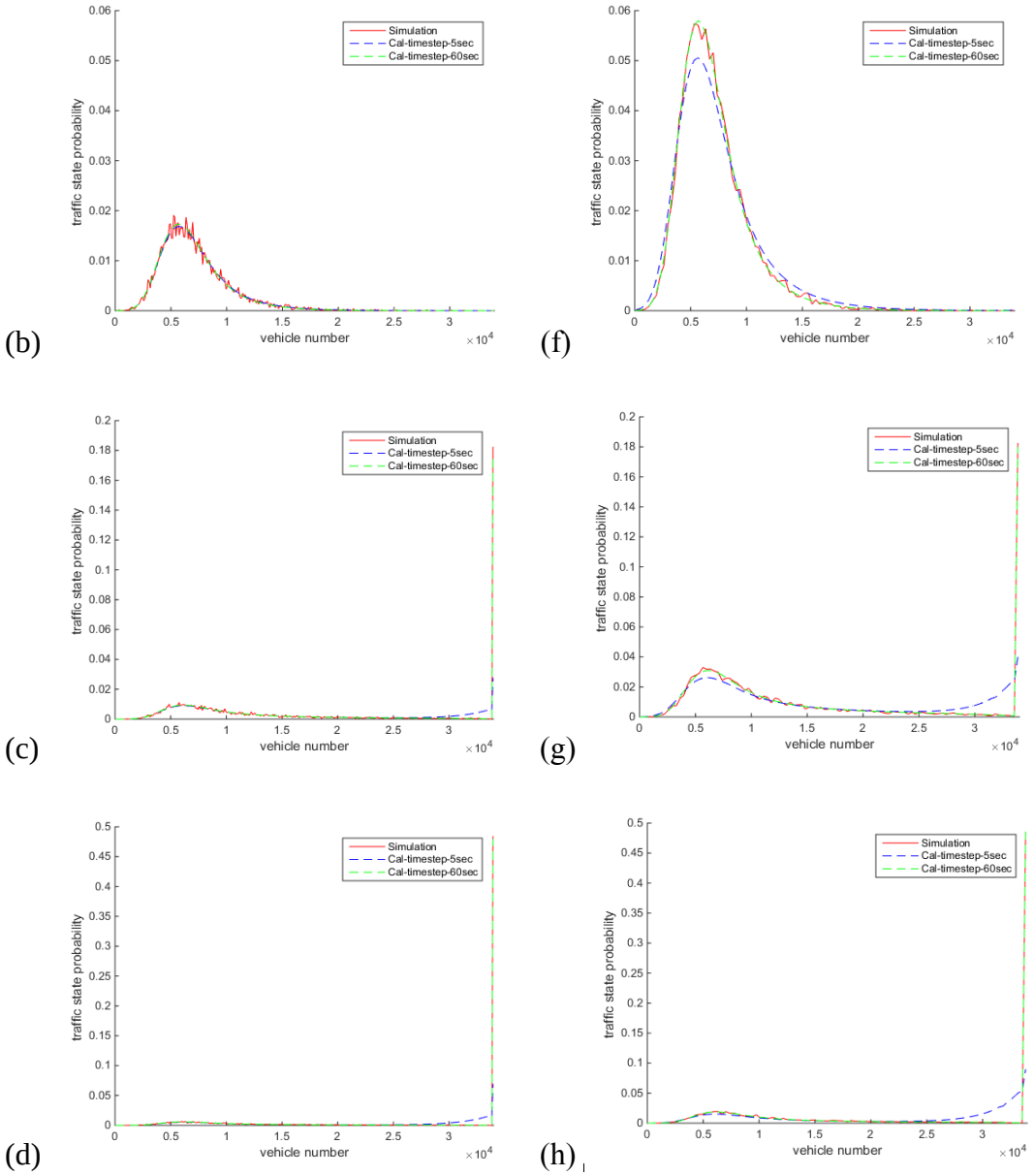


Figure 6-9. Traffic state distributions in scenario 8 at the time (a) t = 0; (b) t = 1hr; (c) t = 3hr; (d) t=5hr with bin size =100; (e) t = 0; (f) t = 1hr; (g) t=3hr; and (h) t = 5hr with bin size =340.

The analytically derived vehicle state distributions are plotted on Figures 6-8 and 6-9 as solid lines while the simulation results with 10,000 iterations are included as dotted lines. The simulations verify that the MC model analytical solutions works very well in the non-jam case and

jam cases with an appropriate selection of bin size and time interval. In fact, a combination of $\Delta n = 100$ [veh] and $\Delta t = 60$ [sec] could produce the analytical solutions of vehicle state distribution, which are the best fit for simulation results. The distributions with $\Delta n = 340$ [veh] and $\Delta t = 5$ [sec] exhibit significant and constant drift away from the analytically predicted states. The phenomena further verifies that MC method works well to derive the analytical solutions of network state in the dynamic process with the optimized bin size and time interval.

Chapter 7

Parametric Study

7.1 Parametric Analysis

The MC model can be used for describing the urban network dynamic process because urban traffic networks follow a Markovian process. In the literature, there exist various Markov chains with different type of parameters. The Markov chain usually refers to a process with a discrete set of times, as the model used in this study. On the other hand, a continuous-time Markov chain is a model in which the time spent remaining in every state depends only on the current state. While the time parameter could be either discrete or continuous, the Markov chain usually employs finite or countably infinite (that is, discrete) state spaces, which have a more straightforward statistical analysis.

The MC model employed in our urban traffic network dynamics analysis is a discrete-state and discrete-time Markov chain. Therefore, the traffic state that is originally provided by the continuous value of vehicle accumulation is redefined as a discrete bin with a predetermined range of vehicle accumulation; see Figure 6-1a. Using this MC method also requires that the time is discretized and the dynamic process of traffic control is modeled by a series of small consecutive time intervals. In each time interval, urban network dynamics follows a linearized NEF traffic model, $f_i(n_t) = f_i' n_t + b_i$, which is described with Equation 7-1. In this way, all the transition probabilities from one state to another in the small time interval can be derived from a set of stochastic differential equations, and this can be used to form a full transition matrix for the MC model.

$$\frac{dn_t}{dt} = E - f_i(n_t) + (G_0 + G_f \cdot \overline{n_1})w_t \quad \textbf{(7-1)}$$

where i is current traffic state with bin $[n_{i_{start}} \quad n_{i_{end}})$, and $\overline{n_1} = 0.5(n_{i_{start}} + n_{i_{end}})$.

The MC model proposed here for urban traffic network dynamics control belongs to the family of Markov chain approximation method (MCMA). Kushner (1990) adopted a discrete-state and discrete-time Markov chain model and demonstrated that this model could work well for numerically approximating the solutions to stochastic control problems. The basic idea of the time and state space discretization of the MC model is inspired by these models adopted in Kushner (1990). This study also pointed out that the time interval Δt has an important role to play for modeling—it should be carefully determined according to state-space parameter Δn to ensure the effectiveness of the approximating Markov chain.

Due to the discretization of both time and space, there are certain differences that arise in the solutions of different MC models. Unfortunately, applying the MC model with poor and careless choice of these parameters would render the model invalid and generate results beyond an acceptable tolerance. For example, in the Figures 6-6 to 6-9, there exist gaps between MC model solutions and simulation results. Therefore, a method to select the parameters Δt and Δn should be proposed to limit the noise in tolerance.

The bin size Δn determines the formation of the discrete state space, which directly affects the traffic state distribution description. Larger bin sizes result in greater difference between the original traffic state, described by the continuous vehicle accumulation, and the discrete state modeled in the MC framework. For example, in all of our scenario tests, it is assumed that the network is initially empty; i.e., $n_0 = 0$. However, in the MC framework the state is represented by the first bin with the range $[0, \Delta n]$. It is obvious here that selecting a smaller bin size would

minimize the error between the original state being modeled and the one assumed in the MC framework. Beside the traffic state description, the bin size also plays a large role on the linearization of the NEF model, as illustrated in Figure 6-1b. As $\Delta n \to \infty$, the piece-wise linear NEF model converges to the smooth and curved model. Thus, again, it is clear that selecting the smallest bin possible would provide the most accurate description of traffic state and traffic model. However, selecting too small of a bin size would increase the computational complexity required to employ the MC framework. Therefore, when selecting the value of Δn that will be used in the MC model, a modeler will have to consider the tradeoff between modeling accuracy and computational complexity.

Bin size Δn is related only to the issue of traffic state and urban traffic model description, rather than the urban traffic network dynamics. Compared with bin size, the time interval Δt plays a more important role in the dynamic process, and its selection is much more complex. As mentioned above, the time parameter should be defined according to state-space parameter. Thus, two extreme cases are used here to study the possible effects of time interval regardless of the setting of the bin size. Assume that the network is initially in the state i; the probability distribution of the next moment traffic state is derived using Equation 7-1. Figure 7-1 plots this distribution for three different values of time interval Δt. If a very large time interval Δt is chosen, the traffic state at the next moment could potentially leave bin i and enter another bin(s) that is far from bin i, as illustrated in Figure 7-1a. In this case, the network dynamic process should follow multiple NEF segments (the blue line). However, the MC model will assume that this dynamic process is described only by the features of a single bin (the red line), and this could result in a large calculation error. The case with the large time interval can be characterized by a very large probability that the traffic state will leave bin i, defined as $leaving\_pp_i = 1 - p_{ii}(\Delta t)$. Conversely, when a very small interval is applied, traffic state at the next moment is highly likely

to stay still within initial bin i (Figure 7-1b); i.e., $p_{ii}(\Delta t) \cong 1$, $leaving\_pp_i \cong 0$. In this case, the traffic state will not change with time and state i would be incorrectly modeled as an absorbing state that is impossible to leave. With extremely small value of time interval, every state would become an absorbing state, and $P(\Delta t)$ would simply become the matrix I_n. This would imply that the urban network does not have any dynamic change and stays at the initial state distribution, which is not in consistent with the real situation and which will inevitably cause calculation errors. According to above analysis, the selection of time interval is closely related to the dynamics of urban traffic: large time intervals would cause this process to be described incorrectly, and small time intervals would create the excessive absorbing states. Therefore, an optimal time interval must be selected that is neither too large nor too small to avoid possible errors and maximize the accuracy of the MC framework. Such time interval is likely to have a value of $leaving\_pp$ that is somewhere in-between the two extreme values discussed in this paragraph; see Figure 7-1c. Note that the selection of time interval would depend upon the bin size selected.

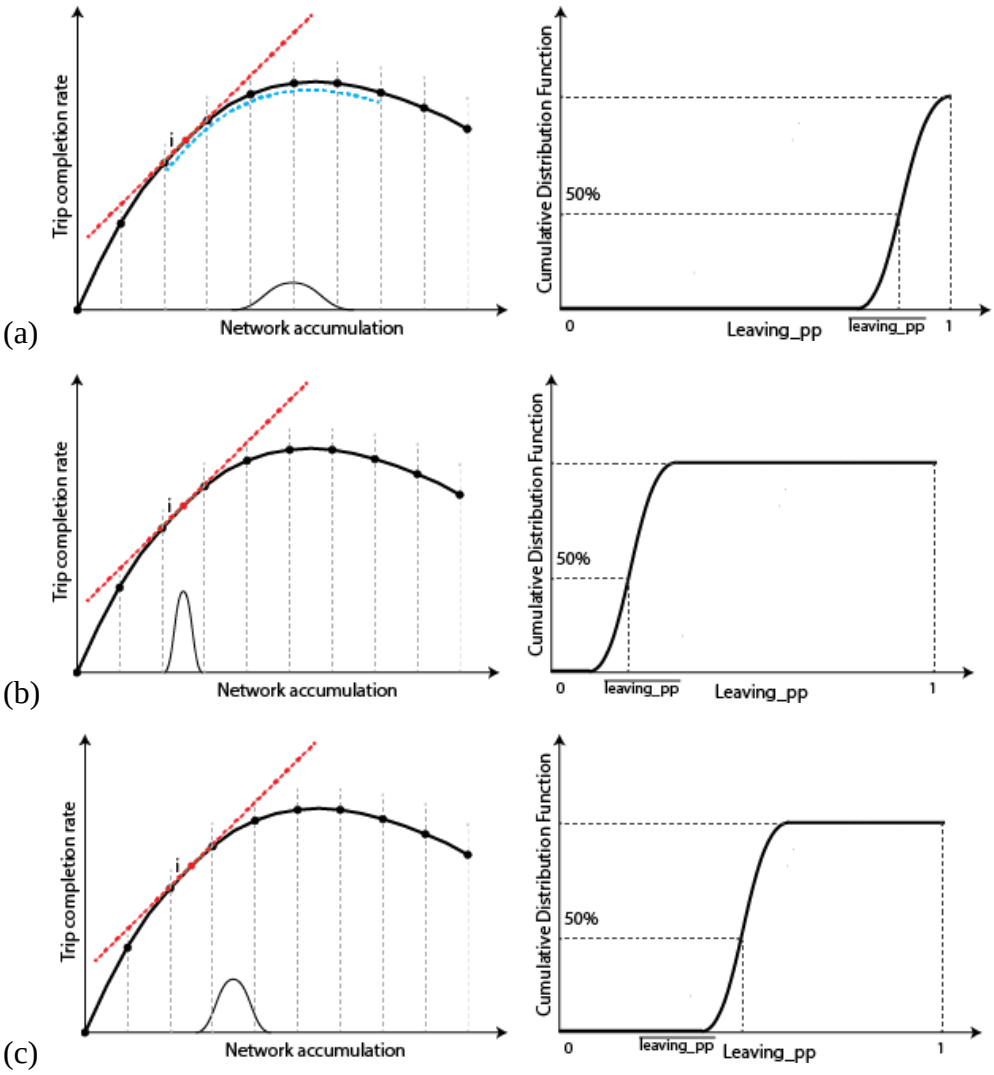


Figure 7-1. The traffic state distribution at the next moment with the initial state at segment *i* and the cumulative probability function of leaving_pp with the same bin size and (a) large time interval; (b) small time interval and (c) appropriate time interval.

The discussion could help to further explain the previous numerical test results with nonlinear NEF models. MC model applications use four different sets of bin size and time interval: 100 [veh] and 5 [sec], 100 [veh] and 60 [sec], 340 [veh] and 5[sec], and 340 [veh] and 60 [sec], see Figures 6-6 to 6-9. Only the Markov chain results with $\Delta n = 340$ [veh] and $\Delta t = 5$[sec] present huge gaps with Monte Carlos simulation results. This might stem from a low level of the $leaving\_pp$ distribution due to the selection of a small time interval and a large bin size. By

choosing a smaller bin size $\Delta n = 100$ [veh] or a larger time interval $\Delta t = 60$[sec], the overall value of $leaving\_pp$ distribution increases, and the errors are reduced.

For the remainder of this section, the index $leaving\_pp_i$ is specified for the property of the traffic state in the bin i. In order to describe the overall level of the $leaving\_pp$ distribution across all the traffic states, a single value is used: the mean $leaving\_pp$ across all bins, as illustrated in Figure 7-1. This can be easily calculated from the transition matrix using the Equation 7-2.

$$\overline{leaving\_pp} = \frac{1}{n}\sum_{i=1}^{n}[1 - p_{ii}(\Delta t)] \quad \textbf{(7-2)}$$

As discussed above, time interval Δt and bin size Δn are two key parameters to build the MC model that affect the accuracy of dynamic analysis and need to be carefully selected. The two parameters should be considered together since their combined effects determine the dynamic change of traffic state, which could be characterized by $\overline{leaving\_pp}$. The optimal set of bin size and time interval would be one that would produce an optimal value or range of $\overline{leaving\_pp}$. Of course, the smallest bins sizes are preferred because they limit the errors between the MC model and the actual traffic description.

7.2 Numerical Simulation Tests

The parametric analysis unveils the significance of time interval and bin size, and suggests that these two parameters should be optimized by controlling the overall level of $leaving\_pp$ distribution, which is done by keeping the mean value, $\overline{leaving\_pp}$, within an optimal range. But this conclusion is built on the theoretical analysis and a limited number of simulation tests. A large number of tests are necessary to confirm this conclusion and provide further guidance on the setting of $\overline{leaving\_pp}$, which could minimize the errors of predicted traffic state distribution in MC model.

The chi-square test is applied here to test whether the MC model could provide the same traffic state distribution as observed from the Monte Carlo simulation. An attractive feature of the chi-square goodness-of-fit test is that it can be applied to binned data, which is exactly the form of the discrete traffic state. But the traffic state distribution in MC model describes the probability that urban network stays at a certain bin, while the chi-square test would measure the difference of the frequency. This is actually not a restriction since traffic state distribution could be converted into frequency by multiplying by the iteration number of Monte Carlo simulation, which is set as 5000 in all scenarios. For the chi-square approximation to be valid, the expected frequency of each bin should be at least 5. In MC model, some bins in the tails with expected counts less than five are combined.

The chi-square statistic is defined as the sum of the squared difference between observed (o) from Monte Carlo simulation and the expected frequency (e) from the MC model, divided by the expected frequency in all possible bins. The formula for calculating chi-square (χ^2) is:

$$\chi^2 = \sum_{i=1}^{n} (o_i - e_i)^2 / e_i \qquad \textbf{(7-3)}$$

The results of chi-square test are measured with the associated p-value obtained from a chi-square table. The p value is the probability that the deviation of the observed from that expected is due to chance alone (no other forces acting). The relative standard commonly used in the research is $p > 0.05$. In this case, using $p > 0.05$, it is expected that any deviation is due to chance alone 5% of the time or more, and the expected frequency with MC method is consistent with the simulation results at the 95% confidence level.

7.2.1 Triangular NEF Model

A series of simulation tests are designed to examine the relationship between the accuracy of MC model and $\overline{leaving\_pp}$. The tests start with the triangular NEF traffic model described with Equation 5-1. Sine it has been demonstrated that MC model could incorporate the uncertainty both in vehicle entry rate and trip complement rate, only one type of uncertainty—uncertainty in vehicle entry rate ($G_0 \neq 0$ and $G_f = 0$)—is considered in these tests for simplicity.

The bin size Δn is set as 100 [veh] and different values of $\overline{leaving\_pp}$ are achieved by varying the value of time interval Δt ranging from 3 to 360 [sec]. For instance, in a scenario such as case 1 with $E = 120{,}000$[veh/hr] and $G_0 = 424{,}000$[veh/sec], a MC model could be built with a set of Δt and Δn to predict the urban network state distribution at each moment. Then, the value of $\overline{leaving\_pp}$ could be derived using Equation 7-2, and the deviations of the expected traffic state and simulation results at each moment could be measured with chi-square tests using the index p. Here, the moment is set as the end of simulation duration which is 5 hours. Thus, MC model with multiple values of time intervals and a fixed bin would produce different pairs of $\overline{leaving\_pp}$ and p, which would form the relationship of these two variables, as illustrated in Figure 7-2.

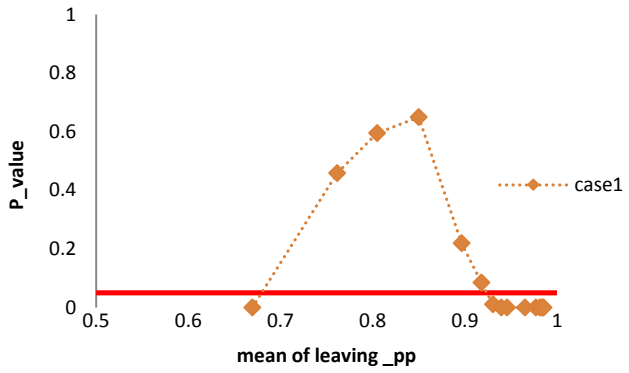


Figure 7-2. The relationships between chi-square tests results and different sets of $\overline{leaving\_pp}$ in the case 1.

Figure 7-1 illustrates that in a certain range of $\overline{leaving\_pp}$, the predicted state with MC model is in consistent with simulation result without significant deviation ($p > 0.05$). Therefore, in case 1. MC model can be optimized by controlling $\overline{leaving\_pp}$ in this optimal range by adjusting the time interval and bin size. The optimal range of $\overline{leaving\_pp}$ appears to be especially close to value 0.85; this is the value for which p is maximum, which indicates the smallest deviation between the MC model and the simulation results. But the optimal range is derived based on one scenario and should be verified with more simulation tests.

Notice that when the probability that the network becomes completely jammed is high, most vehicle state distribution would be concentrated in the n_{th} bin with the maximum vehicle accumulation n_j of the NEF model. The frequencies in the other bins in this scenario would be close to zero, see Figure 6-9d. For these jammed scenarios, the other bins in the left tail should be combined to ensure the sum of frequencies more than 5 for the chi-square test to be applicable. In such a case, there are only two bins involved in the analysis (the jammed bin and the remaining bins), and this is not an applicable situation for chi-square test. Therefore, only the cases with low

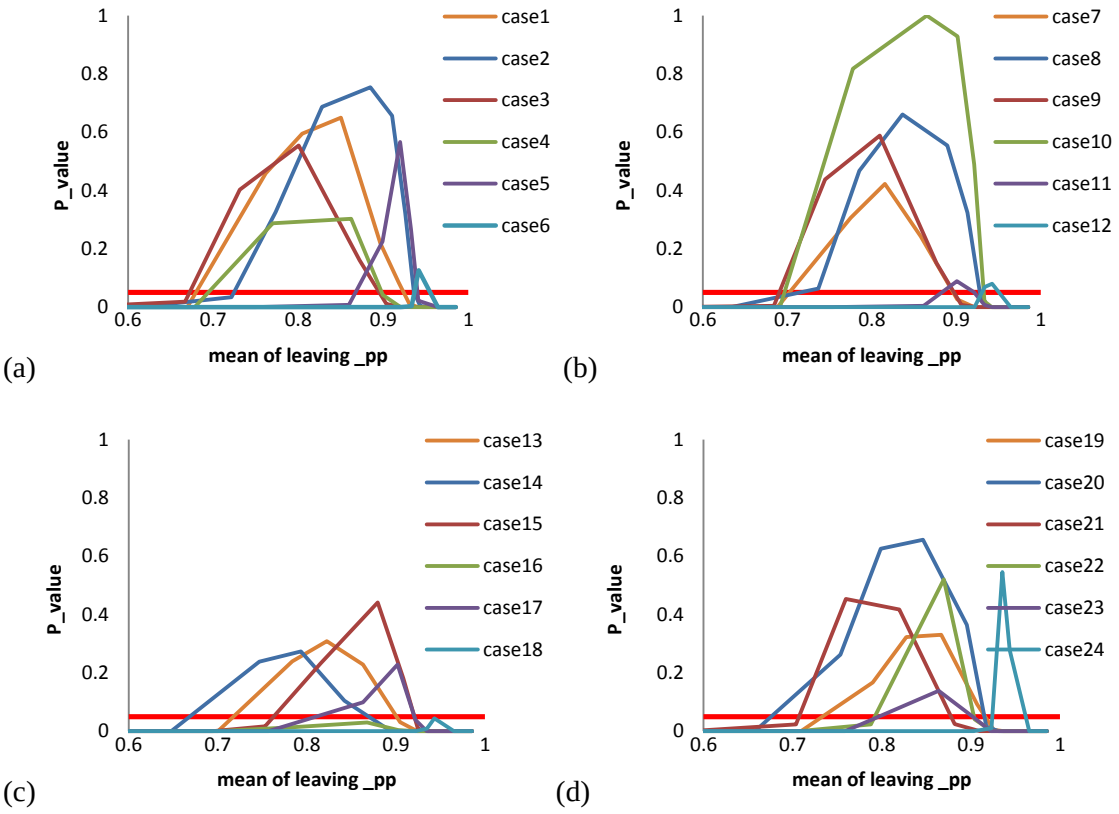
fractions of iterations that are completely jammed at the end of the simulation period (from 5% to 25%) are selected in the numerical tests.

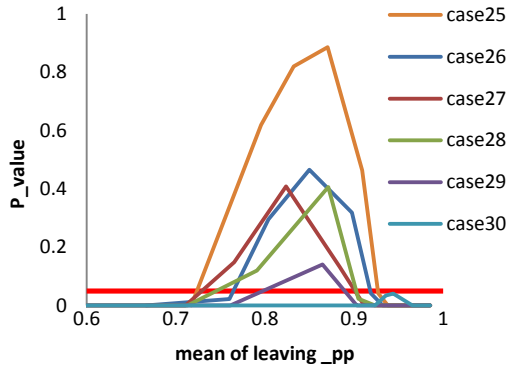
In order to identify the range of $\overline{leaving\_pp}$ that could be widely applied generally, a large number of cases are tested here. The same jam probability at the end of Monte Carlo simulations can be obtained from different combinations of E and G_0. All the parameters of traffic control used in these tests are listed in Table 7-1.

Table 7-1. Parameters of Metering Traffic Flows in the Trianguler NEF Cases

Name	Jam Probability (5hr)	E (veh/hr)	G_0 (veh/sec or 1/60veh/hr)
case 1	5%	120,000	424,000
case 2		130,000	353,000
case 3		140,000	283,000
case 4		150,000	212,000
case 5		160,000	141,000
case 6		170,000	71,000
case 7	10%	120,000	452,000
case 8		130,000	377,000
case 9		140,000	302,000
case 10		150,000	226,000
case 11		160,000	151,000
case 12		170,000	75,000
case 13	15%	120,000	472,000
case 14		130,000	394,000
case 15		140,000	315,000
case 16		150,000	236,000
case 16		160,000	158,000
case 18		170,000	79,000
case 19	20%	120,000	488,000
case 20		130,000	407,000
case 21		140,000	326,000
case 22		150,000	245,000
case 23		160,000	163,000
case 24		170,000	82,000
case 25	25%	120,000	503,000
case 26		130,000	419,000
case 27		140,000	336,000
case 28		150,000	252,000
case 29		160,000	168,000
case 30		170,000	84,000

Similar to the previous numerical tests, the urban network is assumed to be initially empty, i.e., $n_0 = 0$, and controlled with perimeter metering strategies with the parameters E and G_0 during a 5-hr peak period to avoid the congestion. In each case, multiple MC models are built by varying the time interval and maintaining a fix bin size to generate different values of $\overline{leaving\_pp}$. The validities of the predictions from the MC model are measured using the chi-square test. Thus, a series data of ($\overline{leaving\_pp}$, p) can be obtained with different models in each case, as illustrated in Figure 7-3. The minimum acceptable p-value from the tests, $p = 5\%$, is also labeled as red horizontal lines.





(e)

Figure 7-3. The relationships between chi-square tests results and different sets of leaving_pp based on the linear NEF model with (a) 5% jam probability; (b) 10% jam probability; (c) 15% jam probability; (d) 20% jam probability; and (e) 25% jam probability.

The numerical tests have shown that in most cases there exists an optimal range of $\overline{leaving\_pp}$, which satisfies $p > 5\%$. Most lines demonstrate the value of p increases at first and then descends with the maximum value around 85%, where the MC model is most effective to model the stochastic dynamic process of urban network and predict the traffic state distribution. Although those ranges are not identical and show a slightly different in different scenarios, most tests suggests that the optimal value of $\overline{leaving\_pp}$ is in the range [80% 90%].

Notice that some test results fail to verify the existence of an optimal range of $\overline{leaving\_pp}$; an example is case 16 in Figure 7-3c. The optimal solution infeasibility may arise due to the restriction of the fixed bin size used. To explore this further, smaller values of bin size were used to build more MC models and compare the results to the Monte Carlo simulations; see Figure 7-4. As shown, smaller bin sizes (or finer granularity in the description of the traffic state) provides an optimal range of $\overline{leaving\_pp}$. The size of this optimal range increases as the bin size decreases. Therefore, $\overline{leaving\_pp}$ is one of important criteria for controlling the validation of the MC models as long as small enough bins are used to describe the traffic state at a high resolution.

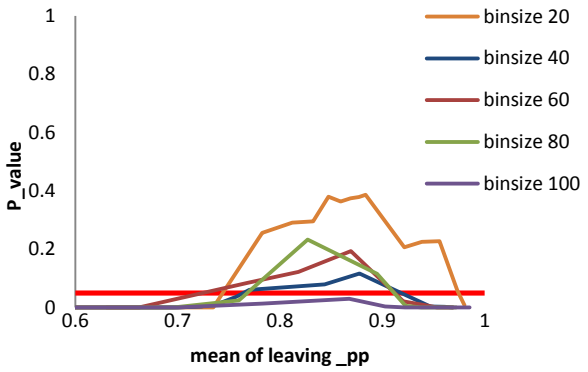


Figure 7-4. The relationships between chi-square tests results and different sets of leaving_pp based on the linear NEF model in the case 16.

7.2.2 Nonlinear NEF Model

The numerical tests are extend to include the nonlinear NEF model. Here, the nonlinear NEF model in previous studies is used, which is described using Equation 5-2. Again, only the uncertainty in vehicle entry rate is incorporated, i.e., $G_f = 0$.

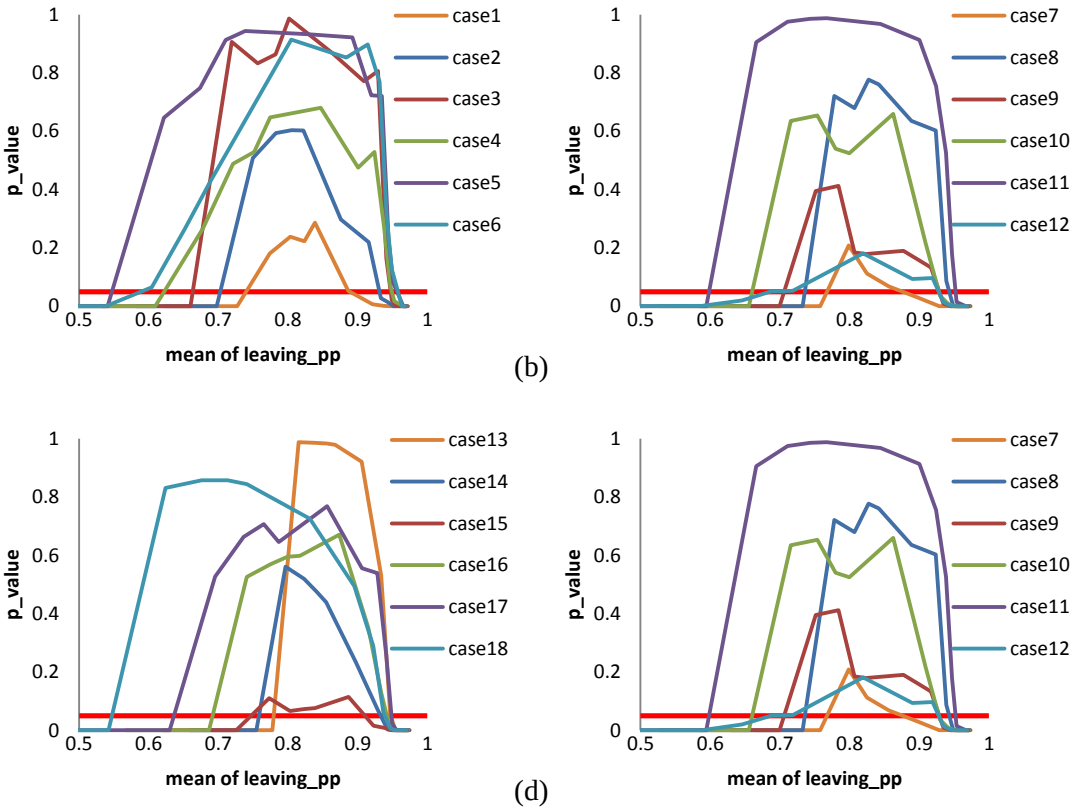
The NEF model is depicted using a set of multiple piece-wise linear segments in MC model application. Compared with the triangular model, a slight change in vehicle accumulation in nonlinear model can change the presentation of the linearized function. Therefore, Markov Chain built on the nonlinear model might demonstrate different characteristics as for the optimal range of $\overline{\_\_\_\_\_}$
$\overline{leaving\_pp}$.

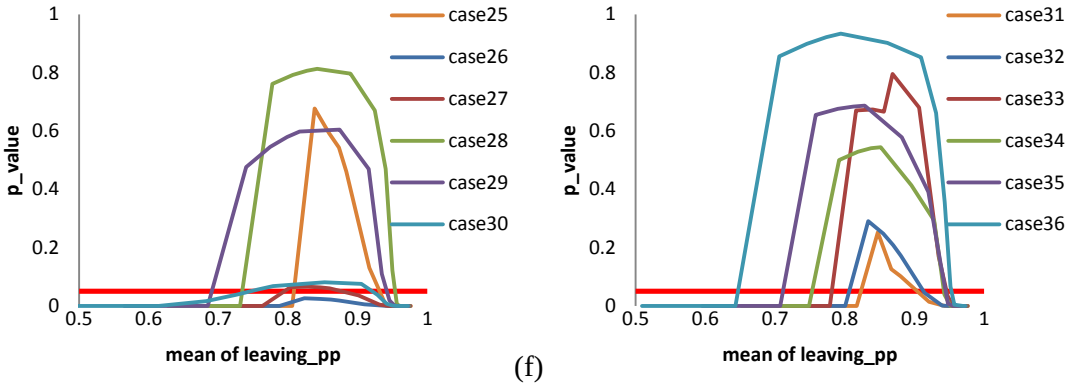
The downtown is initial empty and experiences some large travel demand during a 5-hr peak period. In order to avoid the congestion and traffic delay, the vehicles entry rate is restricted along the perimeter of downtown, which is described with E and G_0 in Table 7-2.

Table 7-2. Parameters of Metering Traffic Flows in the Nonlinear NEF Cases

Name	Jam Probability(5hr)	E(veh/hr)	G_0 (veh/sec or 1/60veh/hr)
case 1	0%	27,000	240,000
case 2		28,000	210,000
case 3		29,000	180,000
case 4		30,000	160,000
case 5		31,000	120,000
case 6		32,000	90,000
case 7	5%	27,000	329,000
case 8		28,000	294,000
case 9		29,000	258,000
case 10		30,000	220,000
case 11		31,000	181,000
case 12		32,000	138,000
case 13	10%	27,000	374,000
case 14		28,000	336,000
case 15		29,000	296,000
case 16		30,000	254,000
case 17		31,000	210,000
case 18		32,000	161,000
case 19	15%	27,000	410,000
case 20		28,000	369,000
case 21		29,000	327,000
case 22		30,000	282,000
case 23		31,000	234,000
case 24		32,000	181,000
case 25	20%	27,000	443,000
case 26		28,000	400,000
case 27		29,000	355,000
case 28		30,000	309,000
case 29		31,000	257,000
case 30		32,000	202,000
case 31	25%	27,000	474,000
case 32		28,000	430,000
case 33		29,000	384,000
case 34		30,000	334,000
case 35		31,000	280,000
case 36		32,000	222,000

In each case, variable MC models are built with a fix bin size $\Delta n = 200$ [veh] and multiple values of time interval. The scenario controlled with a static perimeter metering strategy is also modeled with Monte Carlo simulation with 50,000 iterations. To test the validation of the MC models, the deviation of expected traffic state frequency distribution and simulated result is measured with chi-square test and expressed with the index, p. The relationship of validation of the model, p and $\overline{leaving\_pp}$ in each case is illustrated in Figure 7-5.



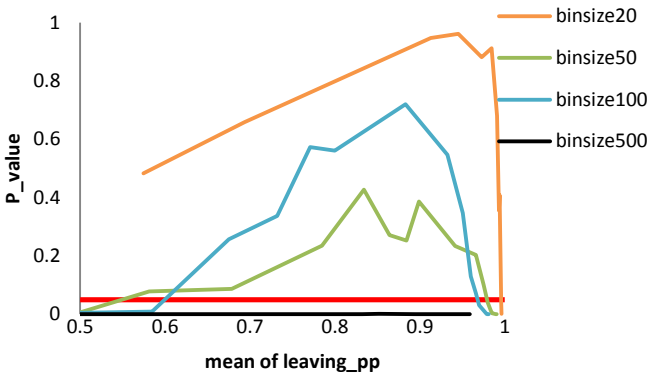


(e) (f)

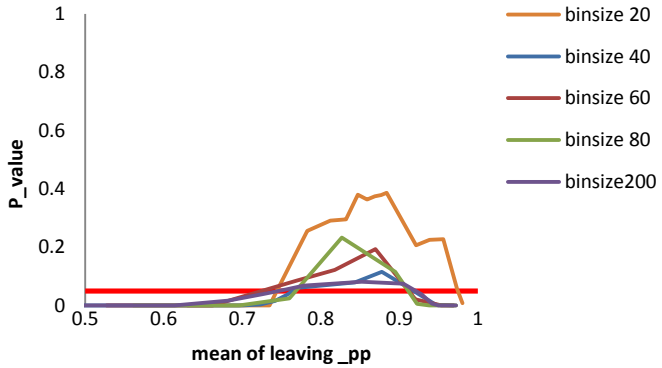
Figure 7-5. The relationships between chi-square tests results and different sets of leaving_pp based on the nonlinear NEF model with (a) 0% jam probability; (b) 5% jam probability; (c) 10% jam probability; (d) 15% jam probability; (e) 20% jam probability and (f) 25% jam probability.

Compared with the numerical tests of the trianguler NEF model, the chi-square test reuslts demonstrate the same pattern of the optimal range of $\overline{leaving\_pp}$. It shows that a stable relationship between p and $\overline{leaving\_pp}$ does exist with the optimal range around in the range of [80% 90%] and the maximium value around 85%.

Among all the scenario tests, only the cases 29 and 30 fail to generate the optimal range of $\overline{leaving\_pp}$, with all values of p less than 5%. The pehonmena might be due to the large value of the bin size. Therefore, another method is used to build the Markov chains using the multiple values of Δn and Δt. In the case 30, several lines are built with the smaller values of bin size, which could easy generate the optimal range of $\overline{leaving\_pp}$, see Figure 7-6b.



(a)



(b)

Figure 7-6. The relationships between chi-square tests results and different sets of leaving_pp based on the nonlinear NEF model in (a) case 6 and (b) case 30.

Those numerical tests confirm the existence of the optimal range of $\overline{leaving\_pp}$ $[80\% \quad 90\%]$. However, $\overline{leaving\_pp}$ is not the only criterion of MC model application. A relatively small bin size is also preferred within the allowable range of computational costs. This is because selecting a bin size that is too large would result in the large deviation between the linearized NEF model and the original model. And, the impacts incurred by the oversized Δn could not be mitigated by simply varying the time interval; e.g., the black line representing $\Delta n = 500$[veh] in the Figure 7-6a always shows a huge deviation of predicted traffic state, even when the $\overline{leaving\_pp}$ in the range of $[80\% \quad 90\%]$.

To sum up, the numerical tests on both triangular and nonlinear NEF confirm the conclusion of the parametric study that a MC model could be optimized by choosing an optimal combination of bin size and time interval. This optimal selection would be one that provides a small enough resolution of traffic states and produces an optimal range of $\overline{leaving\_pp}$. Of course, the smallest bins sizes are preferred in the tolerance of computational costs.

Chapter 8

Conclusions

This chapter provides concluding of the modelling the large-scale uncertainty in the urban network. A summary of the major findings of this work is discussed first. And then potential future work extended from this research are listed.

8.1 Summary of Major Findings

Recent advances in traffic flow theory and network modeling have provided researchers with new macroscopic traffic models that can be used to inform the design of, and study the impacts of, these network-wide traffic control strategies. The most promising approach is the Macroscopic Fundamental Diagram, which can be used to study network dynamics. Unfortunately, these models are deterministic and fail to account for the large amount of uncertainty that might arise in network dynamics due to stochastic driver behavior, data measurement error, and unpredictability of traffic control. Failure to address this uncertainty can lead to suboptimal design and control policies, which would impact the performance of these networks.

To overcome these challenges, this research explores how local traffic control can be used to reduce the uncertainty inherent in macroscopic traffic relationships, and develops a way to incorporate this uncertainty directly in the traffic network dynamics modeling. The first portion of this research has demonstrated the ability of locally adaptive traffic signals to stabilize traffic and produce more reliable MFDs. This is compared with drivers that use adaptive routing schemes, which is known to provide more reliable models but requires that drivers have sufficient

information on current traffic conditions. This studies illustrate the different stabilization mechanisms of adaptive signal control and adaptive routings on congested and unstable urban network, which provide insights of the application and control of those strategies.

The second portion develops a novel probabilistic framework that accounts for the various types of uncertainty that might arise and can be used to optimize large-scale traffic control schemes under uncertainty. The Markov Chain (MC) model is built on some assumptions to conduct the dynamic analysis of an urban network within consideration of stochasticity in driver behavior and instability in NEF traffic models. The MC method adopts the form of a matrix to depict the dynamic process of continuous change of network, and simplify the dynamic analysis. The study shows the MC method with appropriate parameters could work well as a dynamic analysis tool to predict the future traffic state and model large-scale uncertainty of urban network.

The application of the MC model requires two parameters, bin size and time interval to deal with some assumptions, and the analytical solutions with different selections of parameters should demonstrate some deviations. Some unreasonable selection of parameters would overturn these assumptions and throw the deviations out of tolerance. Fortunately, parametric analysis and tests confirm the existence of the optimal selection of those parameters with a minimum deviation. We should carefully select a relatively small bin size to carefully mimic the dynamic process involved with the NEF and select an associated time interval to provide a mean value of $leaving\_pp$ in the range $[80\% \quad 90\%]$ to guarantee the accuracy and validation of MC model.

Overall, this paper seeks a way to mitigate and analyze the large-scale uncertainty of urban network to minimize the disturbance of uncertainty to the efficiency of traffic control schemes in order to avoid the urban network congestion.

8.2 Future Work

There are several ways the work presented in this dissertation can be extended to improve the instability of urban network and decrease the damages occurred by uncertainty phenomena on the large-scale network. Some of these future research areas are:

1. Extending the MC model to multi-region urban networks. Most large cities are designed with more than one city centers and several regions that can be modeling using the MFD and NEF. Traffic agencies could limit the entry rate along the perimeter of central regions and fail to control the vehicular interchange among different regions due to the convenient transportation facilities. Therefore, the MC model should be extended to model the dynamics of urban network with multiple centers or regions. Future work may consider how to build the formulae to form the MC model and test the validation of MC model in multi-central network with a series of numerical tests.

2. Extending MC model into dynamic analysis of the dynamic metering strategy in the urban network control. This research has found that MC model has great flexibility in use and shows a promise of modeling the dynamic metering strategy with fairly low computational costs. However, the research only provides some insights to build the MC model and does not completely examine its application for the dynamic control of urban transportation networks. Future work may consider how the dynamic strategies should be incorporated in MC model, and use several different scenarios to test the accuracy of this Markov Chains.

REFERENCES

Aboudolas, K., & Geroliminis, N. (2013). Perimeter and boundary flow control in multi-reservoir heterogeneous networks. *Transportation Research Part B: Methodological*, *55*, 265-281.

Ardekani, S., Torres-Verdin, V. and Herman, H. (1985). The Two-Fluid Model and Traffic Quality in Mexico City (El Modelo Bifluido y la Calidad del Tránsito en la Ciudad de México). *Revista Ingeniería Civil.*

Ardekani, S., & Herman, R. (1987). Urban network-wide traffic variables and their relations. *Transportation Science*, *21*(1), 1-16.

Arnott, R., De Palma, A., & Lindsey, R. (1990). Economics of a bottleneck. *Journal of urban Economics*, *27*(1), 111-130.

Benesty, J., & Huang, Y. (Eds.). (2003). *Adaptive signal processing: applications to real-world problems*. Springer Science & Business Media.

Bretherton, D., Bowen, G., & Wood, K. (2003). Effective urban traffic management and control: recent developments in SCOOT. In *Transportation Research Board 82nd Annual Meeting*.

Buisson, C., & Ladier, C. (2009). Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record: Journal of the Transportation Research Board*, *2124*(1), 127-136.

Cassidy, M. J., Jang, K., & Daganzo, C. F. (2011). Macroscopic fundamental diagrams for freeway networks. *Transportation Research Record: Journal of the Transportation Research Board*, *2260*(1), 8-15.

Daganzo, C. F. (1985). The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation science*, *19*(1), 29-37.

Daganzo, C. F. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, *28*(4), 269-287.

Daganzo, C. F., & Garcia, R. C. (2000). A Pareto improving strategy for the time-dependent morning commute problem. *Transportation Science*, *34*(3), 303-311.

Daganzo, C. F. (2005). Improving city mobility through gridlock control: an approach and some ideas.

Daganzo, C. F. (2006). In traffic flow, cellular automata= kinematic waves. *Transportation Research Part B: Methodological*, *40*(5), 396-403.

Daganzo, C. F. (2007). Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, *41*(1), 49-62.

Daganzo, C. F., & Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological*, *42*(9), 771-781.

Daganzo, C. F. (2011). On the macroscopic stability of freeway traffic. *Transportation Research Part B: Methodological*, *45*(5), 782-788.

Daganzo, C. F., Gayah, V. V., & Gonzales, E. J. (2011). Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. *Transportation Research Part B: Methodological*, *45*(1), 278-288.

de Gier, J., Garoni, T. M., & Rojas, O. (2011). Traffic flow on realistic road networks with adaptive traffic lights. *Journal of Statistical Mechanics: Theory and Experiment*, *2011*(04), P04008.

de Palma, A., & Lindsey, R. (2011). Traffic congestion pricing methodologies and technologies. *Transportation Research Part C: Emerging Technologies*, *19*(6), 1377-1399.

Dinopoulou, V., Diakaki, C., & Papageorgiou, M. (2006). Applications of the urban traffic control strategy TUC. *European Journal of Operational Research*, *175*(3), 1652-1665.

Edie, L. C. (1963). Discussion of traffic stream measurements and definitions. In *Proc. 2nd Int. Symp on the Theory of Traffic FLOW: London 1963, J. Almond (editor), pp 139-154. Published by OECD*, Paris, France, 1965.

Eliasson, J., & Mattsson, L. G. (2006). Equity effects of congestion pricing: quantitative methodology and a case study for Stockholm. *Transportation Research Part A: Policy and Practice*, *40*(7), 602-620.

Eliasson, J., Hultkrantz, L., Nerhagen, L., & Rosqvist, L. S. (2009). The Stockholm congestion–charging trial 2006: Overview of effects. *Transportation Research Part A: Policy and Practice*, *43*(3), 240-250.

Gal-Tzur, A., Mahalel, D., & Prashker, J. N. (1993). *Signal design for congested networks based on metering* (No. 1398).

Gayah, V. V., & Daganzo, C. F. (2011a). Clockwise hysteresis loops in the macroscopic fundamental diagram: an effect of network instability. *Transportation Research Part B: Methodological*, *45*(4), 643-655.

Gayah, V. V., & Daganzo, C. F. (2011b). Effects of turning maneuvers and route choice on a simple network. *Transportation Research Record: Journal of the Transportation Research Board*, *2249*(1), 15-19.

Gayah, V. V., & Daganzo, C. F. (2012). Analytical capacity comparison of one-way and two-way signalized street networks. *Transportation Research Record: Journal of the Transportation Research Board*, *2301*(1), 76-85.

Gayah, V. V., & Dixit, V. V. (2013). Using mobile probe data and the macroscopic fundamental diagram to estimate network densities. *Transportation Research Record: Journal of the Transportation Research Board*, *2390*(1), 76-86.

Gelb, A. (Ed.). (1974). *Applied optimal estimation*. MIT press.

Geroliminis, N., & Daganzo, C. F. (2007). Macroscopic modeling of traffic in cities. In *TRB 86th annual meeting* (No. 07-0413).

Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, *42*(9), 759-770.

Geroliminis, N., & Levinson, D. M. (2009). Cordon pricing consistent with the physics of overcrowding. In *Transportation and Traffic Theory 2009: Golden Jubilee* (pp. 219-240). Springer US.

Geroliminis, N., & Sun, J. (2011a). Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks. *Transportation Research Part A: Policy and Practice*, *45*(9), 966-979.

Geroliminis, N., & Sun, J. (2011b). Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological*, *45*(3), 605-617.

Geroliminis, N., & Boyaci, B. (2012). The effect of variability of urban systems characteristics in the network capacity. *Transportation Research Part B: Methodological*, *46*(10), 1607-1623.

Geroliminis, N., Haddad, J., & Ramezani, M. (2013). Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. *Intelligent Transportation Systems, IEEE Transactions on*, *14*(1), 348-359.

Ghamann, R., Gettman, D., & Shelby, S. (2004). ACS Lite Project Overview. In *TRB Adaptive Traffic Signal Control Workshop, Washington, DC*.

Gipps, P. G. (1981). A behavioral car-following model for computer simulation. *Transportation Research Part B: Methodological*, *15*(2), 105-111.

Godfrey, J. W. (1969). The mechanism of a road network. *Traffic Engineering & Control*, *8*(8).

Goh, M. (2002). Congestion management and electronic road pricing in Singapore. *Journal of Transport Geography*, *10*(1), 29-38.

Gonzales, E. J., & Daganzo, C. F. (2012). Morning commute with competing modes and distributed demand: User equilibrium, system optimum, and pricing. *Transportation Research Part B: Methodological*, *46*(10), 1519-1534.

Greenshields, B. D., Channing, W., & Miller, H. (1935). A study of traffic capacity. In *Highway research board proceedings* (Vol. 1935). National Research Council (USA), Highway Research Board.

Haddad, J., & Geroliminis, N. (2012). On the stability of traffic perimeter control in two-region urban cities. *Transportation Research Part B: Methodological*, *46*(9), 1159-1176.

Haddad, J., Ramezani, M., & Geroliminis, N. (2012). Model predictive perimeter control for two-region urban cities. In *Transportation Research Board 91st Annual Meeting* (No. 12-1445).

Helbing, D. (2009). Derivation of a fundamental diagram for urban traffic flow. *The European Physical Journal B-Condensed Matter and Complex Systems*, *70*(2), 229-241.

Herman, R., & Ardekani, S. (1984). Characterizing traffic conditions in urban areas. *Transportation Science*, *18*(2), 101-140.

Herman, R., & Prigogine, I. (1979). A two-fluid approach to town traffic. *Science*, *204*(4389), 148-151.

Jin, W. L., Gan, Q. J., & Gayah, V. V. (2013). A kinematic wave approach to traffic statics and dynamics in a double-ring network. *Transportation Research Part B: Methodological*, *57*, 114-131.

Kushner, H. J. (1990). Numerical methods for stochastic control problems in continuous time. *SIAM Journal on Control and Optimization*, *28*(5), 999-1048.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., & Papageorgiou, M. (2012a). Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research Part B: Methodological*, *46*(10), 1393-1403.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., & Papageorgiou, M. (2012b). Congestion control in urban networks via feedback gating. *Procedia-Social and Behavioral Sciences*, *48*, 1599-1610.

Keyvan-Ekbatani, M., Papageorgiou, M., & Papamichail, I. (2013a). Urban congestion gating control based on reduced operational network fundamental diagrams. *Transportation Research Part C: Emerging Technologies*, *33*, 74-87.

Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., & Papageorgiou, M. (2013b). Traffic signal perimeter control with multiple Boundaries for large urban networks. In *IEEE Intelligent Transportation Systems Conference*.

Keyvan-Ekbatani, M., Papageorgiou, M., & Papamichail, I. (2014). Perimeter Traffic Control via Remote Feedback Gating. *Procedia-Social and Behavioral Sciences*, *111*, 645-653.

Knoop, V. L., Hoogendoorn, S. P., & Van Lint, J. W. (2012). Routing strategies based on macroscopic fundamental diagram. *Transportation Research Record: Journal of the Transportation Research Board*, *2315*(1), 1-10.

Knoop, V. L., De Jong, D. M., & Hoogendoorn, S. P. (2014). The influence of the road layout on the Network Fundamental Diagram (poster). In*93rd Annual Meeting Transportation Research Board, Washington, USA, 12-16 January 2014*.

Laih, C. H. (1994). Queueing at a bottleneck with single-and multi-step tolls. *Transportation Research Part A: Policy and Practice*, *28*(3), 197-208.

Lago, A., & Daganzo, C. F. (2007). Spillovers, merging traffic and the morning commute. *Transportation Research Part B: Methodological*, *41*(6), 670-683.

Leclercq, L., Chiabaut, N., & Trinquier, B. (2014). Macroscopic Fundamental Diagrams: A cross-comparison of estimation methods. *Transportation Research Part B: Methodological*, *62*, 1-12.

Li, Y., Xu, J., & Shen, L. (2012). A Perimeter Control Strategy for Oversaturated Network Preventing Queue Spillback. *Procedia-Social and Behavioral Sciences*, *43*, 418-427.

Lighthill, M. J., & Whitham, G. B. (1955). On kinematic waves. II. A theory of traffic flow on long crowded roads. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* (Vol. 229, No. 1178, pp. 317-345). The Royal Society.

Litman, T. (2006). London congestion pricing: Implications for other cities. *Victoria Transport Policy Institute*, *10*.

Luk, J., & Green, D. (2010). Balancing traffic density in a signalized network. *Austroads Research Report* AP-R369/10, Sydney, Australia.

Marchand, M. (1968). A note on optimal tolls in an imperfect environment. *Econometrica: Journal of the Econometric Society*, 575-581.

Maruyama, T., & Sumalee, A. (2007). Efficiency and equity comparison of cordon-and area-based road pricing schemes using a trip-chain equilibrium model. *Transportation Research Part A: Policy and Practice*, *41*(7), 655-671.

Mazloumian, A., Geroliminis, N., & Helbing, D. (2010). The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *368*(1928), 4627-4647.

Mahmassani, H. S., Saberi, M., & Zockaie, A. (2013). Urban network gridlock: Theory, characteristics, and dynamics. *Transportation Research Part C: Emerging Technologies*, *36*, 480-497.

Nagle, A. S., & Gayah, V. V. (2014). Accuracy of Networkwide Traffic States Estimated from Mobile Probe Data. *Transportation Research Record: Journal of the Transportation Research Board*, *2421*(1), 1-11.

Nagel, K., & Schreckenberg, M. (1992). A cellular automaton model for freeway traffic. *Journal de physique I*, *2*(12), 2221-2229.

Ortigosa, J., Menendez, M., & Tapia, H. (2014). Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich. *EURO Journal on Transportation and Logistics*, *3*(3-4), 245-266.

Ortigosa, J., & Menendez, M. (2014). Traffic performance on quasi-grid urban structures. *Cities*, *36*, 18-27.

Panayotounakos, D. E., & Zarmpoutis, T. I. (2011). Construction of exact parametric or closed form solutions of some unsolvable classes of nonlinear ODEs (Abel's Nonlinear ODEs of the first kind and relative degenerate equations). *International Journal of Mathematics and Mathematical Sciences, 2011*.

Papageorgiou, M., Hadj-Salem, H., & Blosseville, J. M. (1991). ALINEA: A local feedback control law for on-ramp metering. *Transportation Research Record*, (1320).

Pigou, A. C. (1912). *Wealth and welfare*. Macmillan and Company, limited.

Prud'homme, R., & Bocarejo, J. P. (2005). The London congestion charge: a tentative economic appraisal. *Transport Policy*, *12*(3), 279-287.

Richards, P. I. (1956). Shock waves on the highway. *Operations research*, *4*(1), 42-51.

Rostami, A. B. (2015). Exact Solution of Abel Differential Equation with Arbitrary Nonlinear Coefficients. *arXiv preprint arXiv:1503.05929.*

Rotaris, L., Danielis, R., Marcucci, E., & Massiani, J. (2010). The urban road pricing scheme to curb pollution in Milan, Italy: Description, impacts and preliminary cost–benefit analysis assessment. *Transportation Research Part A: Policy and Practice*, *44*(5), 359-375.

Saberi, M., & Mahmassani, H. S. (2012). Exploring properties of networkwide flow-density relations in a freeway network. *Transportation Research Record: Journal of the Transportation Research Board*, *2315*(1), 153-163.

Saberi, M., & Mahmassani, H. S. (2013). Empirical characterization and interpretation of hysteresis and capacity drop phenomena in freeway networks. *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington, DC*.

Saberi, M., Mahmassani, H. S., & Zockaie, A. (2014). Network capacity, traffic instability, and adaptive driving: findings from simulated urban network experiments. *EURO Journal on Transportation and Logistics*, *3*(3-4), 289-308.

Schrank, D., Eisele, B., & Lomax, T. (2012). TTI's 2012 urban mobility report. *Texas A&M Transportation Institute. The Texas A&M University System*.

Schroeder, M. (1989). State-of-the-art review of adaptive arterial control systems, *Research Project GC 8286, Task 30, Arterial Control and Integration, Washington State Transportation Center, University of Washington instability. Transportation Research Part B: Methodological 45.4 (2011):* 643-655.

Small, K. A., & Gómez-Ibáñez, J. A. (1997). Road pricing for congestion management: the transition from theory to policy. *Transport Economics*, 373-403.

Small, K. A., & Yan, J. (2001). The value of "value pricing" of roads: Second-best pricing and product differentiation. *Journal of Urban Economics*, *49*(2), 310-336.

Smeed, R. J., & Wardrop, J. G. (1964). *An exploratory comparison of the advantages of cars and buses for travel in urban areas*.

Smeed, R. J. (1967). The road capacity of city centers. *Highway Research Record*, (169).

Smith, M. J. (1984). The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation science*, *18*(4), 385-394.

Thomson, J.M. (1967). Speeds and flows of traffic in Central London: 2. Speed-flow relations. *Traffic Engineering and Control 8 (12)*, 721–725.

Verhoef, E. T. (2002). Second-best congestion pricing in general networks. Heuristic algorithms for finding second-best optimal toll levels and toll points. *Transportation Research Part B: Methodological*, *36*(8), 707-729.

Vickrey, W. (1963). Pricing and resource allocation in transportation and public utilities. *American Economic Review*, *53*(2), 452-465.

Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review*, 251-260.

Wang, B. H., & Wamg, W. X. (2008). Routing strategies in traffic network and phase transition in network traffic flow. *Pramana*, *71*(2), 353-358.

Wang, W. X., Yin, C. Y., Yan, G., & Wang, B. H. (2006). Integrating local static and dynamic information for routing traffic. *Physical Review E*, *74*(1), 016101.

Wardrop, J.G. (1968). Journey speed and flow in central urban areas. *Traffic Engineering and Control* 9 (11), 528–532.

Wood, K., Bretherton, D., Maxwell, A., Smith, K., & Bowen, G. (2002). Improved Traffic Management and Bus Priority with SCOOT. *TRL STAFF PAPER PA 3860/02*.

Wu, X., Liu, H. X., & Geroliminis, N. (2011). An empirical analysis on the arterial fundamental diagram. *Transportation Research Part B: Methodological*, *45*(1), 255-266.

Yang, H., & Huang, H. J. (1998). Principle of marginal-cost pricing: how does it work in a general road network? *Transportation Research Part A: Policy and Practice*, *32*(1), 45-54.

Yang, H., & Huang, H. J. (2005). *Mathematical and economic theory of road pricing*.

Zahavi, Y. (1972). Traffic performance evaluation of road networks by the α-relationship. *Traffic Engineering and Control*, *14*(5-6).

Zhang, L., Garoni, T. M., & de Gier, J. (2013). A comparative study of macroscopic fundamental diagrams of arterial road networks governed by adaptive traffic signal systems. *Transportation Research Part B: Methodological*, *49*, 1-23.

Zhang, J., Tan, K. C., & Chai, K. H. (2003). Systematic innovation in service design through TRIZ. *The TRIZ Journal*, 6-18.

Zheng, N., Waraich, R. A., Axhausen, K. W., & Geroliminis, N. (2012). A dynamic cordon pricing scheme combining the Macroscopic Fundamental Diagram and an agent-based traffic model. *Transportation Research Part A: Policy and Practice*, *46*(8), 1291-130.

VITA

Xueyu Gao

xug103@psu.edu

EDUCATION

Ph.D., Civil and Environmental Engineering (Transportation) graduated Dec. 2015
Pennsylvania State University (University Park)
Advisor: Vikash V. Gayah
Dissertation: Modeling Uncertainty in Large-scale Urban Traffic Networks

M.S., School of Civil Engineering (Transportation) graduated May 2012
Shandong University (China)
Thesis: Traffic Safety in Temporary Work Zone

B.S., School of Civil Engineering graduated May 2010
Shandong University (China)

EXPERIENCE

Modeling Uncertainty in Large-scale Urban Traffic Networks
- Derived the analytical solution of stochastic dynamic process of urban traffic with Markov Chain model and Macroscopic Fundamental Diagram
- Built a probabilistic framework that accounts for the uncertainties that arise on a macroscopic scale

Optimization of a static metering rate for an urban traffic network
- Dynamic analysis of a static metering strategy and a dynamic control
- Optimized a static metering rate when the metering control is subject to uncertainty

Impact Analysis of Locally Adaptive Traffic Signals on Urban Network Stability
- Built models with idealized and more realistic simulation with Java, MATLAB and AIMSUN
- Investigated and compared the effects of adaptive signal to mitigate instability on congested network with adaptive driver routing
- Explored and explained efficiency of adaptive signals on moderate congestion of urban network

PUBLICATION

- Keyvan-Ekbatani, M., Gao, X., Gayah, V.V., Knoop, V.L. (2015). Combination of Traffic-Responsive and Gating Control in Urban Networks: Effective Interactions. Submitted to TRB 2016, Washington DC, USA.
- Gayah, V. V., Gao, X. S., & Nagle, A. S. (2014). On the impacts of locally adaptive signal control on urban network stability and the Macroscopic Fundamental Diagram. Transportation Research Part B: Methodological, 70, 255-268.
- Gayah, V. V., & Gao, X. S. (2014). The Effect of Adaptive Green Duration Control on the Macroscopic Fundamental Diagram. In Transportation Research Board 93rd Annual Meeting (No. 14-0443).
- Shi, B. Y., Gao, X. Y., Ge, Z., & Ma, X. P. (2011). Simulation and Analysis of the Variable Speed Limit Controls on Highway Maintenance. In Applied Mechanics and Materials (Vol. 97, pp. 435-439).