Title
Increasing mobility in cities by controlling overcrowding

Author
Geroliminis, Nikolaos

Publication Date
2007-12-01

Increasing mobility in cities by controlling overcrowding

Nikolaos Geroliminis

DISSERTATION SERIES
UCB-ITS-DS-2007-2

Fall 2007
ISSN 0192 4109

Increasing mobility in cities by controlling overcrowding

by

Nikolaos Geroliminis

PTYCHION (National Technical University of Athens, Greece) 2003
M.S. (University of California, Berkeley) 2004

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering-Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carlos F. Daganzo, Chair
Professor Alexander Skabardonis
Professor Robert B. Cervero

Fall 2007

The dissertation of Nikolaos Geroliminis is approved:

Chair _____ Date_____

_____ Date_____

_____ Date_____

University of California, Berkeley

Fall 2007

Abstract

Increasing mobility in cities by controlling overcrowding

by

Nikolaos Geroliminis

Doctor of Philosophy in Engineering-Civil and Environmental Engineering

University of California, Berkeley

Professor Carlos F. Daganzo, Chair

Various theories have been proposed to describe vehicular traffic movement in cities on an aggregate level. They fall short to create a macroscopic model with variable inputs and outputs that could describe a rush hour dynamically. This dissertation work shows that a macroscopic fundamental diagram (MFD) relating production (the product of average flow and network length) and accumulation (the product of average density and network length) exists for neighborhoods of cities in the order of 5-10km^2. It also demonstrates that conditional on accumulation large networks behave predictably and independently of their oigin-dstination tables. These results are based on analysis using simulation of large scale city networks and real data from urban metropolitan areas. The real experiment uses a combination of fixed detectors and floating vehicle probes as sensors. The analysis also reveals a fixed relation between the space-mean flows on the whole network and the trip completion rates, which dynamically measure accessibility. This work also demonstrates that the dynamics of the rush hour can be predicted quite accurately without the

knowledge of disaggregated data. This MFD is applied to develop perimeter control strategies based on neighborhood accumulation and speeds and improve accessibility without the uncertainty inherent in today's forecast-based approaches. The looking-for-parking phenomenon that extends the average trip length is also integrated in the dynamics of the rush hour.

_____

Carlos F. Daganzo
Dissertation Committe Chair

To my parents with all my love.
They made it all possible.

Contents

List of figures

Acknowledgements

I am greatly indebted to my dear mentor and advisor Carlos Daganzo. His guidance, encouragement, interest, enthusiasm and wisdom made this work possible. The innumerable hours I have spent in discussions with him are irreplaceable and inimitable. He has taught me things that nobody else could and more importantly, I have learned how to ask and answer the right questions from Carlos.

I am also grateful to my dear teacher and friend Alexander Skabardonis for all the knowledge he transferred to me during my years in Berkeley. His enduring advice, thoughtful suggestions and spontaneous support have shaped my career.

My gratitude is extended to the distinguishable pool of professors with whom I had the chance to collaborate during my time in Berkeley, especially, Pravin Varaiya, whose unsolicited support and guidance provided me with a broader prospective to face difficult problems with simple logic; Samer Madanat who taught me in the art of teaching and shared with me a wide variety of academic and non-academic interests. I would like also to thank Michael Cassidy, Robert Cervero, Adib Kanafani and Alexander Bayen for their critique and comments during this dissertation research. I am grateful to Professor Masao Kuwahara from the University of Tokyo for providing me the data analyzed in chapter 3 of this thesis. I am also indebted to Matthew Karlaftis, transportation professor in the National Technical University of Athens, for initially fostering my interest in transportation studies and encouraging me to attend graduate school at Berkeley.

Studying at Berkeley has given me the opportunity to work with an amazing group of fellow students. I would like to acknowledge a number of students who have been both colleagues and friends: Stella So, Ipsita Banerjee, Offer Grembek, Hwasoo Yeo, Tasos Nikoleris, Josh Pilachowski, Anthony Patire and Eric Gonzales. Above all, I should mention Gautam Gupta for his deep friendship during all these years.

Closer to home, I am grateful to my parents, Vasiliki and Anastasios, for enfolding me with love and for strongly encouraging me to pursue my goals.

Finally, I want to thank my love, Kalliopi (a Muse in ancient Greek Mythology), for her unwavering love and support the last three years. She has been my inspiration and without her patience and tenderness I would not have been able to achieve this goal.

This research was supported by the University of California, Berkeley's Center for Future Urban Transport (a Volvo International Center of Excellence). Special thanks to the University of California Transportation Center (UCTC) and Partners for Advanced Transit and Highways (PATH) for financial support throughout my studies.

Chapter 1

Introduction

This dissertation seeks to shed some light in the macroscopic modeling of traffic flow for overcrowded cities. Understanding the physics of congestion is a necessary first step towards developing policies for mobility improvements, such as pricing and control.

Traffic congestion is increasing in major cities. The Texas Transportation Institute (2005) estimated that in 2000 the 75 largest US metropolitan areas experienced 3.6 billion vehicle-hours of delay, resulting in 5.7 billion gallons in wasted fuel and $67.5 billion in lost productivity. Furthermore, between 1980 and 1999, the total length of roads in the United States increased by only 1.5 percent, while the total number of miles of vehicle travel increased by 76 percent (FHWA, 2004). The construction of new infrastructure is not a feasible solution to decreasing congestion, not only because of the tremendous cost to keep pace with population increases and the resulting increase in travel demand, but also because of the phenomenon of induced demand (Small, 2004).

From the above, it is apparent that to decrease congestion in large cities and improve urban mobility society has to focus on better utilization of the existing

infrastructure. To do this models of traffic in cities are needed. Existing models can be coarsely divided in two main categories: microscopic and macroscopic.

For the last 50 years, the development and evaluation of transportation policies for urban mobility improvement has relied heavily on forecasting models and the trend has been toward microscopic models. Given correct inputs, modern computer models can produce much detailed information on a disaggregate basis (link flows, travel time in each route etc) on any large transportation network. Unfortunately, this approach turns out to be quite fragile when dealing with congested systems.

Firstly, these models require disaggregated time-dependent origin-destination (O-D) tables. But for a model with a reasonable spatio-temporal resolution the entries of the time dependent O-D tables are so numerous that in some cases they can exceed the whole population of a city. For example, a metropolitan area with 4 million people could be modeled with 400 zones (approximately 10000 people per zone). To model the morning and evening peak hour in 24 time intervals the total number of entries would have to be almost equal to the population of the city. Moreover, even if this information was available, the stochasticity and variations in individual driver behavior would make the estimation of derived vehicle movements unreliable.

To further complicate matters, even if we were able to predict this huge amount of input information, detailed performance measures such as vehicle hours of travel on specific links would be hard to predict because the outputs of congested systems are hyper-sensitive to the inputs. It is known that small perturbations to O-D tables or small changes to drivers' route choices can drastically change link flows; see Daganzo (1998).

We conclude that existing microscopic models for large scale urban networks are not a panacea to deal with crowded conditions.

Despite the intense complexity of a city's transportation system, averages of pertinent variables could be functionally related as a consequence of collective effects. The idea of describing a complex physical system with averages of variables is not new. For a simple physical system consisted of few particles that are in interaction the dynamical laws can describe accurately the state of the system in a future time. But, physical systems containing many particles, complex to be described in details, can often been characterized by averages of few variables (such as pressure or temperature in the Ideal Gases Law). We shall show that the same idea applies to urban traffic systems. Perhaps, a macroscopic approach could offer more insights. After all, the research on this dissertation pursues this idea: we develop observation-based models that alleviate the fragility problems of traditional models. In this approach, monitoring replaces prediction, and the system is repeatedly modified based on observations. To succeed this goal a city is modeled at an aggregated manner and relations between state variables are developed. Macroscopic feedback control strategies are introduced which rely on real-time observation of relevant spatially aggregated measures of traffic performance.

1.1 Macroscopic modeling background

The literature review that follows examines past efforts with macroscopic approach, but it is not an exhaustive examination. Rather, it illustrates the general approaches that have been used and focuses on the limitations that will be addressed by this dissertation.

Much of the work over the past 40 years has been empirical: observing and examining general traffic characteristics among various cities and using these data to estimate an assumed functional dependence among network-wide traffic variables, such as average speed, road density, average signal spacing etc. Existing theories of this class propose a variety of macroscopic traffic relations but the results are tentative since they are based on few data points.

Early work by Wardrop (1952) and Smeed (1968) dealt with the development of macroscopic models for arterials, which were later extended to general networks. Smeed (1966) theorized based on dimensional analysis that the maximum flow that can enter the central area of a city should be a function of the area of the city, the fraction devoted to roads and the capacity of the roads, expressed in vehicles per unit time per unit width of road. Although seminal, this work does not say what happens to speeds and trip completion rates when demand exceeds capacity during a rush hour.

Thomson (1967) found from data collected from streets in central London for many years that there seemed to be a linear-decreasing relationship between average speed and flow.[1] On a more theoretical approach, Wardrop (1968) proposed a generic relation between average speed and flow, which depended on average street width and average intersection spacing, but it still decreased monotonically. Zahavi (1972) analyzed relations for various cities in United Kingdom and United Stated by combining data across different regions of a city for the same time period (one day or peak period), and proposed that speed was inversely related to flow; i.e., still monotonically. Monotonicity only makes sense if traffic is light, since it cannot capture crowded states with very low

[1] This suggests that the streets in the data set were not very congested.

speeds and flows; e.g., approaching gridlock. Thus, these models cannot be used to describe the rush hour in a congested city.

A helpful work which establishes macroscopic relations in vehicular traffic in large cities is the "two-fluid model of town traffic" by Herman and Prigogine (1979). The authors assumed that the speed distribution splits into two parts; one part corresponding to moving vehicles and the other to the vehicles that are stopped due to local conditions as congestion, traffic control devices, accidents etc., but not parked cars since there are not components of the moving traffic. The basic postulate of the theory relates the average speed of the moving vehicles v_r, to the fraction of moving vehicles, f_r as follows:

$$v_r = v_m \cdot f_r^k, \tag{1.1}$$

where v_m is the average maximum running speed and k is a parameter denote the quality of traffic service in the network. The parameter v_m can be interpreted as the average speed of a test vehicle moving randomly in the network without any interference with other vehicles, but stopping only because of the control devices.

The two-fluid model assumes that a city traffic network is ergodic. In other words, the fraction of stopped time of a single vehicle circulating in the network over a sufficiently long period of time is equal to the mean fraction of the stopped vehicles over the same period of time f_s. Under this assumption the following relation is obtained in terms of average travel times:

$$T_r = T_m^{\frac{1}{k+1}} \cdot T^{\frac{k}{k+1}}, \tag{1.2}$$

where T_r and T_m are the inverse of v_r and v_m respectively and T is the average total trip time (including stopped time T_s), all per unit distance. The log transform of equation (1.2)

allows the least square estimation of k and T_m, given observations of stopped and moving time in the network.

Herman and Ardekani (1984) further developed and validated the two-fluid model for data collected in Austin, Houston and other US cities with chase cars following randomly selected cars in designated networks. Later, Herman et al. (1988) studied the influence of driver behavior in the two-fluid parameters. They found that the two-fluid model is sensitive to driver's behavior (aggressive vs. conservative drivers).

Williams et al. (1987) and Mahmassani et al. (1987) investigated through simulation, network level relationships between the three fundamental variables of traffic flow, speed, flow and density. To derive these relationships they assumed an additional relation between average fraction of vehicles stopped from the two-fluid theory and the network concentration (density).

One aspect of the two-fluid model that has not been analyzed yet is its validity under different O-D demands. If these relationships are valid and insensitive they could be realized as network properties and be used as an important tool to monitor a large system and test different control strategies. Also, this model allowed for a more realistic representation of crowded conditions in the steady state, but the idea was not sufficiently developed to create a macroscopic model with variable inputs and outputs that could describe a rush hour dynamically.

More recently, Daganzo (2007) proposed a simpler and more general model that could bridge this gap. The initial conjecture is that if a network is homogeneously loaded and congestion is uniformly distributed over the network, then travel production P, (the number of veh-km traveled per unit time) under steady state conditions can be expressed

as a function of the total accumulation of the network, n, independently of the disaggregate link data:

$$\sum P_i(n_i) \cong P\left(\sum n_i\right) \equiv P(n), \tag{1.3}$$

where P_i is the travel production for the individual link i with accumulation n_i. While the 1970 models yield monotonic curves between travel production and accumulation (see figure 1.1), the new proposal was that this relationship (which we call the macroscopic fundamental diagram—MFD) has 3 different regimes, as it happens for a single link: Uncongested when few vehicles are there (regime I), congested when n is large (regime III) and capacity (regime II) when n is in a "sweet-spot" accumulation range.

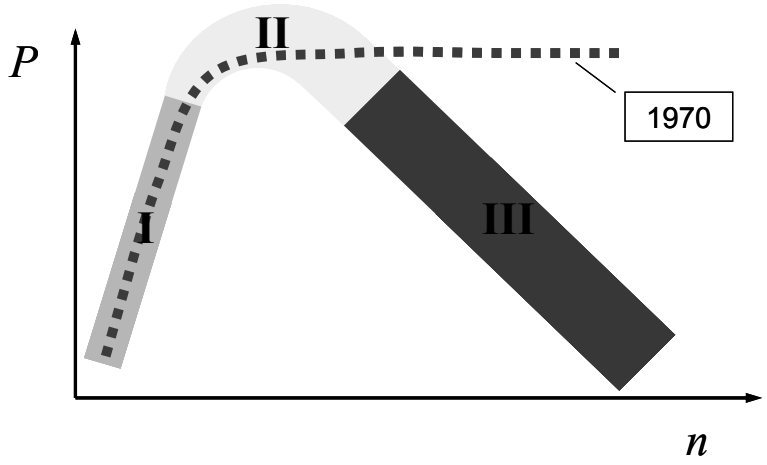


Figure 1.1: The shape of the MFD

The same reference showed that if production-accumulation conjecture were true, and drivers were to make trips of random lengths with a given average, l, then drivers would be exiting the network at a rate $O(n)=P(n)/l$. It was conjectured that since drivers naturally flock to underused parts of a network, entrances and exits should not disturb the distribution of speed and that city traffic could be treated macroscopically as a single-

neighborhood dynamic (queueing) system with n as a single state variable. The dynamics of the system are described by this ordinary differential equation:

$$\frac{dn}{dt} = f(t) - O(n(t)), \quad \text{for } t \geq 0. \tag{1.4}$$

where $f(t)$ is the input flow to the system. Daganzo (2007) also showed that these types of queueing systems exhibit a form of congestion (called gridlock) that can be eliminated with equitable control policies that require no demand data if the state can be observed. The general idea of this policy is to "meter" the input flow to the system and holding vehicles outside the system if necessary. In Daganzo (2007), the model has never been tested. Therefore, whether it is a good representation of the real world is unknown. In particular, it is not known if the MFD exists, and if so, whether it is a property of the city's infrastructure or also depends on the demand characteristics. We also do not know whether the $O(n)/P(n)$ relation is stable over time. On the theoretical side, the dynamics of the rush hour and control strategies will be developed for cities that are not uniformly congested. This will be done by introducing a multi-neighborhood model. The thesis will also discuss how to estimate the models using existing technologies. These experimented questions will be addressed in this dissertation.

There are some examples of cities around the world, where macroscopic control strategies have been applied. Therefore, the results of this thesis can help refine these efforts. For example, a novel traffic light operating system designed for active management of the high demand central area was developed and set up in Zurich. The traffic lights are internally coordinated based on the logic that the flow towards an overloaded area should be restricted while the flow towards an underutilized area should

be promoted. Prevention of overcrowding on the city centre is achieved by metering of access to maintain the mobility of cars at a stabilized level. For example, longer red times are applied in the periphery of the centre during peak hours for phases which direct vehicles to the centre (Joos, 2000). A congestion charging scheme was introduced in central London, covering 22 km^2, in 2003. The main aims of the scheme are to reduce traffic congestion in and around the charging zone, to improve the bus services, journey time reliability for car users and to make the distribution of goods and services more reliable, sustainable and efficient. The reductions in both traffic and congestion that had been observed in the charging zone are around 30 percent. There are 65,000 fewer car trips into or through the zone per charging day as result of the scheme. The reliability of buses in and around the charging zone had also improved significantly (Transport for London, 2004). The models in this thesis can predict the outcomes of policy actions, such as those in Zurich and London with more accuracy than previously provided.

1.2 Dissertation overview

1.2.1 Main contributions

The main contributions of this research include:

- Demonstration of existence and characterization of a macroscopic fundamental diagram for two cities with different demand profiles.
- Development of the invariant relationship between production and output for these cities.
- Development of a methodology to estimate the macroscopic fundamental diagram by combining data from fixed detectors and floating vehicle probes.

- Development of the dynamics for cities with an inhomogeneous structure.
- Development of the first model of cruising-for-parking with realistic physics of overcrowding. This model shows that the effects for users with destinations outside the “cruising region” are significant, and provides tools to estimate the total system delays due to parking effect.
- Development of perimeter control strategies for cities with complicated structures, and the study of their effectiveness and limitations.

1.2.2 Organization

The thesis is organized in a series of self-contained chapters. Chapter 2 scrutinizes (i) if an MFD exists for one-dimensional arterials and urban networks, using micro-simulation tools, and (ii) if this MFD is sensitive to different demand patterns. It also examines how the dynamics of the rush hour for simple systems, can be predicted without the knowledge of disaggregated link by link measures and O-D tables. Chapter 3 designs and analyzes a natural experiment with real traffic in Yokohama city (Japan) to ascertain whether a MFD exists in that city. It also explains how data from the fleet of taxi cabs can be fused with detector data to obtain consistent city-wide estimates of average speeds, flows and densities and how an MFD can be estimated it in other cities. Chapter 4 develops the dynamics of cities with more complicated structures. It also presents a model, consistent with the physics of traffic, for the cruising-for-parking phenomenon and provides tools to estimate the delays for different types of users. Chapter 5 extends the analysis of chapter 4 to provide perimeter control strategies in cities with complicated structure. It also examines under what cases perimeter control can

have a significant positive effect in mobility improvements and when its efficiency is limited. Finally, chapter 6 presents some conclusions and discusses possible extensions and future work.

Chapter 2

A Macroscopic Fundamental Diagram – Simulation Evidence

This chapter demonstrates using micro-simulation tools that a Macroscopic Fundamental Diagram (MFD) relating production (the product of average flow and network length) and accumulation (the product of average density and network length) exists for one-dimensional arterials and urban networks and that this MFD is the same for vastly different demand profiles. It shows that the trip completion rate (rate at which vehicles leave the network) is proportional to the production. It also shows that the dynamics of the rush hour can be predicted quite accurately without the knowledge of disaggregated link by link measures and Origin-Destination (O-D) tables. Section 2.1 describes the study sites, Sections 2.2 and 2.3 present the simulation results and Section 2.4 provides discussion.

2.1 The study sites

The existence of a Macroscopic Fundamental Diagram was initially tested by simulation in two arterial test sites: (i) Lincoln Avenue in Los Angeles and (ii) Downtown San Francisco Area. The first site is typical of one dimensional multilane

major arterials on urban/suburban areas; the second site is a two dimensional network with varying geometric characteristics.

2.1.1 Major Arterial

This test site is 1.42 mile long stretch of a major urban arterial (Lincoln Avenue) north of the Los Angeles International Airport, between Fiji Way and Venice Boulevard, in the cities of Los Angeles and Santa Monica. The study section includes 7 signalized intersections with link lengths varying from 500 to 1,600 feet. The number of lanes for through traffic per link is three lanes per direction for the length of the study area. Additional lanes for turning movements are provided at intersection approaches. The free flow speed is 40 mph. Traffic signals are all multiphase, operating as coordinated under traffic responsive control, as part of the Los Angeles Automated Traffic Surveillance and Control (ATSAC) system. System cycle lengths range from 100 seconds early in the analysis period (6:00 to 6:30 am) to a maximum of 150 sec during the periods of highest traffic volume (7:30 to 8:30 am).

A field study was undertaken obtain a comprehensive database of operating conditions in the study area. First, basic data on intersection geometrics, spacing and free-flow speeds were obtained from field surveys. Next, manual turning movement counts at each intersection and floating car studies were undertaken for a four hour period (6-10 am) on Wednesday May 26th, 2002. Finally, signal timing data for the study period were obtained from the ATSAC database. Figure 2.1 shows a sketch of the study site.

Figure 2.1: Study site I: Lincoln Avenue, LA

2.1.2 Downtown City Center

This test site is a 2.5 square mile area of Downtown San Francisco (Financial District and South Of Market Area), with more than 100 intersections and link lengths varying from 400 to 1,300 feet. This site consists of more than 800 link-lanes, and the total length of roads is about 76km. The number of lanes for through traffic varies from 2 to 5, and the free flow speed is 30 miles per hour. Traffic signals are all multiphase, fixed-time, operating on a common cycle length of 100 seconds for the west boundary of the area (The Embarcadero) and 60 seconds for the rest. Data on the study network, including geometrics and traffic volumes, were available from previous research studies on traffic control.

2.2 Simulation Results for One-Dimensional Major Arterial

Traffic was simulated in the Lincoln Ave. study site with CORSIM microscopic simulation model (FHWA, 2003) for a period of 4 hours (6am-10am) with time- and space-dependent demand.[2] Data for the simulation were obtained from the field study, described in subsection 2.1.1. During this study period a wide range of traffic conditions was observed: (i) low volume off-peak conditions (6am-7am), (ii) peak period conditions (7am-9am) and (iii) post-peak mid-day flow conditions (9am-10am). Traffic demand was high especially during the peak hour. Traffic volumes were heavily directional with the higher through and turning volumes in the northbound direction. Observed travel speeds

[2] Turning movement percentages and flows generated from origins change every 15 min during the simulation

on the test section were 25 mph during the off-peak times and dropped to about 10 mph during the peak hour in the heavily traveled northbound direction.

During the simulation, the vehicle-meters travelled (VMT), vehicle-hours travelled (VHT) and exit rates (vehicles exiting Lincoln Ave. either from the downstream end or turning in one of the cross streets) every 15 seconds were recorded for every link in the network. These quantities were aggregated every two cycles (~4min) and travel production P, output (trip completion rate) O, average flow $\overline{q}$, speed $\overline{v}$ and density $\overline{k}$ according to Edie's definitions (1963) were calculated for the whole network. The reason for this aggregation with time is to smooth the variations of traffic during a cycle.

Figure 2.2 shows scatter plots of ($\overline{q}$, $\overline{k}$) and ($\overline{v}$, $\overline{k}$) pairs. It is clear that a Macroscopic Fundamental Diagram (MFD) holds for the study site. The three different regimes described in Section 1.3, are clearly noticeable. While ($\overline{q}$, $\overline{k}$) pairs observed in regime I, flow-density pairs for individual links mostly observed in regimes A and B (shown in figure 1.2). For ($\overline{q}$, $\overline{k}$) pairs in regime II most of the links operate at capacity, while in regime III individual links are observed in regimes B and C of figure 1.2. Congestion would be unevenly distributed over the network if individual links in regimes A and C persisted simultaneously. This could create points beneath the curve of figure 2.2a.

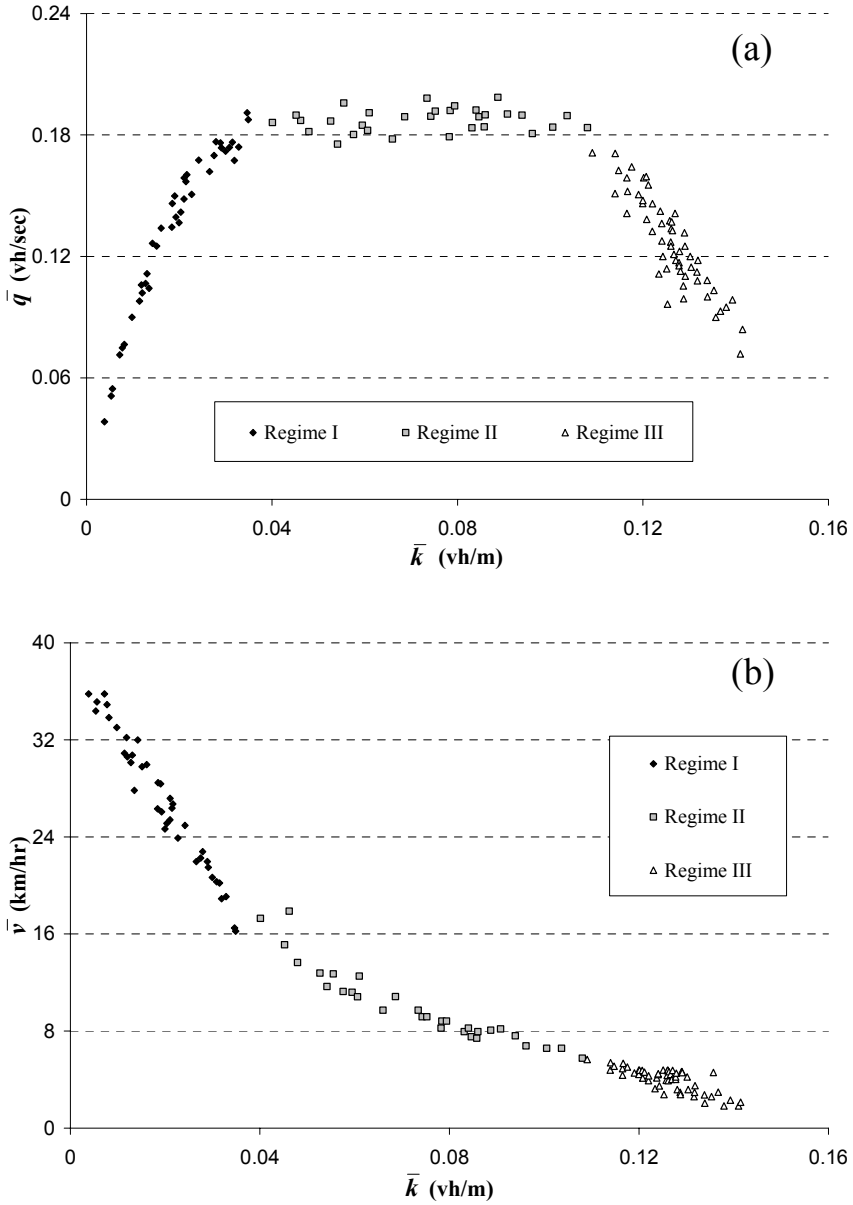


Figure 2.2: MFD for Lincoln Ave. (a) $\bar{q}$ vs. $\bar{k}$; (b) $\bar{v}$ vs. $\bar{k}$

Figure 2.3 shows total output and production pairs for the major arterial network. Notice that they are linearly related. The inverse of the slope of this line describes the average distance travelled in the network per trip completion and can be considered as the average trip length (approximately 1760m). The high degree of correlation (R^2=0.98)

suggests that the traffic production is a good proxy for the aggregate output, which cannot be easily observed.

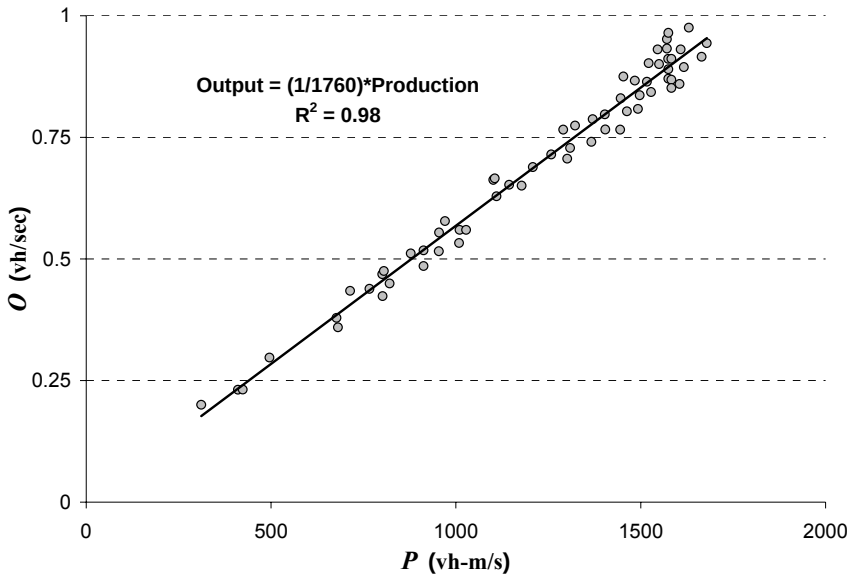


Figure 2.3: Output vs. Production pairs per 2 cycles for Lincoln Ave.

2.3 Simulation Results San Francisco Business District (SFBD)

2.3.1 The demand

Traffic was simulated in the second study site for many periods (~10) of 4 hours with time- and space-dependent demand. We scaled up the estimated real demands by a factor to make sure that the city was severely congested. The simulation package did not allow importing time-dependent origin-destination (O-D) tables. This was possible by changing the turning movements at intersections and the generated flows at the "edge nodes" of the network with time. To test the sensitivity of the proposed macroscopic relations to different O-D tables, many runs were made while varying turning movement percentages between runs, and within each run the generated flows were changing every 15 minutes.

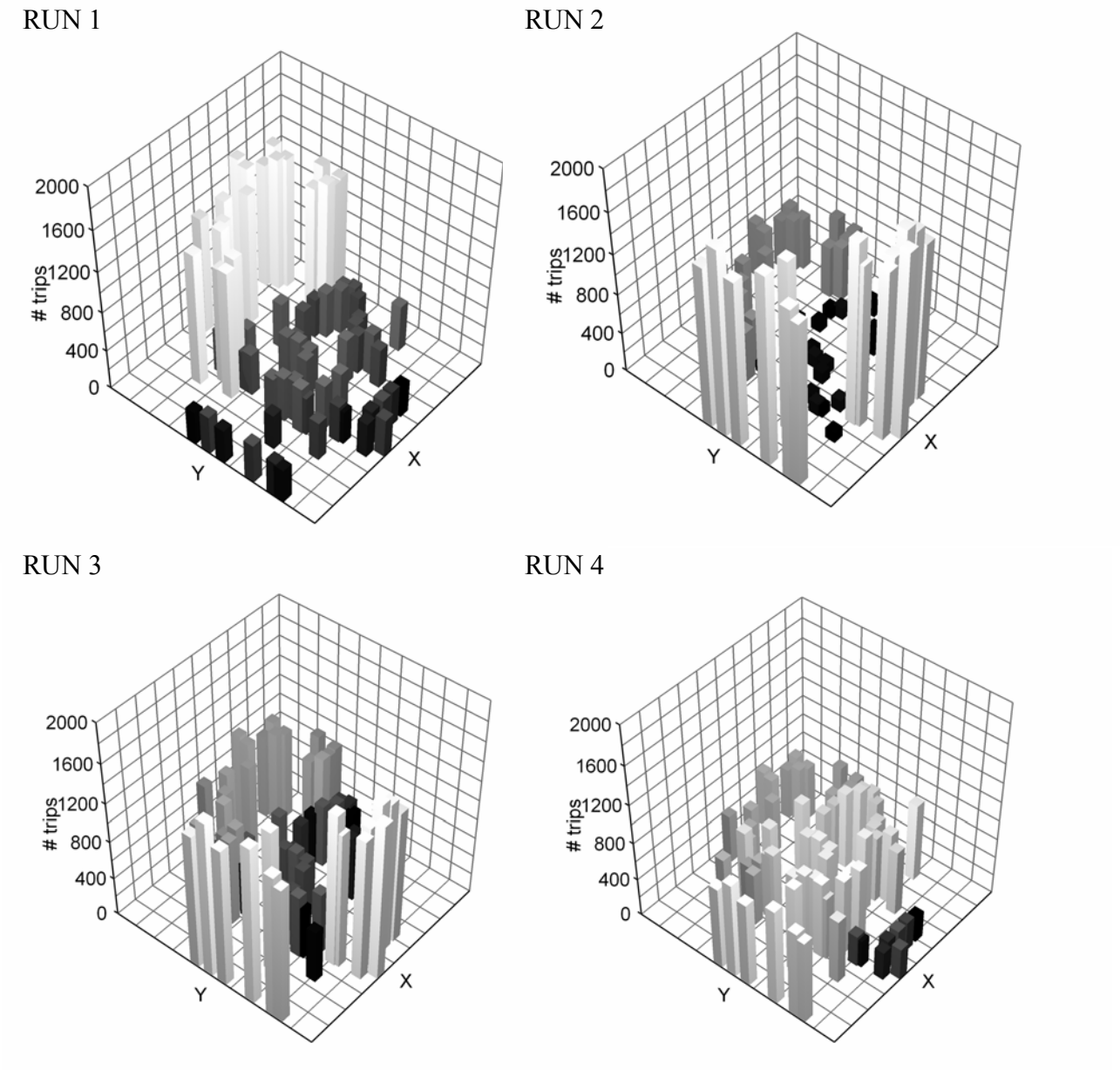
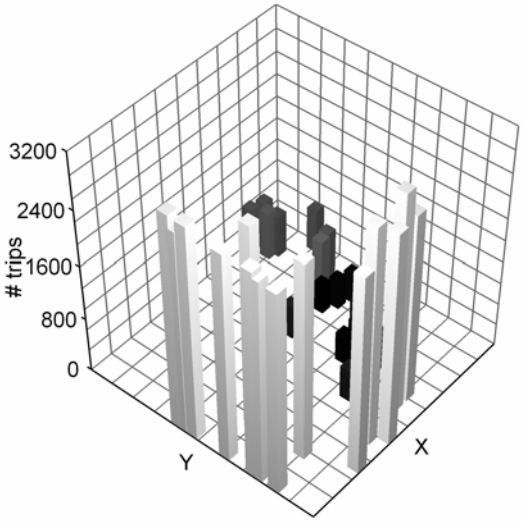


Figure 2.4: Total Trips Originated per node per run

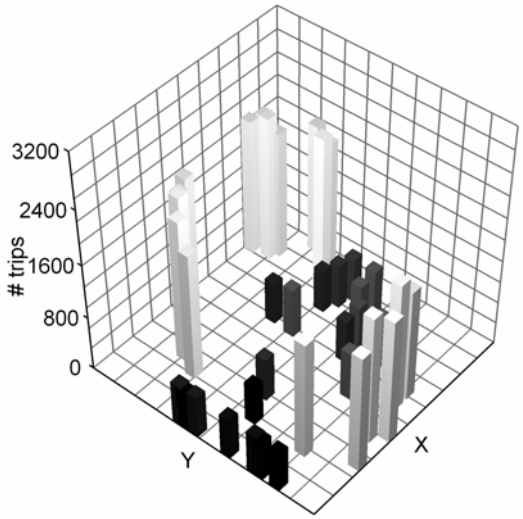
Figures 2.4 and 2.5 shows the total number of trips generated and completed per node in a period of 4 hours for four different runs. Runs differed vastly in the geographical distribution of demand. For example, in some of the runs (run 4) more than 70% of the demand origins were internal to the SFBD whereas in others (runs 1 and 2), 80% of the traffic was external, entered the SFBD from either the north or the south. Trip endings

followed similar profiles (mostly external in runs 1 and 2, evenly distributed in run 3, mostly internal in run 4).

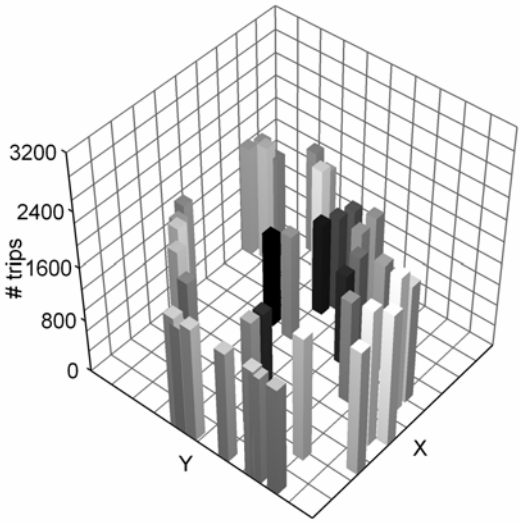
RUN 1



RUN 2



RUN 3



RUN 4

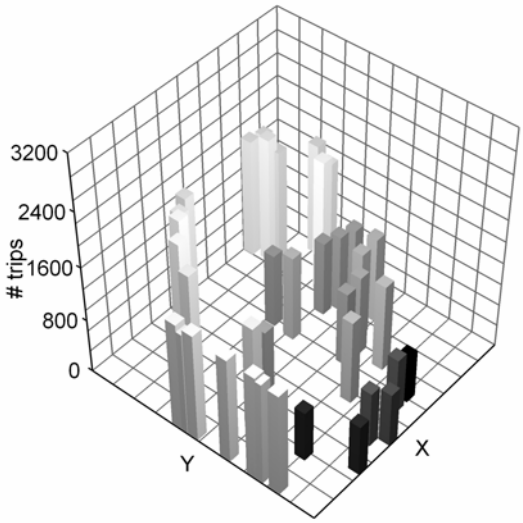


Figure 2.5: Total Trips Ended per node per run

2.3.2 A Macroscopic Fundamental Diagram for SFBD

It is now shown that the SFBD network has an MFD with less scatter than for individual links and that the MFD is reproduced under different demand conditions.

First, figure 2.6 shows scatter plots of flow-density pairs for two single representative links of the network with a 2 min sampling interval for a period of 4 hours (one run). Although one can discern the trapezoidal edge (representing steady state conditions) of an one-link Fundamental Diagram, it is clear that transitions between points in the steady state regimes occur by following different paths, creating a high disorder and scatter especially when the flows are maximal ($k \cong 0.05$vh/m). But, when all the links are aggregated this disorder disappears.

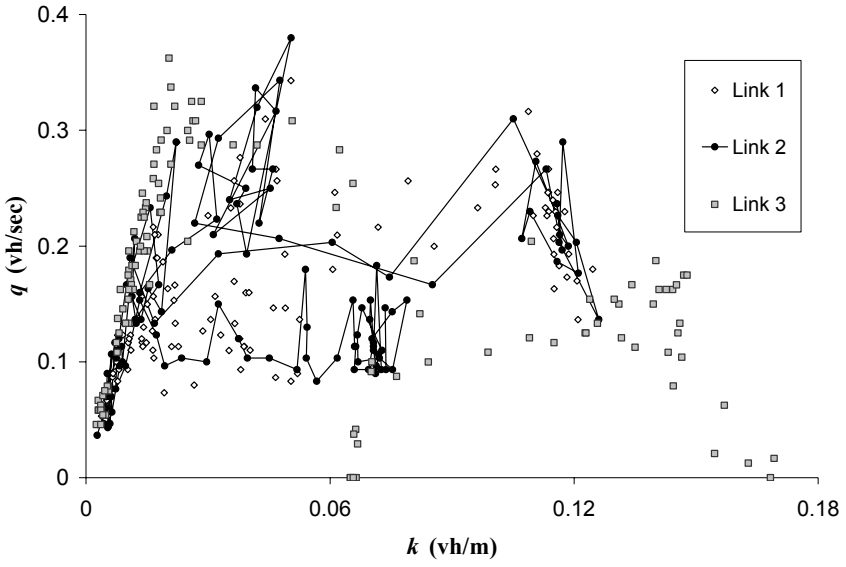


Figure 2.6: Scatter plot of q vs. k for three individual links (without aggregation)

Figure 2.7 shows travel production-accumulation pairs and speed-accumulation pairs aggregated per 2 cycles for the whole network for four different runs. Note that the points line along a well defined Macroscopic Fundamental Diagram (the orderliness is clear).

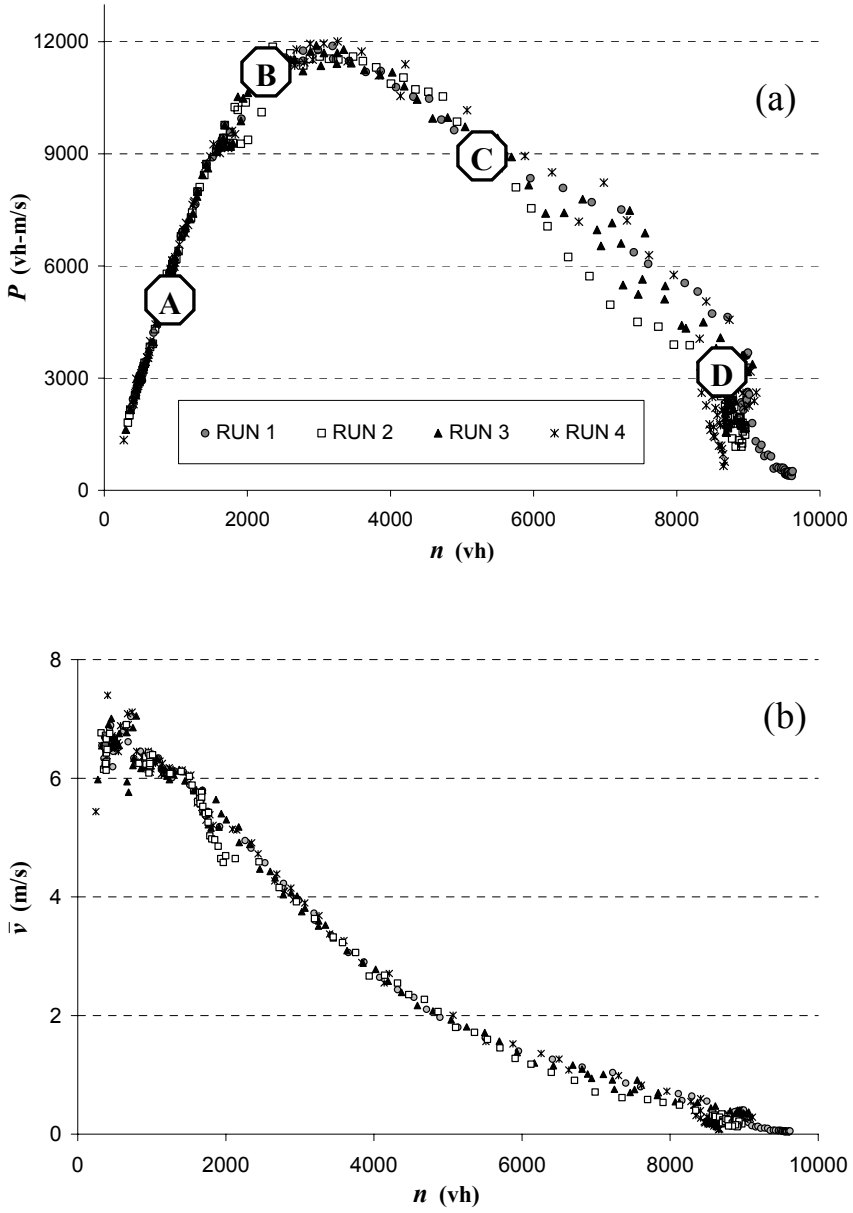


Figure 2.7: MFD for different runs in SFBD
(a) Travel Production vs. Accumulation; (b) Speed vs. Accumulation

Particularly, since Figure 2.7 includes points for runs with vastly different O-D tables it suggests that this diagram is property of the network and is independent of the demand and the disaggregated link data. Also, the maximum value of travel production, reached for an accumulation of about 3200 vehicles, is the same for all the runs. Note as well that production values are near the maximum (difference <5%) for a wide range of accumulation (2500-4000 vehicles). Four different states of the system (A, B, C and D) have been highlighted in Figure 2.7a.

Figure 2.8 shows snapshots of the network corresponding to these four states. State A belongs to regime I when the system is undersaturated and the average speed is about 25km/hr. As demand increases system moves to state B where the vehicle – travel production is near the maximum and the average speed is 17km/hr. In state C congestion is broad, long queues are observed and the average speed drops to 7km/hr. In state D output is near to zero and the majority of vehicles are stopped. It was found that when traffic conditions are very near gridlock (point D), the system does not easily return to better traffic states (B or C) even if the demand decreases significantly. This means that it is necessary to prevent traffic in cities to move to states of very high accumulation and apply control in a preventive form.

We also tested whether the ratio of the output O to production P is insensitive to demand by plotting (P, O) pairs, sampled in 2 min periods, for the runs of figure 2.7. Figure 2.9 shows the result. The resulting curve suggests that the ratio is indeed invariant. Statistical tests cannot reject the hypothesis that the P/O ratio is 1743m of travel, per trip completed throughout the day.[3]

[3] We fitted a 5th degree polynomial to the points of Figure 2.9 and only the linear coefficient is significant.

Figure 2.8: Views of the San Francisco network for different traffic regimes
(white dots are vehicles; black sections are vacant portions of the pavement)

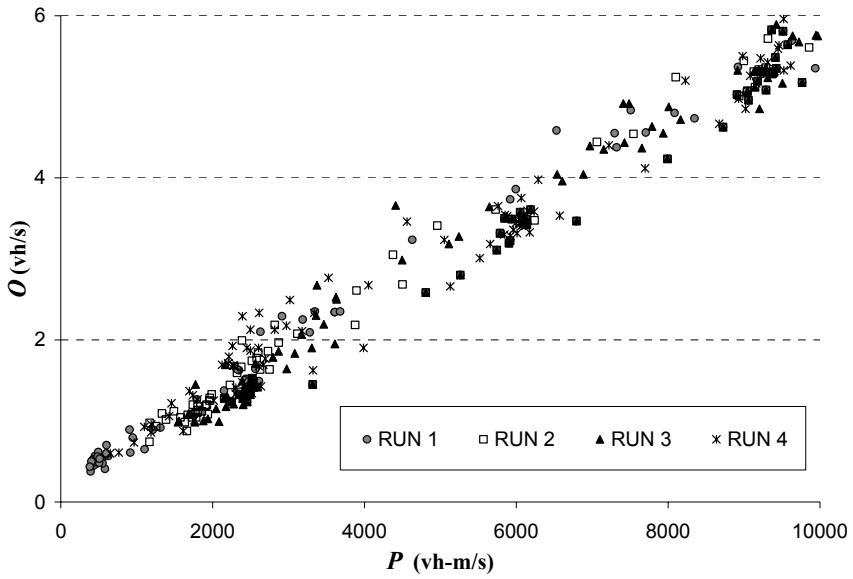


Figure 2.9: Output vs. Travel production for different runs in San Francisco network

2.3.3 Description of Dynamics

The existence of a Macroscopic Fundamental Diagram connecting accumulation and output in SFBD implies that if the boundary conditions change slowly with time we can model the dynamics of accumulation as if this network was a single reservoir, described by equation 1.4. To test this hypothesis, queueing diagrams with cumulative input-output curves where constructed from the simulated records for 4 runs with vastly different demand. The diagrams were also constructed with the single reservoir method, using the following discrete version of equation 1.4: [4]

$$n_{t+1} = n_t + f_t - O\left(n_t\right), \tag{2.1}$$

where n_t, f_t are accumulation and input flow at time t, and $O(n)$ is the MFD function that gives the output for a given accumulation. The function $O(n)$ was estimated by least

[4] Given n_0, $G(n)$ and q_t^{in} for every t, we apply finite differences and solve Equation 2.1 for all $t>0$.

squares method using data points from all the runs.[5] The result is shown in figure 2.10a. Figure 2.10b shows the residuals, i.e. the differences between the observed and predicted pairs.

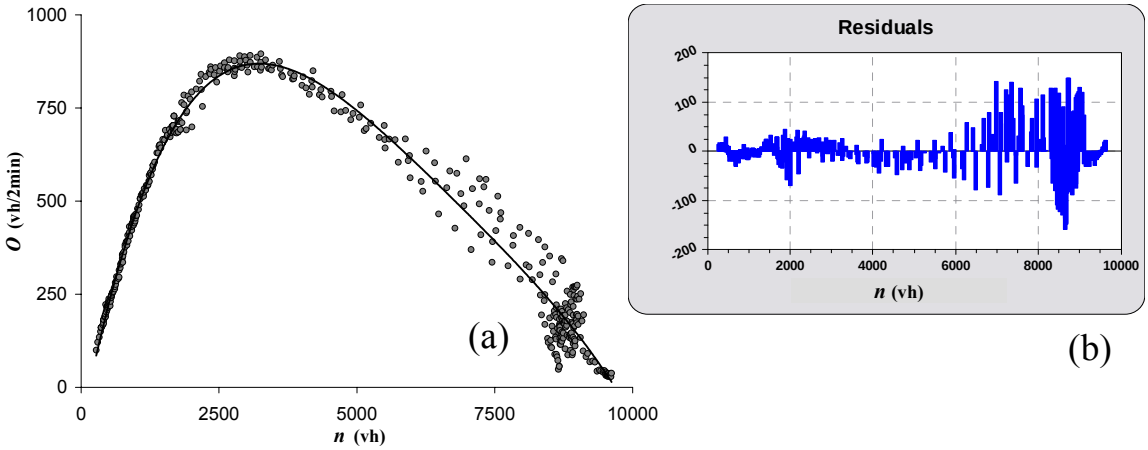


Figure 2.10: a. Output vs. Accumulation for all runs
b. Residuals

We tested the dynamic model predictions under different traffic conditions. Figure 2.11 shows the resulting queueing diagrams, including the simulated and predicted output values and the accumulation during each run. In all cases, the two output curves fit quite closely and they are almost perfectly superimposed when accumulation values do not belong to the decreasing branch of the diagram (Regime III). Predictions of regime III are more difficult, as expected, because congested systems are chaotic (Daganzo, 1998). This is only a minor problem because any control strategy should try and avoid states in this regime.

[5] A 4th degree polynomial was applied. The best fit curve was (R^2=0.97):
$O(n)= -1.0718\times10^{2}+7.5225\times10^{-1}\, n + 1.9213\times10^{-4}\, n^2 + 1.8558\times10^{-8}\, n^3 + 6.8381\times10^{-13} n^4$

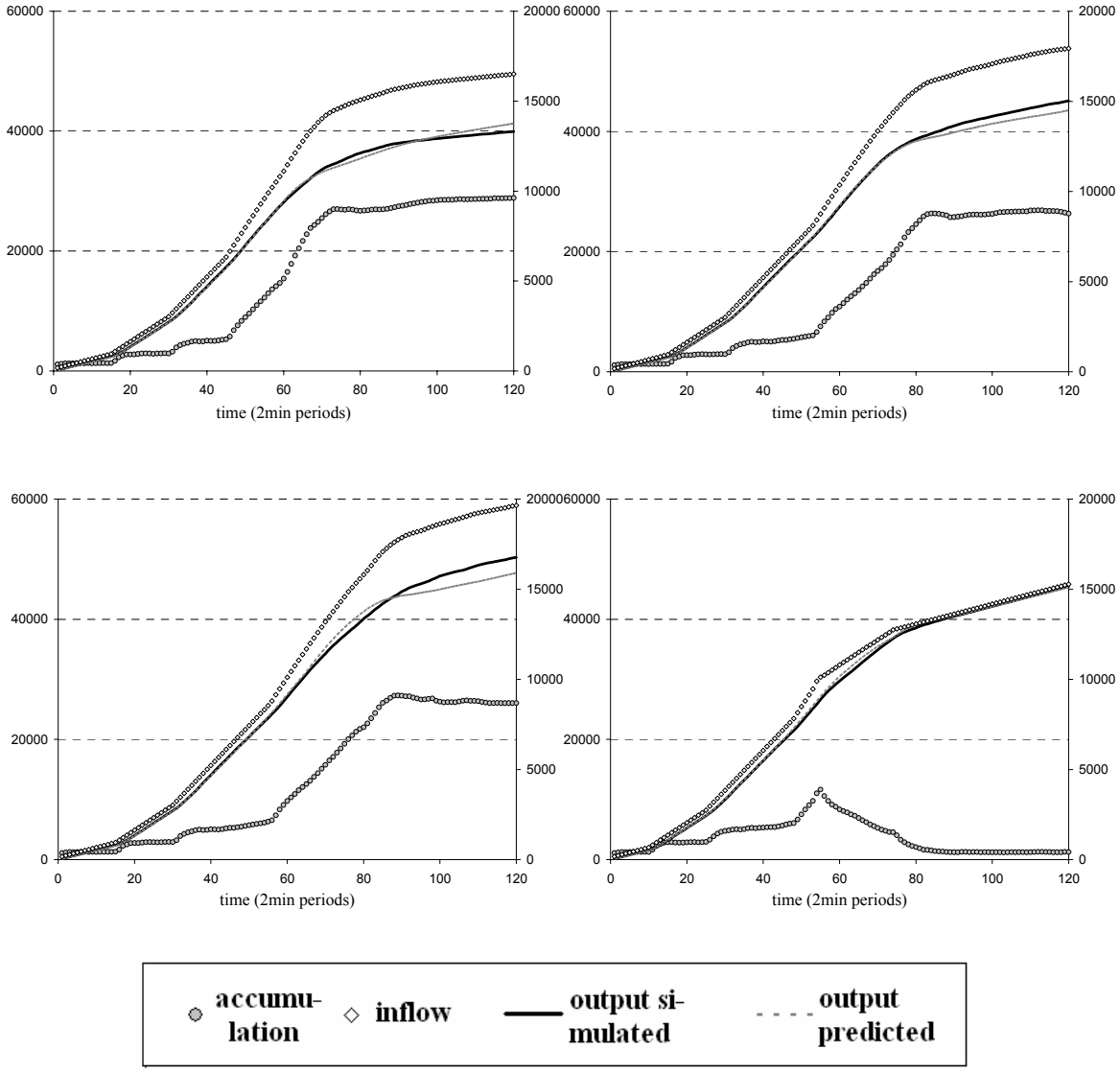


Figure 2.11: Queueing diagrams for different runs in San Francisco network
(Left axes: Cumulative inflow or output (vehicles), right axes: Accumulation (vehicles))

2.4 Discussion

The outcomes of this chapter are encouraging. Our simulation experiment suggests that a macroscopic fundamental diagram exists for a network with a hundred signalized intersections. It also suggests that, conditional on accumulation, large networks behave predictably and independently of their Origin-Destination tables.

If these results hold up to further tests, practitioners will have reliable tools to anticipate the consequences of "simple and smart" traffic management policies and plan without the uncertainty of today's forecast-based approaches. These further tests should include field experiments since simulations must assume particular forms of driving and navigating behavior, which may or may not be realistic.

On the other hand, these findings are of limited use because accumulation and output are not easily observable (i.e. measurable) in real life. Thus, it is fair to ask (i) whether an MFD can be observed in real networks (even if it exists) and (ii) whether we can develop the necessary tools for estimating the relevant quantities to an MFD. Chapter 3 answers both of these questions by presents a natural experiment from Yokohama, Japan.

Chapter 3

Existence of urban-scale macroscopic fundamental diagrams: Experimental findings

The outcomes of Chapter 2, even though encouraging, are based on simulation. Chapter 3 describes a field experiment in Yokohama (Japan), and reveals that a macroscopic fundamental diagram (MFD) linking space-mean flow, density and speed exists on a large urban area. The experiment uses a combination of fixed detectors and floating vehicle probes as sensors. The space-mean speeds and densities at different times-of-day are estimated for the whole study area using relevant parts of the detector and taxi data sets. The analysis also explores the relation between the space-mean flows on the whole network, and the trip completion rates, which dynamically measure accessibility. Section 3.1 describes the study site and the data; Section 3.2 presents some results using only the detector data; Section 3.3 extends these results for the whole study site (including streets without detectors) by combining all the data and Section 3.4 displays final remarks.

3.1 Site and data description

Yokohama is a major commercial hub of the Greater Tokyo Area. It developed rapidly as Japan's prominent port city with a population of 3.6 million. Its road network includes streets of various types, with closely spaced signalized intersections at its center (~100-300m), and a few elevated freeways. Streets have 2 to 4 lanes in each direction. The speed limit is 50 km/hr on arterials. Major intersections are centrally controlled by multiphase traffic signals, with a cycle time that responds to traffic conditions: 110-120 sec long at night and 130-140 sec during the day.

The part of downtown Yokohama examined in this paper is approximately a 10 km^2 triangle with corners at Yokohama Station, Motomachi-Chukagai Station and the Shin-Hodogaya Interchange. Yokohama's center is congested during peak hours with speeds considerably less than 50% of the average speed during the off-peak; average speeds under 10 kilometers per hour are observed for extended periods during morning and evening peak in weekdays in the arterial network. Figure 3.1 shows a sketch of Downtown Yokohama (the perimeter of the study site is shown as a dashed line). Data from the following two sources, and spanning one month (December 2001), were available:

- *Fixed sensors*: 500 ultrasonic and loop detectors positioned midblock on arterial lanes of most major intersections in the study area. The detectors are located about 120m upstream of the stopline in case of long links and about 50m in case of short links. They provided 5-min vehicle counts and occupancy measurements.

- *Mobile sensors*: 140 taxis, called Internet Protocol probe car System (IPCar), equipped with GPS and a data logger reported their position and other data (all the data were time stamped). These data included activations and deactivations of the parking brake, the left or right blinker and the hazard lights, as well as the beginning and end of all stops lasting more than a few seconds. The format of the data is shown in Figure 3.2.

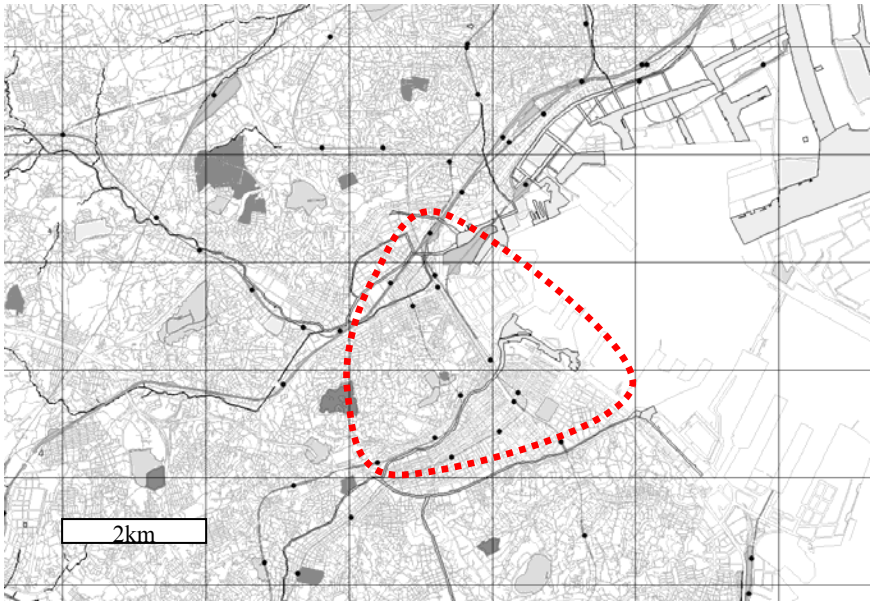


Figure 3.1: A map of Downtown Yokohama

For each taxi ID and date the IPCar System does not store the GPS position at regular intervals. Instead it logs the *time* (Column 1 in Figure 3.2 - hh:mm.ss) and position (Column 6 - *Latitude* and Column 7 - *Longitude*) every time that the *state* S of the vehicle changes, away from an old state (Column 2). The possible states are: short trip (S=1), short stop (S=2), right blinker (S=4), left blinker (S=8), hazard light (S=16), parking brake (S=32). A change from short trip to short stop/short stop to short trip occurs when vehicle speed drops below/increases above 3km/hr. There is a maximum of 30 sec

duration for a short trip, but there is no time limit for short stop.[6] Column 3 and Column 4 display the time (*duration*) and *distance* that the corresponding state had persisted. Note that this duration matches the change in time from Column 1, Δ(time), for S=1 and S=2, but not in case S>2. This happens because when a taxi uses some special equipment (S>2), it should be also reported if it is moving (S=1) or stopping (S=2), i.e. the taxi is at *two* states at the same time. For example (see Figure 3.2) from 9:18.50 to 9:19.39 vehicle was moving with speed continuously <3km/hr (short stop, S=2) and traveled 23m in 49sec, and from 9:19.39 to 9:20.09 was moving (short trip, S=1) for 30sec and traveled 210m. In the mean time it had on the right blinker (S=4) for 114sec and traveled 99m ending S=4 at 9:19.43. Thus, the total distance traveled for one hour is estimated by summing all the numbers from Column 4 for S=1 and S=2, ignoring S>2. The system also records if the *GPS* works properly or not (Column 5).

Taxi ID	1046					
Date	12/14/2001					
1	2	3	4	5	6	7
Time	State	Duration	Distance	GPS	Latitude	Longitude
091645	1	28	296	1	32676902	128678449
091650	2	5	26	1	32676997	128678395
091720	1	30	335	1	32678413	128675699
091750	1	30	316	1	32679908	128673160
091759	1	9	9	1	32680059	128673096
091841	2	42	16	1	32680213	128673208
091850	1	9	6	1	32680062	128672440
091939	2	49	23	1	32680282	128672640
091943	4	114	99	1	32680464	128672827
092009	1	30	210	1	32681685	128673859
092024	2	15	30	1	32681744	128674079
092041	8	6	23	1	32682259	128675520
092054	1	30	202	1	32682371	128675369
092119	4	6	11	1	32681982	128673969

Figure 3.2: Format of mobile sensors data

[6] If a taxi is moving (with speed >3km/hr) for 2 min, four consecutive short trips will be recorded.

The taxi data were not linked to a digital map of Yokohama, so it was not possible to tell without human intervention and great effort whether or not a taxi had passed over a particular detector at a particular time. More details about the IPCar System can be found in Sarvi et al. (2003).

3.2 Results from detector data

This section shows that the part of Yokohama's network covered by detectors has a Macroscopic Fundamental Diagram (MFD) with less scatter than for individual links, and that the MFD is reproducible under different demand conditions.

3.2.1 Existence of the MFD

The initial data consist of time-series of flow and occupancy at about 500 locations upstream of signalized intersections in the city center. Denote by i a road lane segment between intersections, and by l_i its length. Let q_i and o_i, be the flow and occupancy measured by the corresponding detector in a particular time slice. We use A for the set of lane segments in the study area, and $A' \subset A$ for the subset with detectors. As is well known, the density at a detector location is $k_i = o_i/s$, where s is the space-mean effective vehicle length, which is about $s \cong 5.5$ m (Kuwahara, 2007). We are interested in patterns of these variables produced by both individual detectors and all the detectors combined.

To do this, data have been aggregated according to Edie's definition (1963) as follows:

- Weighted average flow: $q^w = \sum_i q_i l_i / \sum_i l_i$ for all lane segments i with detectors ,

- Unweighted average flow: $q^u = \sum_i q_i / \sum_i 1$,
- Weighted average occupancy o^w and density k^w: $o^w = k^w s = \sum_i o_i l_i / \sum_i l_i$ and
- Unweighted average occupancy o^u and density k^u: $o^u = k^u s = \sum_i o_i / \sum_i 1$.

Note that the numerator of q^w is the production P, defined in Chapter 1. The unweighted averages are representative of the part of the network covered by detectors; i.e., space-means for this part, which we denote *A"*. Note that *A"* is disjoint and that $A'' \subset A'$. The weighted averages would be space-means for *A'* if the detectors are at representative locations within each link. This can only be guaranteed for flows (for all the locations along a link), on time slices large compared with a traffic cycle. This is because on this time scale link, flows are roughly the same. But obviously, the same is not true for density or speed, as it matters where these are measured within a link.

Consider a single link for the moment. Figure 3.3 is a scatter-plot of q_i vs. o_i, for a whole weekday with time slices Δt = 5 min, for two different detectors. It is not clear that the points belong to a well defined reproducible curve, especially under congested conditions. The disorder persists at the link level, after aggregating data for the lanes of a single link. For the same value of occupancy, variations in flows are high (Root Mean Square >0.2, for values of occupancy that capacity occurs). Non steady-state conditions, variations in turning movements or different types of platoon arrivals at the intersections are usual phenomena that can cause this kind of disorder in the traffic data.

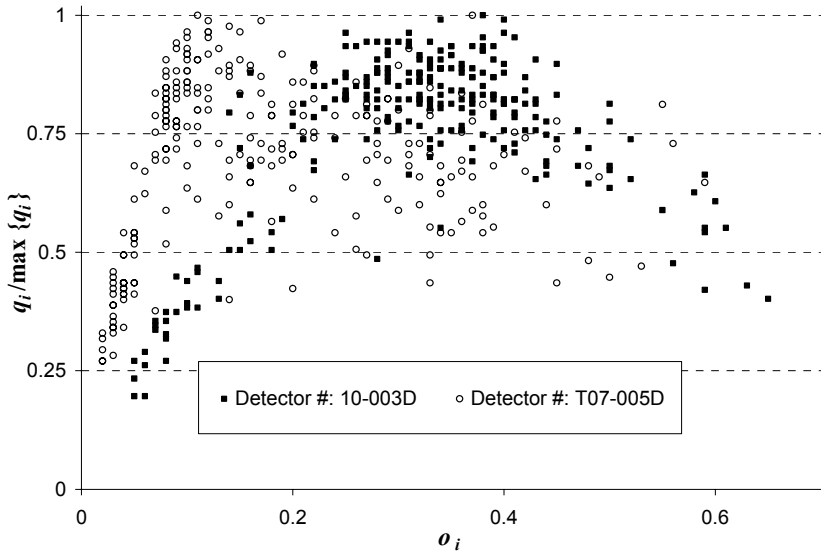


Figure 3.3: Flow vs. occupancy pairs for two single detectors across a day

To see what happened to this pattern when data were aggregated, we considered two different days: a weekday (12/14/2001) and a weekend day (12/16/2001). Figures 3.4a and 3.4b show the time-series of unweighted average flows and occupancies, q^u and o^u, that were observed. The one day period is divided in 4 time periods A, B, C, D based on the occupancy series for the weekday representing different demand profiles (AM and PM peak growth and decay). Based on this categorization, we code the data according to 8 different time periods (4 per day) and we present the results of aggregation in figures 3.5 and 3.6. Note that there is severe congestion in the weekday's afternoon rush hour, but not on the weekend; e.g., at the time of maximum occupancy (around 17:00 hrs) the average flow is below the maximum on the weekday -- but this does not happen on the weekend. Note as well that flow and occupancy varied considerably by time-of-day, on both days. These time periods represent not only various values of total demand but also strong variability across origins and destinations as people make different types of trips during AM or PM (entering vs. exiting the city center), during weekdays or weekends

(work vs. leisure trips).These substantial variations within and across days suggest that the demand rates and origin destination (O-D) tables varied considerably.

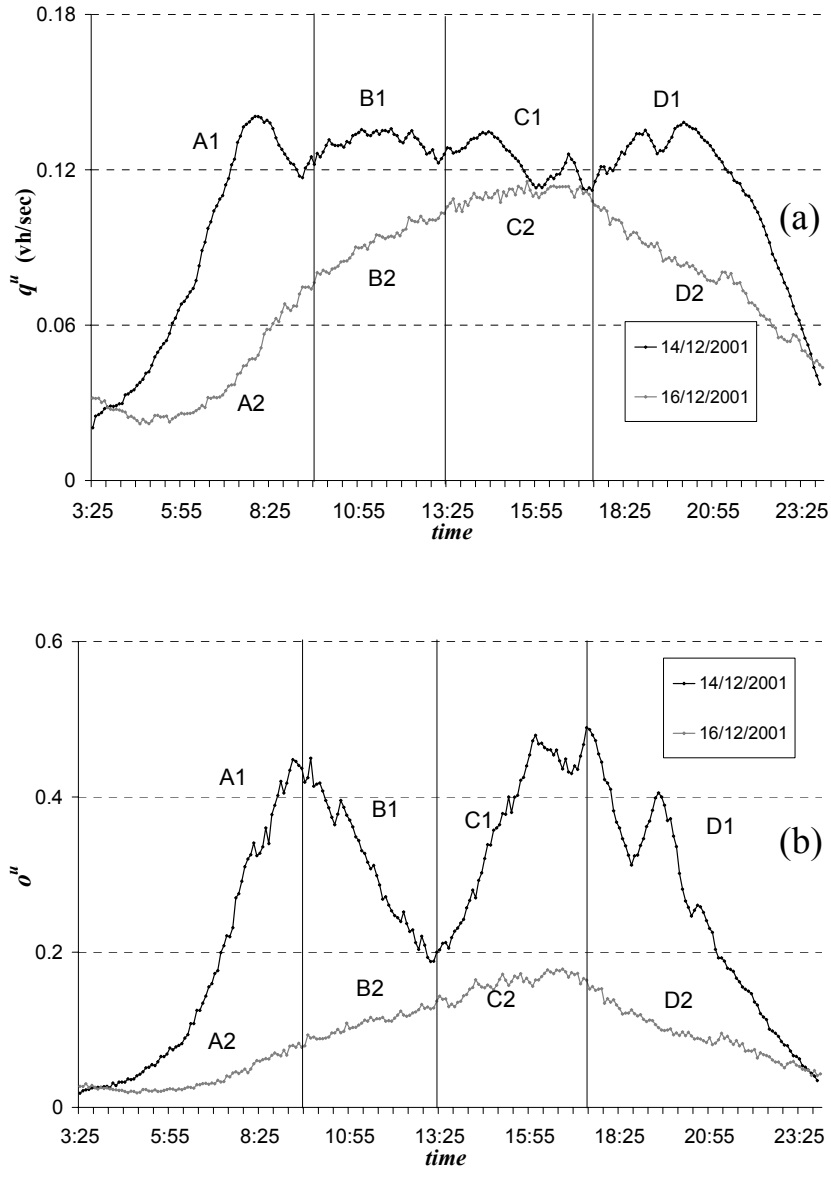


Figure 3.4: Loop detector data, Time-series: (a) average flow; (b) average occupancy

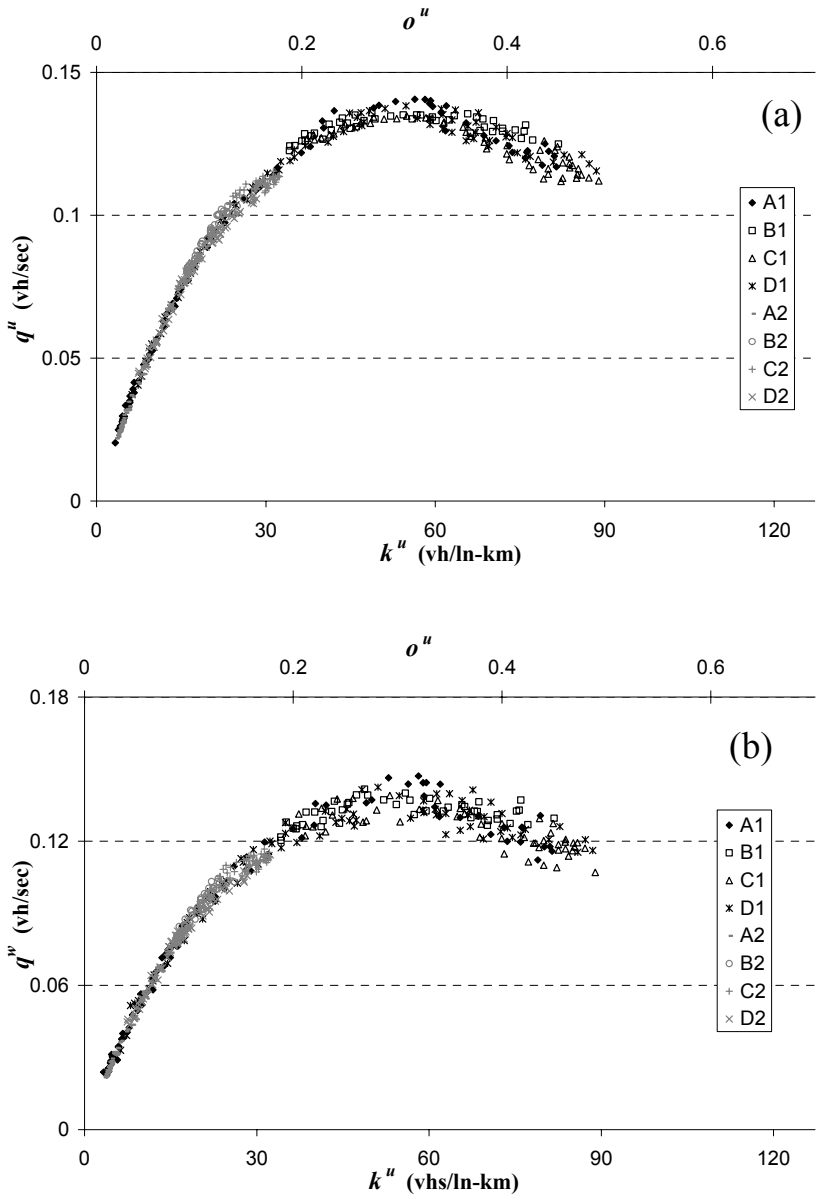


Figure 3.5: Scatter plot of average flow vs. average occupancy:
(a) unweighted average flow; (b) weighted average flow

Figures 3.5a, 3.6a and 3.6b are scatter plots of pairs of: q^u, $k^u = o^u/s$ and $v^u = q^u / k^u$. These are the averages representative of the detector locations *A''*. The high degree of ordering (compare with Figure 3.3) strongly suggests that an MFD exists on *A''*, i.e. the disjoint portion of the network covered by detectors. Note that each set of symbols cannot

be distinguished from the others despite the substantial variations in O-D demands across time periods. Plots involving weighted averages exhibit similar behavior. These plots are of great interest because they extend these results to *A'*, i.e. the part of the network including all the streets with detectors. Figure 3.5b shows how well q^w , the production per unit length on all of *A'*, can be predicted from the detector data.

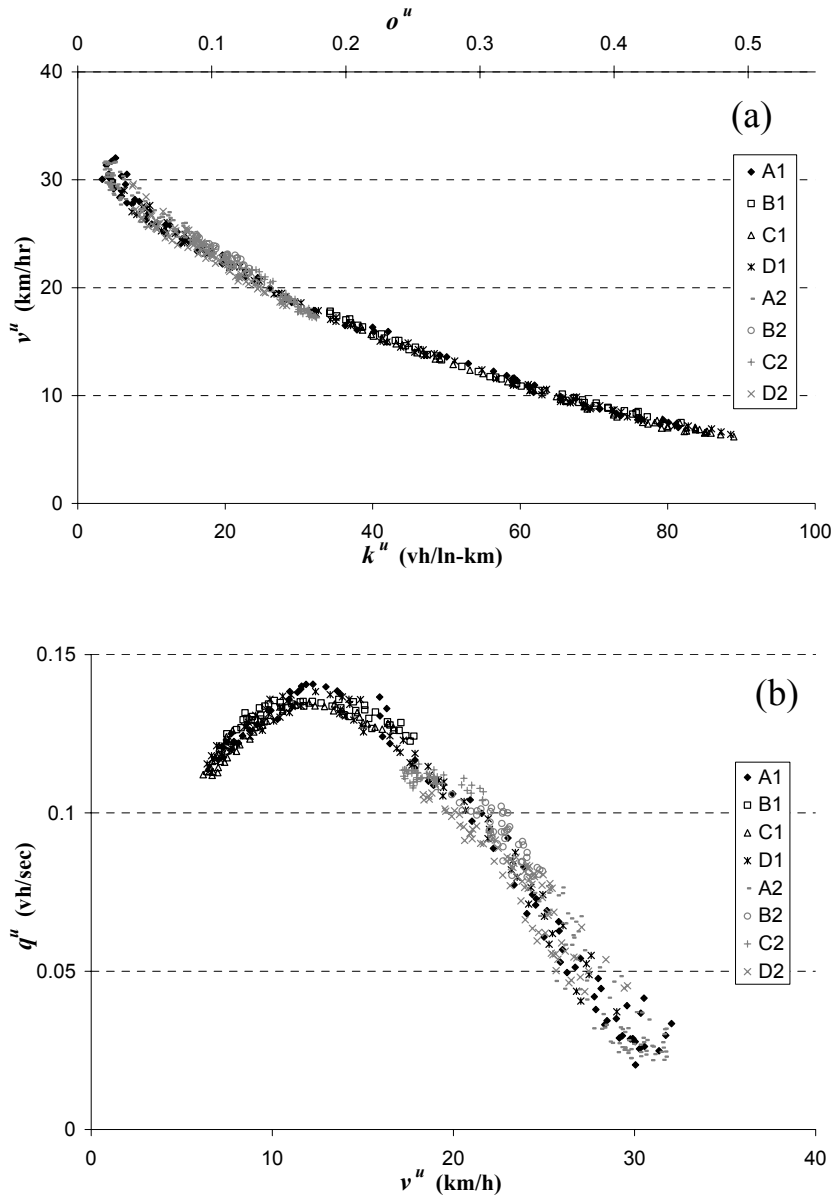


Figure 3.6: Loop detector data across two different days: (a) average speed vs. average occupancy; (b) average flow vs. average speed.

Note as well (see Figure 3.6a) (i) that the flow-speed relation is not monotonic as mentioned in Chapter 1; (ii) that the maximum flow is reached for an average occupancy of 0.3 (Figure 3.5) and an average speed of 13 km/hr (Figure 3.6b) and (iii) that these values are consistent across time periods (morning and evening peaks). This confirms that "aggregation is good" and O-D tables do not matter in predicting vehicle-kilometers traveled in an area from occupancy data.

There is also no hysteresis, a common phenomenon clearly observable from experimentally obtained flow-density plots (e.g. Treiterer and Myers, 1974). Average flow and occupancy pairs during AM and PM do not separate and instability in drivers behavior during acceleration and deceleration does not create any difference between the growth and decay of demand.[7]

The consistency of these results indicates that if the trip completion rate in *A'* is linearly related to q^w then dynamics models based on the MFD can both provide an accurate representation of a city's traffic and suggest realistic ways of improving it. Based on these diagrams we can develop simple perimeter control strategies for which production *P* or exit flow is maximized. This can be achieved by preventing average occupancy measured from the detectors to exceed sweet-spot occupancy of figure 3.5 (approximately 30%) or average speed to decrease below "sweet-spot" speed of figure 3.6b (about 13km/hr). Given Yokohama's heavy congestion, its residents would benefit

[7] Daganzo (2002) attributed the hysteresis phenomenon to lane changing and the non-conservative nature of flow in a single lane. The fact that hysteresis disappears when we aggregate lanes supports this interpretation.

substantially from this kind of applications.[8] Therefore, we now examine the relationship between the trip completion rate and q^w.

3.2.2 Existence of a linear relation between exit flows and network flow

To describe the dynamics of the rush hour for *A'*, as this expressed by equation 2.1, it is necessary to estimate a relation between trip completion rate *O'*, i.e. output for *A'*, and average density k^w. This would be easy if we were able to know a relation between the average network flow q^w and rate *O'*. Rate *O'* is the sum of (i) the rate at which cars exiting the perimeter from streets with detectors and (ii) cars that leave internally *A'*. But, cars are most readily observed along the perimeter where there are detectors. In that sense, we focus for now on this observed outbound perimeter flow (veh/hr), *D'*. This perimeter flow includes a considerable part of the trip-ends. So, a connection between *D'* and q^w aims to shed some light on relation between q^w and rate *O'*, although it excludes internal trips.

Figure 3.7a shows the time series of *D'* and q^w for the weekday of 12/14/2001 where the flows are sampled in 5 min intervals. They appear to be correlated. Other weekdays and weekends are similar. Figure 3.7b confirms this fact; it reveals that the ratio q^w/D' is close to 0.033 in every 5-min time slice of the day. Note there is no trend. The best-fit 5-degree polynomial deviates from the straight line by less than 0.2% (Root Mean Square) and all the coefficients, except the intercept, are statistically insignificant (see Figure 3.8 for *t*-statistic and *P*-values). The residuals can be explained by statistical variations in trip lengths across individual cars.

[8] Reference (Daganzo, 2007) showed that these kinds of strategies are Pareto optimal for single neighborhood cities, which is a valid assumption if the whole city is uniformly congested.

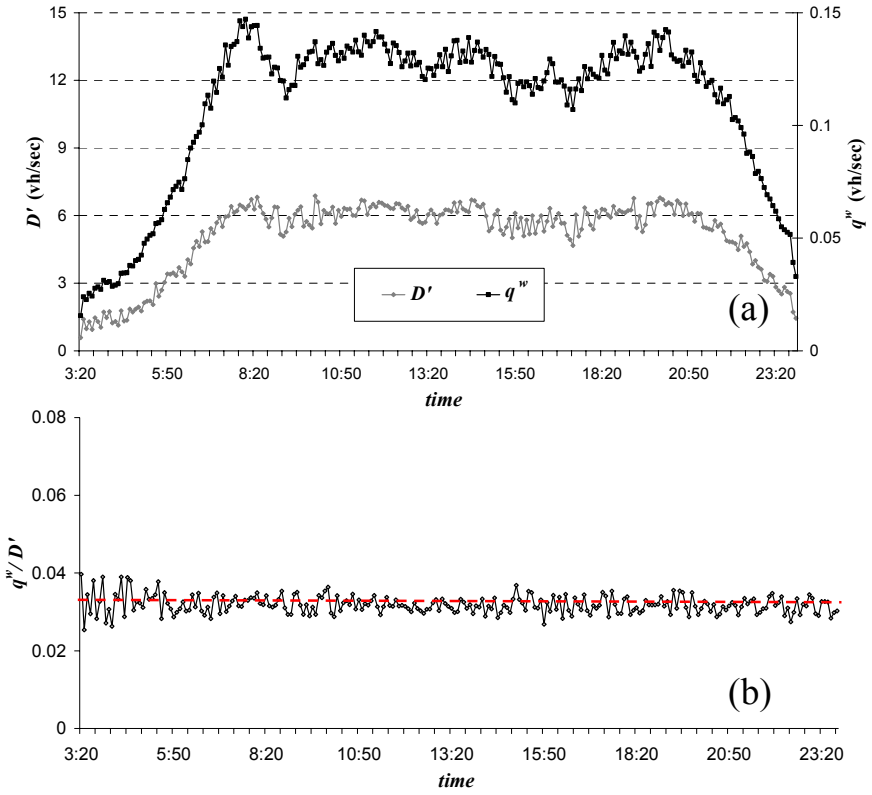


Figure 3.7: Average network flow and exit rates measured by detectors on 12/14/2001 (5-min time slices): (a) Time series of the two variables; (b) time series of their ratio.

	Coefficients	*t* Stat	*P*-value
Intercept	0.0328	36.01	$<10^{-20}$
t	-7.31×10^{-6}	-0.10	0.92
t^2	-2.74×10^{-7}	-0.15	0.88
t^3	3.77×10^{-9}	0.20	0.84
t^4	-1.72×10^{-11}	-0.21	0.83
t^5	2.63×10^{-14}	0.20	0.84

Figure 3.8: Statistical test for a 5th degree polynomial of q^w/D' with time t

We conclude that an MFD exists for the set of streets where detectors are implemented. These detectors are typically placed near traffic signals, where observed speeds are lower. Thus, the results may not be representative of the whole network. It is

fair to ask whether these relations can extend to the whole neighborhood of a city and to trips that end internally. Section 3.3 below utilizes taxi data to expand the results from *A'* to *A*. The section also demonstrates that a trip completion relation similar to Figure 3.7b also holds for *A*, including trips that end internally and leave *A* from the perimeter.

3.3 Results from taxi data

Fixed detectors' data do not provide information about the average speed, number of vehicles moving or average flow on all the streets of the study area, *A*. Therefore, we cannot guarantee that the results of the previous section apply to *A*. Taxis are useful markers because they can be identified; if e.g. 10% of the vehicles are taxis we would conclude that 10% of vehicles are observed in exiting from the boundary of a region. Therefore we can have an unbiased estimator of the number of vehicles in a city.

Nevertheless, taxis do not behave identically as ordinary vehicles because they stop more frequently to pick up passengers, or follow circuitous routes when searching for passengers (e.g. trip between points A_4 and A_5 shown in figure 3.9). We use passenger-carrying taxis as markers that behave identically as ordinary vehicles as far as (i) the distribution of speeds and (ii) average trip completion rate per unit length traveled, are the same.

We can estimate the flow of all cars in *A* by scaling up the observed flow of all cars in *A'* with a factor determined from the observed flows of passenger-carrying (full) taxis in *A* and *A'*. With this scaling method we can also estimate the trip completion rate, the vehicular accumulation and the space-mean speed in *A*. To do this effectively, we need a set of full taxi trips. Subsection 3.3.1 presents a filtering method to exclude time periods when taxis do not carry passengers; subsection 3.3.2 describes how to estimate

macroscopic traffic characteristics (accumulation, and speed) for the whole area *A*; and subsections 3.3.3 through 3.3.5 show the results.

3.3.1 Filtering method for passenger-carrying taxis

We assume that passenger-carrying taxis follow routes through Yokohama's center similar to those of cars. Although the events in the data set did not include "boarding" or "alighting" moves, the data set included other information (as described in section 3.1) that revealed whether a particular stop was a passenger move in or out of the taxi. From this we identified the passenger-carrying (full) trips.

To tag a stop as a passenger move, *at least one* of the following hypotheses (H) should be satisfied:

- H_1: A short stop (S=2) is greater than 60 sec or
- H_2: Taxi turns on the hazard lights or
- H_3: Taxi uses the parking break or
- H_4: Taxi turns on the left turn light and stops for more than 45 sec[9]

Hypothesis H_4 was used to capture stops where the taxi driver does not activate any safety devices when serving a passenger, while filtering out stops due to traffic congestion. The path traveled by a taxi between two consecutive passenger moves is a trip. A taxi route is an alternation of full and empty trips. A trip is considered as passenger-carrying (full) if *all* of the following hypotheses are satisfied:

[9] Vehicles in Japan drive on the left side of the streets

- H_5: Duration of trip is greater than 5min and

- H_6: Distance of trip is greater than 1.5 km and

- H_7: Distance of trip is less than twice the Euclidean distance between its end points.

Hypotheses H_5 and H_6 rule out short trips, which are unlikely to be filled with a passenger, and are not be representative of car trips. Hypothesis H_7 disregards empty taxi trips cruising for passengers and not following shorter path routes.

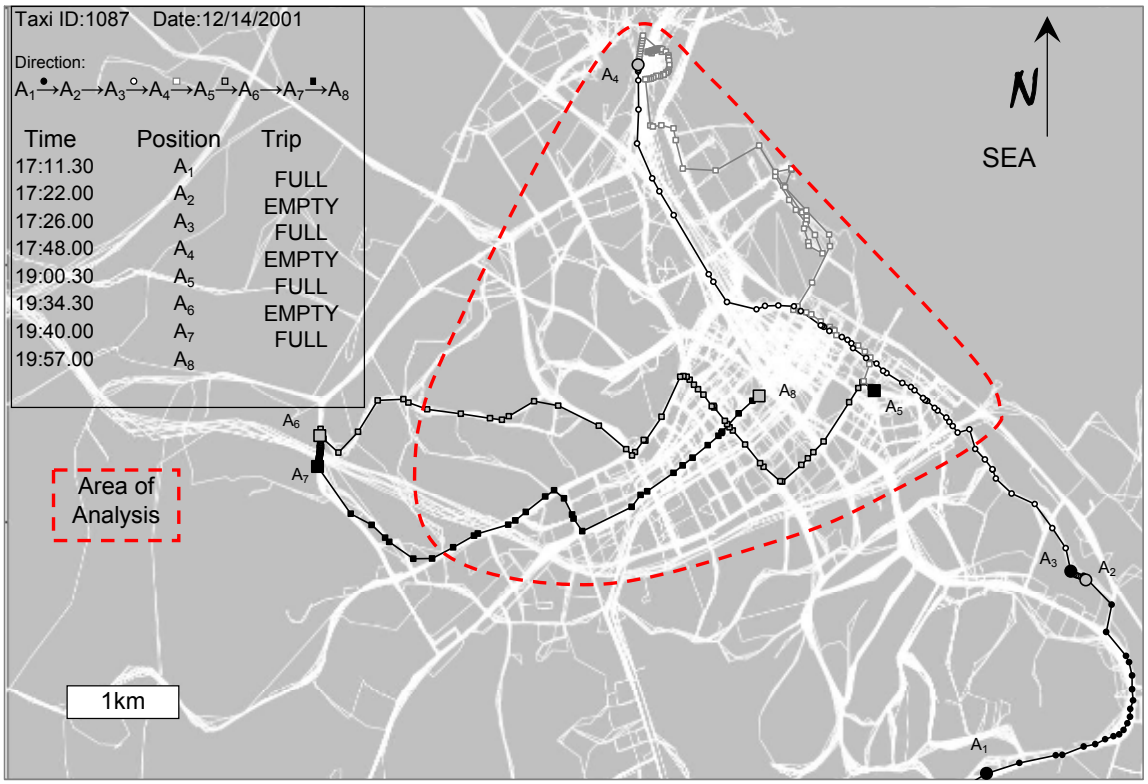


Figure 3.9: Trajectory of taxi 1087, and area map (in white) produced by a superposition of all the taxi trajectories.

Next, we apply the filtering method to the whole IPCar dataset for the same days detector data were analyzed. This method may disregard some full trips, but with high probability the subset of censored trips will be full ones. Only on very rare occasions, as

happens in reality, our method identified consecutive full trips. This indicates that the method was effective in eliminating circuitous routes or stops not because of traffic congestion. Many taxi routes were plotted and the patterns looked realistic. Figure 3.9 shows the trajectory of taxi-1807 for 3 hours. The set of streets, i.e. lines in white color, have been constructed using the complete set of taxi trajectories for a period of one week and is an accurate representation of the area when compared with a real map. The perimeter of *A* is shown by a dashed line. The large symbols A_1 to A_7 illustrate passenger moves (black for boarding and grey for alighting). Note how full and empty trips alternate. Smaller symbols mark the position of the taxi every 30 sec. Note how the distance between consecutive symbols is greater when the taxi is full than when it is not.

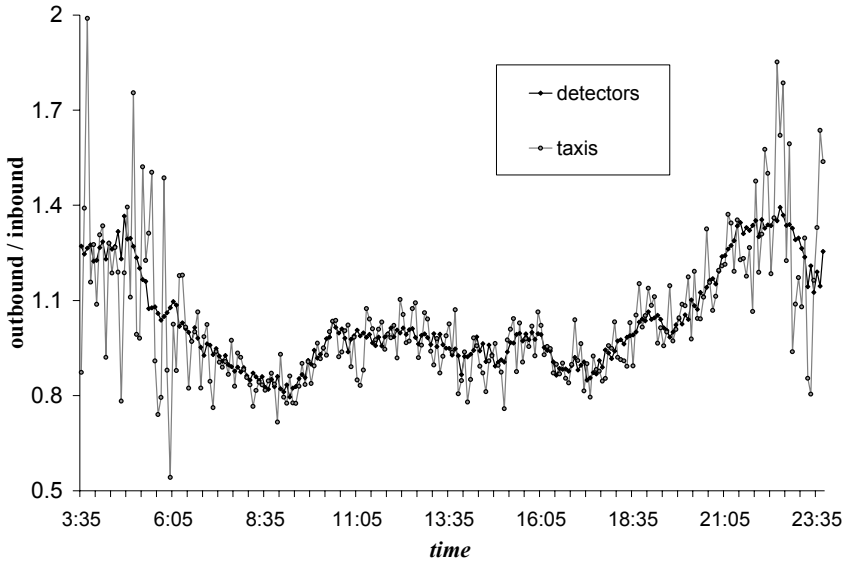


Figure 3.10: Observed outbound/inbound flow ratio for all cars (detectors) and full taxis

We have stated that full taxi trips are representative of car trips. Figure 3.10 tests this hypothesis with the only possible way given the available data; the ratio of outbound vs. inbound flows crossing the perimeter of *A* for full taxis and cars. Time series of both

ratios for full taxis and cars are shown. Note how the two curves remain close to each other throughout the day, although they vary with time. Significant fluctuations arise only when the number of full taxis is so low (<10) that fluctuations are unavoidable.

3.3.2 Estimation of speed and accumulation

The censored taxi data were aggregated into 5-min intervals. The reasons for this time aggregation are (i) to smooth the variations of traffic during a cycle because of the signals (ii) to match detector data time aggregation resolution. From this information we calculated for each time slice and for all the full taxis:

- the total distance δ traveled in A;
- the total time τ spent in A;
- their space-mean speed in A, $v_T = \delta/\tau$;
- their number in A, $n_T = \tau/\Delta t$;
- the number N_T that exited A along its perimeter and
- the number M_T that finished a passenger-carrying trip inside A.

These data are used to estimate the space-mean speed v and accumulation n of all cars in A to test if, despite the statistical errors due to the low number of taxis, an MFD as in Figure 3.6b appears. We use v_T as our estimate $\hat{v}$ for v, since valid taxi trips are assumed to be typical trips; i.e.:

$$\hat{v} \equiv v_T \cong v \tag{3.1}$$

The estimation of accumulation is not so straightforward. If now "full" taxis behave identically with ordinary cars, then for any time t the probability π that a randomly selected car moving in area A is a taxi is the same across (i) all streets in area A, (ii) for points belonging in the perimeter ∂A and (iii) for points belonging in the perimeter for streets with detectors $\partial A'$. If n is the number of vehicles in A; N is the number of vehicles exiting A along its perimeter; N' and N'_T are the numbers of vehicles and taxis exiting A' along streets with detectors, then the following three, are unbiased estimators for the estimation of probability π: the ratio of cars and taxis observed (i) in area A (ii) in the perimeter ∂A and (iii) in the perimeter $\partial A'$: [10]

$$\hat{\pi} \equiv \frac{n_T}{n} \cong \frac{N_T}{N} \cong \frac{N'_T}{N'} \ . \tag{3.2}$$

Thus, an estimate $\hat{n}$ was constructed as follows (Figure 3.11 illustrates this estimation): Firstly, N' is measured by the detectors, and N'_T is approximated as a fixed proportion of N_T, which is also observed. Measurements of N'_T could not be automatically extracted from the data base because the detector positions and taxi data were not linked to a digital map of Yokohama. We used $N'_T \cong 0.7N_T$. The factor "0.7" was estimated manually, after following for a whole day the routes of 10 taxis on the map of Figure 3.9, and determining from this map whether each exit point belonged either to a street with detector or not. Figure 3.12 displays the result of this effort. Note how the fraction of exits from streets with detectors varies little from 0.7. Therefore, we can estimate n with $\hat{n}$, as follows:

[10] Note that if accumulation fluctuates rapidly during the time of a trip the conditions of equation (3.2) may not be satisfied. But it is intuitive that this is not the case for an area A large comparable with a street link.

$$n \cong n_T N' / N'_T \cong n_T N' / [0.7 N_T] \equiv p n_T \equiv \hat{n}\,, \tag{3.3}$$

where $p \equiv N'/[0.7N_T]$ is an observable expansion factor, which approximates the ratio of vehicles vs. full taxis exiting the zone from streets with detectors. Next subsection presents the estimation results.

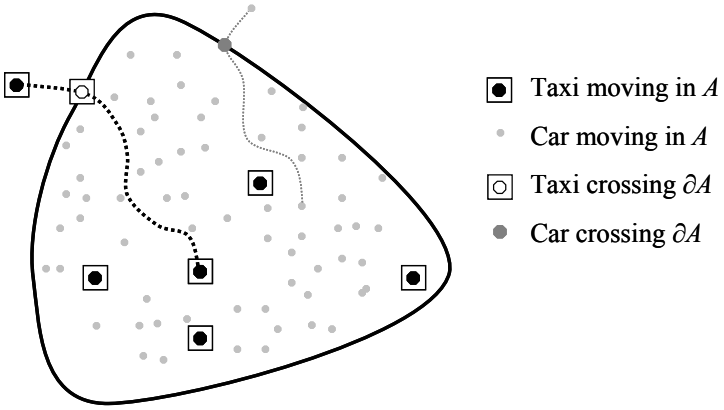


Figure 3.11: Moves of taxis and cars along the network

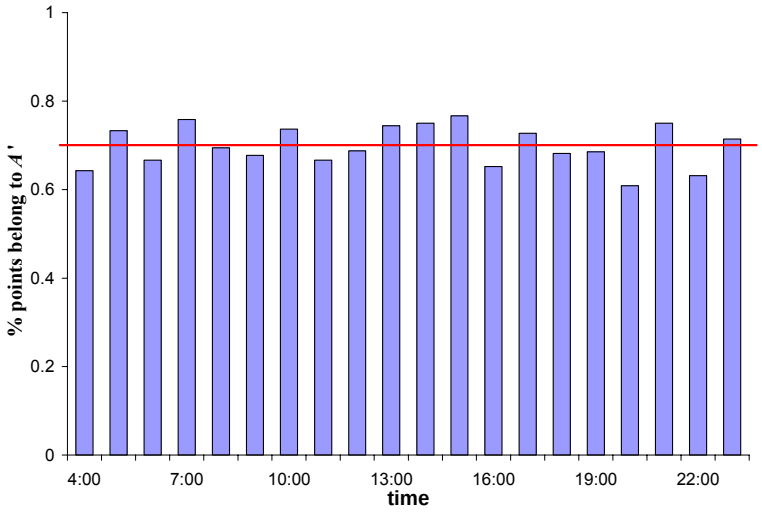


Figure 3.12: Fraction of full taxis exiting the perimeter of A that exit through A' at different times of the day

3.3.3 Existence of an MFD in A: Estimation results

Subsection 3.3.2 described the methodology to estimate accumulation and average speed for the whole area A based on the available data. However, if δ and τ are low (as occurs at night) the estimate could have a significant error due to small sample size. To alleviate this problem we aggregated our data and used Δt = 30 min. Average speed of taxis and exit flow from the detectors are smooth enough to allow a 30min aggregation with no loss of information regarding the traffic patterns. This aggregation ensures that the coverage of A is reasonable when the number of full taxis in A is on the order of 10 -- a value that is consistently exceeded during the day time.

Figure 7a is a scatter plot of $\hat{v}$ vs. $\hat{n}$ with Δt = 30 min. Time slices are displayed every 5 min. Triangles are used for the morning and circles for the afternoon and evening. Triangles of consecutive time slices are linked by a dark line. The figure shows that the pattern is the same at all times of the day, without hysteresis. We conclude that an MFD exists in A.

But, what about the errors in the estimator $\hat{n}$? This can be explained by considering the low number of taxis. These errors are of three types: (i) error in the estimation of speed; (ii) error in the estimation of N'_T from N_T and (iii) error in the estimation of π from N'_T /N'. The first error is not important because average speed over 30min interval is representative of a region which is roughly homogenously congested. The second error is not substantial, as shown in figure 3.12.

The most severe error is the third one. We can approximate the number of taxis leaving randomly A' as a Poisson distribution. The reason for this is that the probability

of more than one taxis exiting at the same interval (t, $t+\Delta t$) with $\Delta t \to 0$ is about zero. Thus, it is assumed that the number of occurrences (taxis exiting) follows the *law of small numbers***,** which describes events that happens rarely, but have very many opportunities to happen (Bortkiewicz, 1898). We consider that this approximation is valid for $N'_T > 25$, because when zero taxis are exiting accumulation cannot be estimated from equation 3.3.[11]

We now show that the average error for $\hat{n}$ is about $N'^{-1/2}_T$ .Both the mean and the variance for the number of taxis exiting are equal to $n_T N'/\hat{n}$. Then, a reasonable range for the number of taxis exiting is the average value plus or minus one standard deviation:

$$N'_T \cong \frac{n_T \cdot N'}{\hat{n}} \pm \sqrt{\frac{n_T \cdot N'}{\hat{n}}} \quad . \tag{3.4}$$

Solving for $\hat{n}$ using the top of the range (equation 3.4 with a plus sign) we get:

$$\hat{n} \cong \frac{4n_T \cdot N'}{\left(-1+\sqrt{4N'_T}\right)^2} \cong \frac{n_T \cdot N'}{N'_T - \sqrt{N'_T}} = \frac{\dfrac{n_T \cdot N'}{N'_T}}{1-\dfrac{1}{\sqrt{N'_T}}} . \tag{3.5}$$

The second approximate equality is accurate since $\sqrt{4N'_T} \gg 1$ for $N'_T > 25$. Finally, by applying a first order Taylor expansion in $1/\left(1-N'^{-1/2}_T\right)$ we get:[12]

[11] The probability of zero taxi exits in 5min interval for a Poisson Distribution with mean equal to 25exits/30min is <1.5%.

[12] We know that $\frac{1}{1-x} \cong 1+x+\ldots$ for 0<x<1

$$\hat{n} \cong \frac{n_T \cdot N'}{N'_T} - \frac{n_T \cdot N'}{N'_T} \cdot \frac{1}{\sqrt{N'_T}} . \qquad (3.6)$$

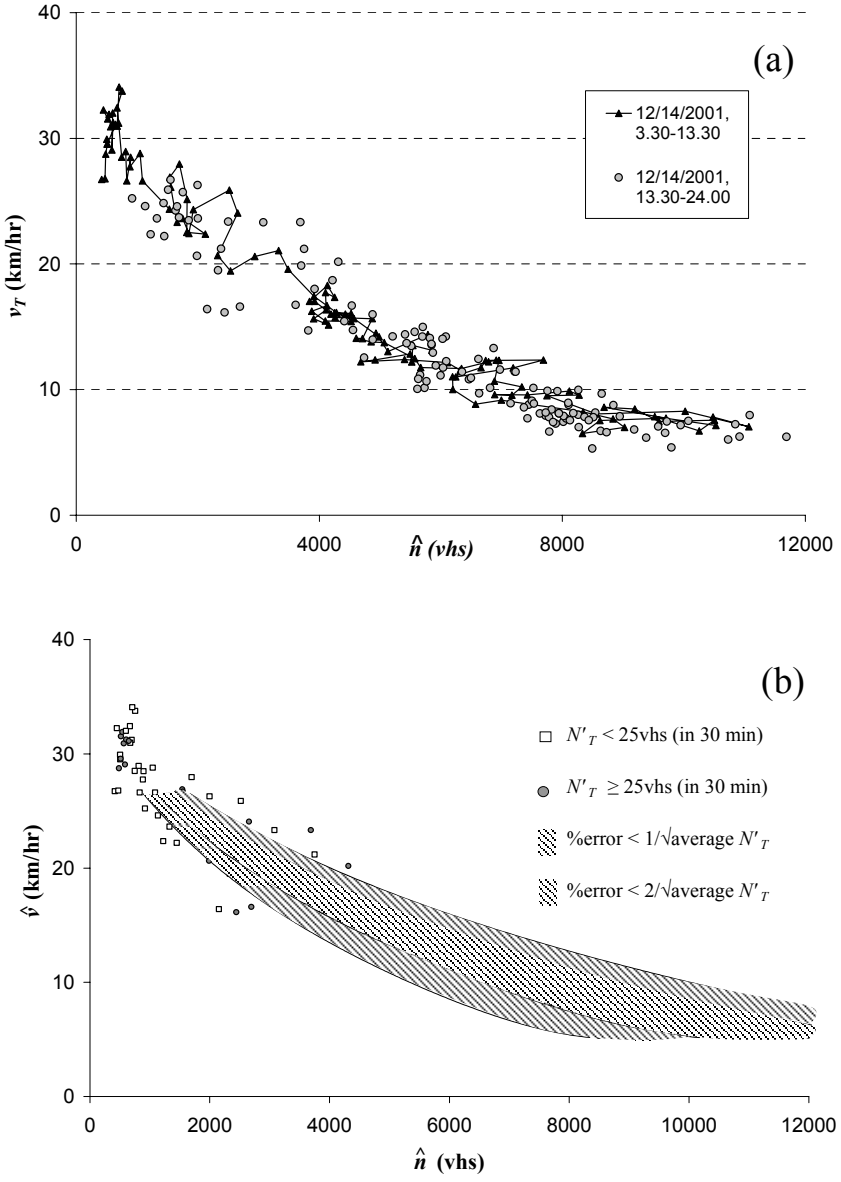


Figure 3.13: Yokohama's estimated MFD: (a) Scatter plot of $\hat{v}$ vs. $\hat{n}$; (b) 1- and 2-standard deviation bands.

By repeating this procedure for Equation 3.4 with a minus sign, and solve for $\hat{n}$, we get the same result with Equation 3.6 with a plus sign. Figure 3.13b shows the 1- and

2-standard deviation bands arising from this formula on each side of a fitted curve. Regarding the fitted curve, we estimated the best 4th degree polynomial that fits Figure 3.6a, and scaled it horizontally by the factor L= 131 lane-km, that resulted in the best agreement. This factor, however, is not necessarily the true length of A. The bands only apply to the points with $N'_T > 25$.

The white squares arise mostly during the night when the number of taxis is small (<10) and their higher scatter should be expected both because the approximation does not hold when N'_T is low, and also because speed estimation for small number of taxis introduces significant errors in the vertical direction.

3.3.4 Existence of a linear relation between the trip completion rate in *A* and total production in *A*

Using the detector data, we showed in Subsection 3.3.2 that the outbound perimeter flow D' is highly correlated with the average network flow of A', q^w. This perimeter flow included a considerable part of the trip-ends, but not the internal ones. To estimate the trip completion rate O at which vehicles depart the network including trips that end within the study area, we expand the rate at which full taxis exit A or finish a trip within A as follows:

$$\hat{O} = p \cdot \left(N_T + M_T \right) / \Delta t \tag{3.4}$$

The space-mean network flow q cannot be estimated in the same way because we do not know the total network length L required to express the space-mean flow of full taxis in A, which is $\delta / \Delta t L$. Therefore, instead of average flow we estimate travel production P

and focus on the relation between production and the trip completion rate. Since the full-taxi production in A is $\delta/\Delta t$, we estimate total production in A as:

$$\hat{P} = p\delta/\Delta t. \tag{3.5}$$

Figure 3.14 is a time-series of $\hat{P}/\hat{O}$, which supports the hypothesis that $\hat{P}/\hat{O}$ is relatively constant and equal to about 2.3 km, i.e. a trip is completed on average for every 2.3 km traveled. Again a 5-degree polynomial does not improve the fit in a statistically significant way, and the best fitted 5th degree polynomial has an RMS < 0.1% (See figure 3.15 for values of coefficients and statistical tests).This constant is the average vehicular trip length in A.

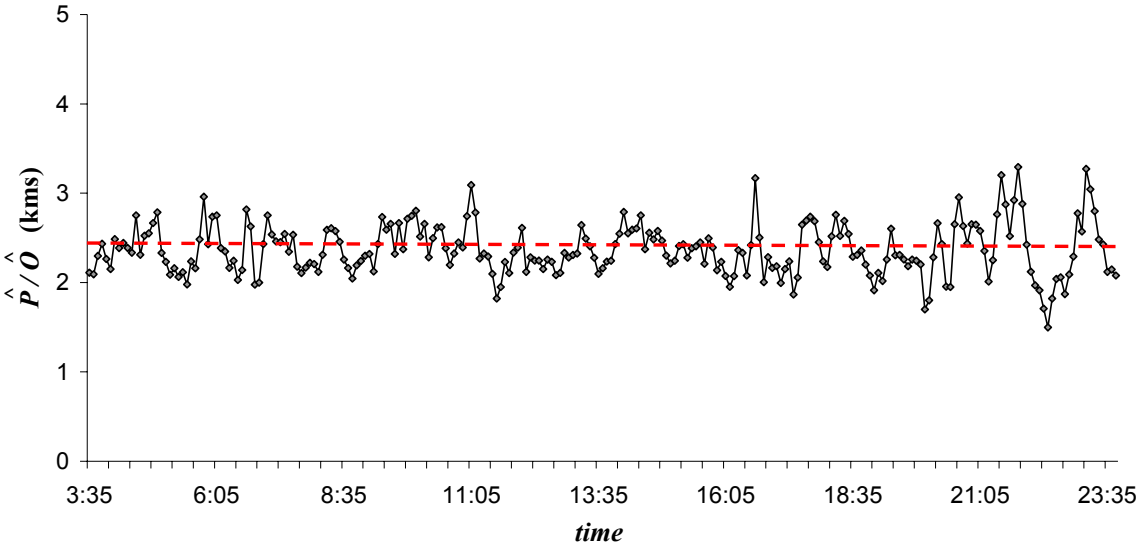


Figure 3.14: total production over trip completion rate time-series (12/14/2001)

	Coefficients	t Stat	P-value
Intercept	2.333	20.23	$<10^{-20}$
t	4.07×10^{-4}	0.04	0.97
t^2	5.18×10^{-5}	0.22	0.83
t^3	-9.35×10^{-7}	-0.38	0.71
t^4	5.17×10^{-9}	0.46	0.65
t^5	-9.13×10^{-12}	-0.50	0.62

Figure 3.15: Statistical test for a 5th degree polynomial of $\hat{P}/\hat{O}$ with time t

3.3.5 Prediction of outbound perimeter flows

As the macroscopic fundamental diagram for Downtown Yokohama is reproducible among different days and for different demand profiles, it seems that we can predict exit flows (a measure of accessibility) from either occupancy or accumulation or speed data. Results shown in figure 3.16 demonstrate this postulate.

Figure 3.16a shows a scatter plot of exit flows (outbound perimeter flows) measured from the detectors and average speed of taxis for 12/14/2001. First, we estimated the best-fit 4th degree polynomial to these points. Next, we applied the filtering method of section 3.3.1 for a different day (12/6/2001) and calculated average speed of passenger-carrying taxis. We also predicted exit flows for this day using speed of full taxis, v_T, and the fitted curve, and we measured the exit flow from streets with detectors in the perimeter of area *A'*. Figure 3.16b shows an oblique plot of the predicted and the measured cumulative exit flow. The two curves are quite close (maximum error less than 2%).

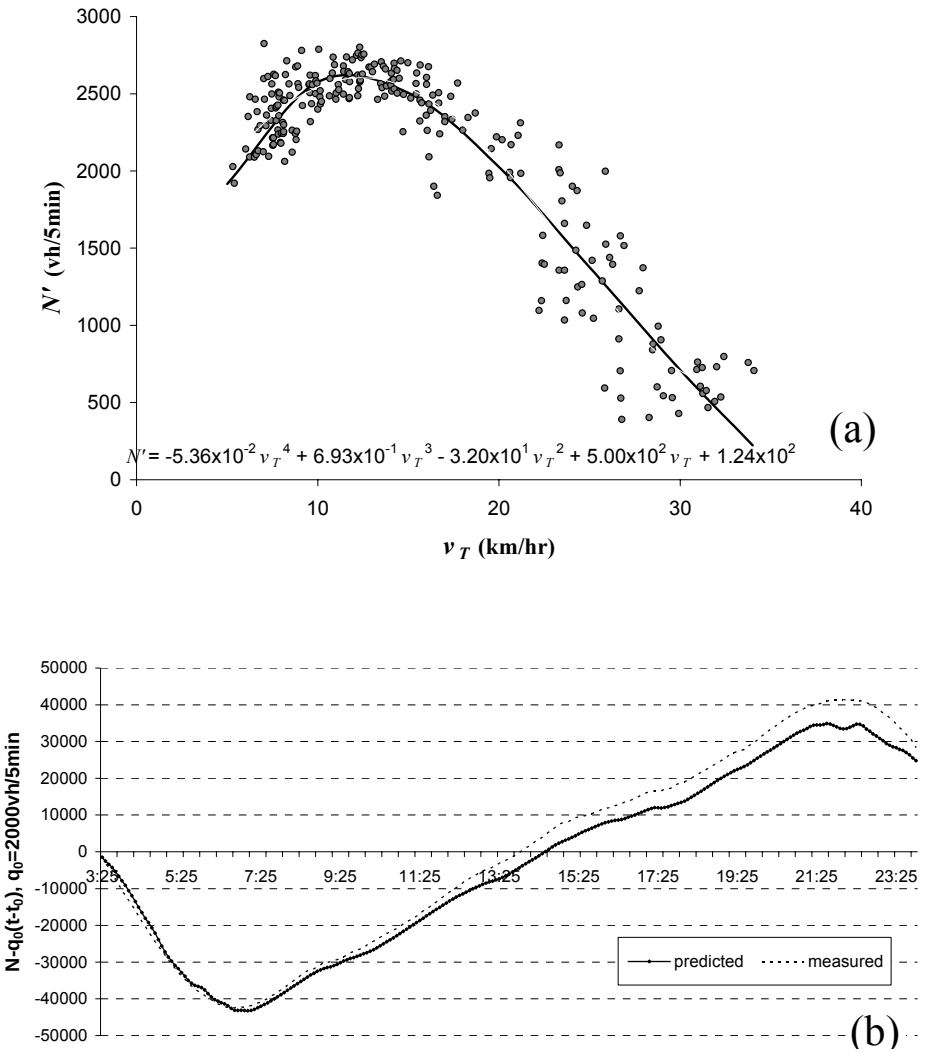


Figure 3.16: Prediction of exit flows from streets with detectors:
(a) exit flow vs. average taxi speed (12/14/2001); (b) Oblique plots of predicted vs. measured exit flow (12/6/2001)

3.4. Final Remarks

In summary, the results of chapters 2 and 3 show that (i) an MFD exists on neighborhood-sized sections of cities independently of the demand and (ii) that it can be used to control demand and improve mobility. Also, the findings of section 3.3 are important (i) because they establish that the MFD is not a property of the detectors'

locations (*only of the network itself*); (ii) because they show how to estimate accumulation and speed for the whole network and (iii) because perimeter control can be more easily applied to the whole network than to the part which includes only streets with detectors. A question that arises from the above is whether cities experience congestion with production-accumulation pairs in regime III, with speeds considerably less than 50% of the average speed during the off-peak.

The answer is yes. In many European cities, speeds in the centre are under 10 kilometers per hour at midday and slower during the rush. For example, the Hellenic Institute of Transportation Engineers (HITE) reports that during year 2005 average speeds in major arterials in the city of Athens during peak hours were three to five times smaller than the off peak average speeds (e.g. Alexandras Avenue – 5km/hr for peak hour, 25km/hr for non peak, Mesogion Avenue – 8km/hr for peak hour, 40km/hr for non peak) (HITE, 2006). Developing countries face up similar problems. Drivers in Bangkok spend the equivalent of 44 days a year in gridlock (Ressler, 1999). This suggests that the development of control strategies to relieve congestion and increase mobility could have a significant payoff.

This can be done with pricing, rationing and/or perimeter control strategies based on neighborhood accumulation and speeds. Simple versions of these strategies are already being used in many cities around the world: e.g., in London, Stockholm and Singapore (congestion pricing); in Beijing -- a test in anticipation of the 2008-Olympiad -- and Mexico City (rationing); and in Zurich (perimeter traffic control). But by knowing the MFD and monitoring the state of traffic continuously, transportation managers can now

see whether their system is in a state that is producing the desired levels of mobility at all times. Therefore, existing strategies can be redefined.

Building in the knowledge of Chapters 2 and 3, macroscopic control strategies are explained in Chapter 5. But, prior to developing efficient control strategies to improve mobility and avoid overcrowdness, we need to model the dynamics of the rush hour. This is described in details in Chapter 4.

A question that arises after the analysis of this chapter is that, unlike taxi trips, some automobile trips with internal destinations include a “looking-for-parking” portion that extends their length, and this extension increases accumulation. This effect should be minor in this study because 70% of the trips were found to have external destinations and, surely, a significant portion of internal trips have pre-assigned parking. But, this is not necessarily the case. The looking-for-parking phenomenon will also be presented and analyzed in Chapter 4.

Chapter 4

Dynamics of multi-neighborhood cities

Prior to developing control strategies, which will improve the overall mobility of a city, we need to describe the rush hour in a congested city dynamically. Daganzo (2007) derived an ordinary differential equation (Equation 1.4) for the system dynamics of an "one neighborhood (reservoir)" city. This chapter develops the dynamics of cities with more complicated structure. Section 4.1 presents a general model of an N-neighborhood city and an example for N=2. In this model, it is assumed that when vehicles reach their destination find a parking spot with zero delay. Section 4.2 models the cruising-for-parking phenomenon that under certain conditions can lead to heavy congestion. This model is consistent with the physics of traffic and its variables are observable quantities. We show that we can obtain some additional insight about the key parameters that matter during congestion. Finally, section 4.3 provides some final remarks.

4.1 Dynamics of the rush hour

One can model a city as a single or multi-reservoir system depending on the geometry, the demand patterns and the distribution of trip destinations among the city. The requirement for homogeneity in traffic loads and slow variations in time should determine the number N of required reservoirs and the time scale of the model. Fine time scales and large $N>10$ are not recommended because they: (i) make the accuracy of the macroscopic relations questionable and (ii) require the knowledge of detailed origin-destination tables.

Consider now a city partitioned in N reservoirs. Denote by $i=1,\ldots, N$ a reservoir in the system, n_i its accumulation and n_{ij} the number of vehicles in i with final destination reservoir j, at a given time. Let P_i, $V_i \doteq P_i/n_i$ and $O_i = P_i/l_i$ be the production, average space-mean speed and output at a specific time, where l_i is the average trip length for reservoir i. This average is assumed to be independent of time and destination, internal or external, in i. Output O_i is the sum of the exit flows O_{ij} ($\forall j\neq i$) from i with final destination j, plus the internal output O_{ii} (internal trip completion rates at i).

We assume that for each reservoir i: (i) there exists a Macroscopic Fundamental Diagram (MFD), $P_i(n_i)$, between accumulation n_i and production P_i, which describes the behavior of the system when it evolves slowly with time and (ii) that Little's formula (Little, 1961) stands for each of the reservoirs, i.e. output is a function of accumulation, $O_i(n_i) = P_i(n_i)/l_i$. Also, by definition, we have that: $V_i(n_i) \doteq P_i(n_i)/n_i$.

It is also assumed that there exists an entrance function $C_{ij}(n_j)$, which describes the maximum inflow or *inflow capacity* to reservoir j from an adjacent reservoir i as a

function of accumulation its n_j. The causality of function C_{ij} is that sufficiently large accumulations in j restrict its inflow along the periphery. If the inputs, demand and controls change slowly with time the system should be near equilibrium all the time, and functions G, O, C and V should also, describe the system in the dynamic case. We observe that the entrance function $C_{ij}(n_j)$ is roughly invariant for a range of accumulation up to a critical value and then decreases for larger values. But, how C_{ij} is related to P_i?

Figure 4.2 shows how C_{ij} changes with accumulation for the San Francisco network, described in Chapter 2, for different runs. Note, however by comparing figure 2.4a and 4.2 that inflow capacity stays nearly constant for accumulations well into the undesirable regime III. Thus, the system should not be allowed to self-regulate. The figure also shows that inflow capacity is insensitive to different demand patterns for accumulations beneath the critical value, but is less predictable for values of accumulation in regime III.

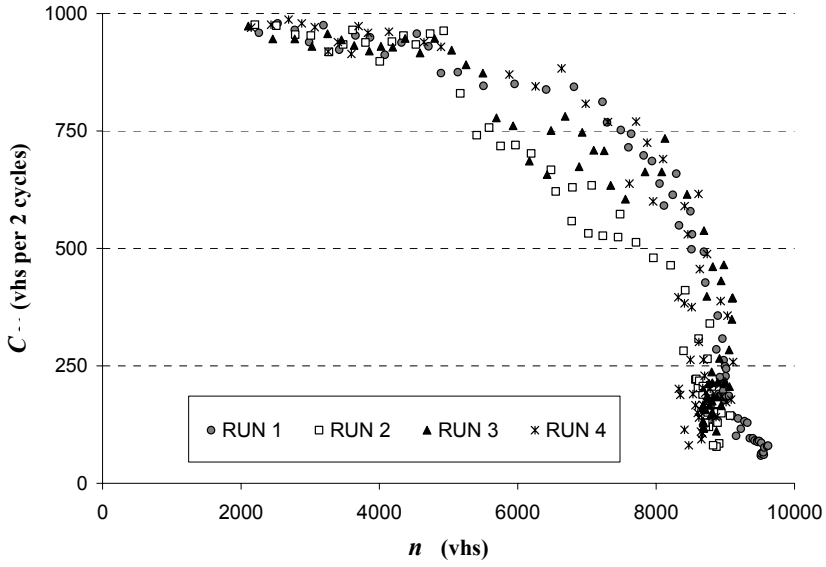


Figure 4.2: Inflow capacity vs. accumulation for the San Francisco network for different runs

Theorem 4.1: *The fraction of exits from i, that correspond to destination j (O_{ij}/ O_i), are expressed as* $\left(n_{ij}/n_i\right)\cdot a_{ij}$, *where* a_{ij} *is a constant inversely proportional to trip length* l_i.

Proof: Using Little's formula: (i) for the total accumulation of reservoir *i* and (ii) only for vehicles in *i* with final destination *j*, we have that (i) $n_{ij} = O_{ij}\cdot l_{ij}/V_i$ and (ii) $n_i = O_i\cdot l_i/V_i$, where l_{ij} is the average distance traveled in *i*, by all vehicles in *i* with destination *j*, and V_i is the common speed for both families of vehicles.[13] Thus, we have:

$$O_{ij} = O_i\cdot\frac{n_{ij}}{n_i}\cdot\frac{l_i}{l_{ij}} \doteq O_i\cdot\frac{n_{ij}}{n_i}\cdot a_{ij} = V_i\cdot\frac{n_{ij}}{l_{ij}}, \tag{4.1}$$

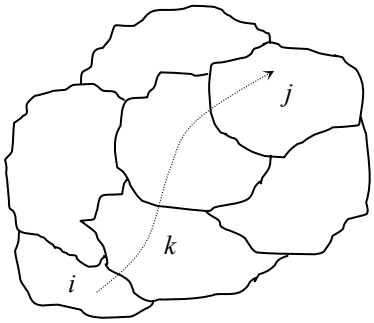


Figure 4.1: A multi-reservoir system

Let $q_{ij}(t)$ be the exogenous flow generated in reservoir *i* with destination reservoir *j* at time *t* and $\delta^j_{i\to k}$ be an exogenous binary variable with value equal to 1 if a trip from reservoir *i* to *j*, passes through *k* immediately after leaving *i*, and 0 otherwise. It is assumed that all trips generated from *i* with destination *j*, follow the same sequence of reservoirs in their route. We could now list endogenous variables. Let $q^j_{i\to k}(t)$ be the transferring flow from reservoir *i* to *k* at time *t*, with final destination *j* (see Figure 4.1);

[13] Evidence from the Yokohama's experiment (Chapter 3) showed that α_{ij}=1 for different types of vehicles, which simplifies equation 4.1.

$x_{ij}^{*}(t)$ be a control variable for the inflow capacity for movements from i to j that an engineer would set, then the dynamic equations between the state variables n_{ij}, the auxiliary variables $q_{i\to k}^{j}$, the control variables x_{ij}^{*} and the input q_{ij} of a system with N reservoirs are (time t is omitted from the equations; i, j, k= 1,2…, N):

$$\frac{dn_{ij}}{dt} = q_{ij} - \sum_{k=1}^{N} q_{i\to k}^{j} + \sum_{k=1}^{N} q_{k\to i}^{j}, \tag{4.2}$$

where

$$q_{i\to k}^{j} = \begin{cases} \min\left(x_{ik}^{*} \cdot a_{i\to k}^{j}, C_{ik}\left(n_k\right) \cdot a_{i\to k}^{j}, V_i\left(n_i\right) \cdot \dfrac{n_{ij}}{l_{ij}} \right), & \text{if } \delta_{i\to k}^{j} \neq 0, \\ 0, & \text{otherwise.} \end{cases} . \tag{4.3}$$

Equation (4.2) states that the rate n_{ij} changes over time equals to the internally generated flow from i to j minus the outflow from reservoir i with final destination j plus the inflow from all reservoirs adjacent to i with final destination j. Given our assumption, the factor $a_{i\to k}^{j}$ in equation (4.3), which is the fraction of inflow capacity C_{ik} or applied control rate x_{ik}^{*} from reservoir i to k, for vehicles with final destination j, is (one can derive this by applying Little's formula):

$$a_{i\to k}^{j} = \frac{n_{ij}/l_{ij}}{\sum_{m=1}^{N} \delta_{i\to k}^{m} \cdot n_{im}/l_{im}} \tag{4.4}$$

Equation (4.3) merely explains that the transferring flow from reservoir i to k with final destination j is the smallest of (i) the applied control rate (ii) the inflow capacity for these movements and (iii) the exiting flow from reservoir k with destination j, as given from the exit function $O_i(n_i)$.

Let us look how dynamic equations (4.2) and (4.3) are simplified in a case of a city partitioned in two reservoirs, R_1 and R_2. This could be a case of a two-nested-reservoirs city where the internal reservoir R_1 (city center) attracts most of the trips during the morning commute and the external reservoir R_2 (periphery) generates most of the trips at the same time period. If $l_{ij} = l_i, \forall i, j = 1,2$ and vehicles do not reenter in a reservoir they exited from, the dynamic equations for this system are:

$$\frac{dn_{ij}}{dt} = q_{ij} - \min\left(x_{ij}^*, C_{ij}\left(n_j\right), \frac{n_{ij}}{n_i} \cdot O_i\left(n_i\right) \right), \quad i \neq j, \tag{4.5}$$

$$\frac{dn_{ii}}{dt} = q_{ii} + \min\left(x_{ji}^*, C_{ji}\left(n_i\right), \frac{n_{ji}}{n_j} \cdot O_j\left(n_j\right) \right) - \frac{n_{ii}}{n_i} O_i\left(n_i\right), \quad i \neq j. \tag{4.6}$$

The second term of the right hand side, for both equation (4.5) and (4.6), represent the intertransfers between the reservoirs; while the third term of equation (4.6) represents the rate vehicles finish their trips inside reservoir *i*.

An important remark is that the middle term in the parenthesis of (4.5) and (4.6) can be ignored. The reason is that C_{ij} is invariant for values of accumulation in regimes I, II and part of regime III. But, any control strategy will try to avoid accumulations in the undesirable regime III. Thus, (i) $x_{ij}^* < C_{ij}$ when the system reaches states in regime III and (ii) $O_i < C_{ij}$ for regimes I and II.

There is a caveat in the model of the two-reservoir system. If we control the input by restricting the perimeter capacity, we may induce uneven distribution of vehicles on the sending reservoir. This may invalidate the homogeneity assumption of traffic loads within a reservoir (if the reservoir is too big) and reduce the accuracy of the model. In

that case, accuracy can be improved by using smaller reservoirs, i.e. an "onion structure" with 3 or 4 nested reservoirs. The issue does not arise with other forms of control, e.g. pricing.

Next section analyzes the effect of looking-for-parking in the output $O_i(n_i)$ and the dynamics of the rush hour.

4.2 The effect of cruising-for-parking

Cruising for parking creates a mobile queue of cars that are waiting for curb vacancies and is an important source of congestion (Shoup, 2006). For example, during peak hour in the area around Harvard square in Massachusetts, 30% of moving vehicles are searching for parking, with an average searching time of 12 min (O' Malley, 1985). Most of the literature in the effect of parking in traffic is (i) policy oriented--see Shoup (2005) for a detailed description; or (ii) analyzes the economics of parking (e.g. Douglas (1975), Glazer and Niskanen (1992), Arnott et al (1992), Arnott and Rowse (1999), Calthrop et al (2000), Anderson and De Palma (2004) etc).

Some useful research in identifying interactions between cruising for parking and traffic congestion from an economic prospective was made by Arnott and Inci (2006) and Arnott and Rowse (2007). They modeled parking in the steady state by considering three different types of vehicles: (i) moving, (ii) cruising and (iii) parked. Although they provide some useful relations between these three types and the economic impacts of the phenomenon, they fall short in describing (i) the dynamics of parking with realistic physics and (ii) how cruising-for-parking can lead to congested traffic conditions. This

section shows how the effect of parking can be easily modeled with the framework of section 4.1.

4.2.1 Modeling the cruising-for-parking phenomenon

The model presented in this section fills three important gaps of the existing policy- and economics-oriented models: (i) it describes the physics of overcrowding; (ii) shows that cruising-for-parking affects all the users of the system, even those with destinations outside the "limited parking region" and (iii) provides tools to estimate the direct costs of all users, as these expressed by additional vehicle-hours traveled. Also, it differs from the existing economic models because (i) it can describe the dynamics of parking phenomenon and most importantly (ii) it contains variables and functions that are readily observable or can estimated with field experiments (e.g. this is not possible for a demand curve that most of the economic models assume). Therefore the laws of behavior can be verified, and the models are expected to produce reliable outputs.

Consider a multi-reservoir city, with infinite parking availability for all the reservoirs, except one (reservoir R), where vehicles may have to cruise for parking because of limited on-street parking. We focus the analysis on on-street cruising for parking at R. Its accumulation n is the sum of three families of vehicles: (i) vehicles searching-for-parking n_s (family $\boldsymbol{s}$); (ii) vehicles moving towards their destination internal to the reservoir, but not yet searching for parking, n_m (family $\boldsymbol{m}$), and (iii) vehicles moving with external destinations, n_o (family $\boldsymbol{o}$):

$$n = n_s + n_m + n_o . \tag{4.7}$$

Denote by n_p the vehicles parked on street (family ***p***) and by N_p the total number of parking spots. Also, let $p \doteq n_p / N_p$ be the percentage of available parking spots, and d_1 be the average distance traveled between two adjacent spots (from N_p). We assume that p and d_1 do not vary drastically in the dimension of space. When a vehicle reaches its destination, it is assumed that it does Bernoulli trials with probability of success p, until finding an available spot. The number of trials until the 1st success follows a geometric distribution with mean $1/p$. In each trial a vehicle travels, on average, distance d_1 and the average distance traveled while searching for parking is $l_s=d_1/p$. To model the effect of parking, the reservoir R is divided in three sub-reservoirs:

(i) A moving reservoir R_{mo} with $n_m + n_o$ vehicles

(ii) A searching reservoir R_s, where vehicles transfer from R_{mo}, when they arrive close to their destination and

(iii) A parking reservoir R_p, where vehicles transfer from R_s when they find a parking spot. Also, trips generated in R_p, transfer from R_p to R_{mo} and lead to their destinations. Figure 4.3 shows this partitioning and movements between different families.

The output for sub-reservoirs m, s, o, is estimated using Little's formula (see also equation 4.2) in R:

$$O_x\left(n_x, n\right) = \frac{n_x}{n} \cdot \frac{P(n)}{l_x}, \quad x \in \{m, s, o\}. \tag{4.8}$$

Output from family ***m*** is input to family ***s***, output from ***s*** is input to ***p***, output from ***p*** is input to ***m*** and ***o***. Output from reservoir i is the sum of outputs from ***s*** and ***o***.

In the case of $x=s$ (searching), equation (4.8) becomes:

$$O_s = \frac{n_s}{n} \cdot \frac{p}{d_1} \cdot P(n). \tag{4.9}$$

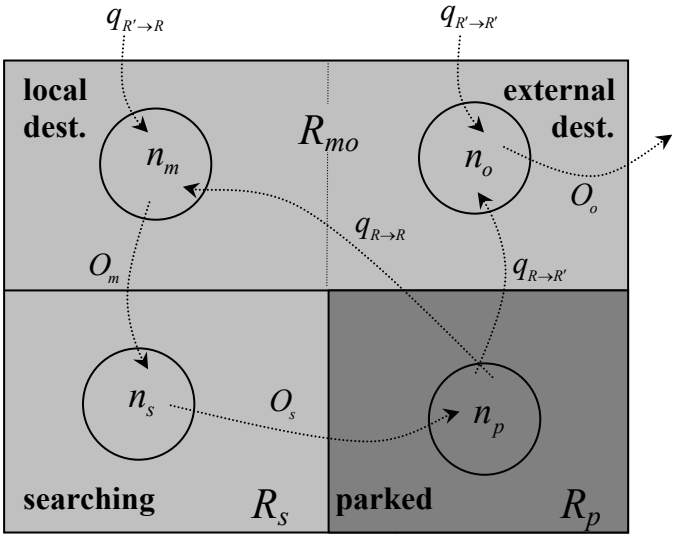


Figure 4.3: Different movements to model cruising-for-parking phenomenon

The average cruising time is $l_s/V(n) = d_1/\left(p \cdot V(n)\right)$. The state of reservoir R (n_m, n_s, n_o, n_p) is described in the dynamic case as follows (time t is omitted from the equations):

$$\frac{dn_m}{dt} = q_{R'\to R} + q_{R\to R} - O_m(n_m, n) \tag{4.10}$$

$$\frac{dn_s}{dt} = O_m(n_m, n) - O_s(n_s, n) \tag{4.11}$$

$$\frac{dn_o}{dt} = q_{R'\to R'} + q_{R\to R'} - O_o(n_o, n) \tag{4.12}$$

$$\frac{dn_p}{dt} = O_s(n_s, n) - q_{R\to R} - q_{R\to R'} \tag{4.13}$$

where $q_{R\to R}$ and $q_{R\to R'}$ is the rate trips starting from R_p with internal (inside R) or external destinations (outside R) ; $q_{R'\to R}$ and $q_{R'\to R'}$ are the flows entering from outside with internal or external destinations.[14]

When conditions change slowly with time($dn_s/dt \simeq 0$), note that equations (4.11) and (4.8) yield:

$$\frac{n_s}{n_m} = \frac{1}{p}\cdot\frac{d_1}{l_m}. \tag{4.14}$$

This suggests that by measuring p, the % of available parking spots and estimating the two lengths of the RHS of equation (4.14), we approximate the fraction of vehicles searching for parking. As expected, small values of p, lead to a high number of vehicles searching for parking, and as a result the average distance traveled per trip completion in R, l, increases.

Why do we look for l? The cost of cruising-for-parking, in term of additional vehicle miles traveled can be substantial. If, for example, 20% of vehicles search for parking, distance traveled per trip completion is 25% higher, which results to 20% smaller outputs, even for the same average speed. But the effect is even worse, as average speed V is a decreasing function with accumulation and smaller outputs cause higher accumulations. Especially, when the system enters regime III of the Macroscopic Fundamental Diagram, production decreases with accumulation and system can lead to gridlock. Quantitatively, we can estimate l using Little's formula for the total output O, i.e. the sum of O_s and O_o:

[14] This model neglects trips starting from R but not from R_p (e.g. trips from garages). This assumption can be easily relaxed.

$$O = O_s + O_o \Leftrightarrow \frac{n \cdot V}{l} = \frac{n_s \cdot V}{l_s} + \frac{n_o \cdot V}{l_o} \Leftrightarrow$$
$$\frac{1}{l} = \frac{n_s}{n} \cdot \frac{p}{d_1} + \frac{n_o}{n} \cdot \frac{1}{l_o} \tag{4.15a}$$

Now using (4.14) and assuming l_o=l_m, we get

$$\frac{1}{l} = \left(1 - \frac{n_s}{n}\right) \cdot \frac{1}{l_m}. \tag{4.15b}$$

We now estimate the total delay, d_p, caused by the cruising-for-parking effect by comparing the actual vehicle hours of travel with those when $N_p \to \infty$. This delay is not only experienced by vehicles cruising for parking, but also by all moving vehicles**,** because of decreased speeds. The delay d_p during a time period [0,T], can be estimated using the queueing diagram of figure (4.5), where $n(t)$ is the actual accumulation in i, at time t and $n^*(t)$ the accumulation when $N_p \to \infty$. This graph merely explains that (i) the difference in accumulation with and without cruising, at time t, is the difference between their cumulative outputs $\sigma(t)$ and $\sigma^*(t)$, at time t, and (ii) delay d_p is the shaded area between the cumulative outputs.

While $n(t)$ can be estimated for real networks (e.g. using methodology from section 2.3), the state of the system for when $N_p \to \infty$ is unknown. Nevertheless, $n^*(t)$ can be approximated as:

$$n^*(t) = n(t) - \int_0^t \left(\frac{P(n^*(\tau))}{l_m} - \frac{P(n(\tau))}{l(\tau)} \right) d\tau \tag{4.16}$$

where $l(t)$ is given by Equation 4.15. The term inside the integral is the decrease in output, at time t, due to cruising-for-parking phenomenon. Next subsection presents a

simulation of the phenomenon and gives additional insights in the causes and the results of the phenomenon.

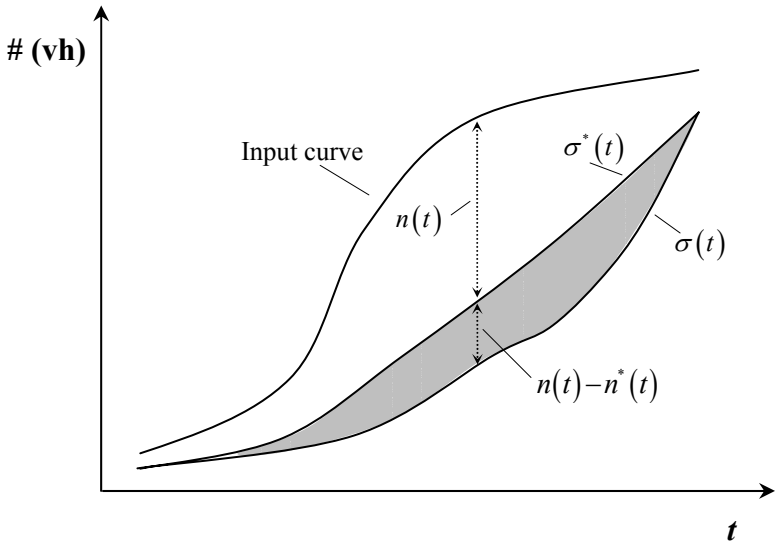


Figure 4.4: A queueing diagram for reservoir *R*

4.2.2 A macroscopic simulation of cruising-for-parking

To illustrate numerically the model described in the previous section and see if it produces logical results, we now present a cruising-for-parking simulation for different supply of on-street parking spots. The simulated network is a macroscopic representation of the geometric and traffic characteristics of the San Francisco micro-simulated site (see Chapter 2).[15] First, a production function $P(n)$ was estimated as the best fit 4th degree polynomials in all the (P, n) pairs of figure 2.7a. Next, time-series of total demand rates were calculated for "run 4" of figures 2.4 and 2.5. This run did not create severe congestion in the system (see figure 2.10). Also, averages for the fraction of demand rates with internal or external destinations were calculated across the whole run. On-street

[15] In this site, all the internal trips were completed in garaged parking areas and cruising was absent.

parking spots were evenly distributed in both sides of all streets, without restricting the lanes used for traffic movements. Figure 4.5 summarizes the values of the most important parameters used in this simulation.

parameter	value	description
L	56.25	total steet length (km)
f_{11}	0.1	
f_{12}	0.2	f_{ij} = fraction of demand generated from i with destination j,
f_{21}	0.4	i, j=1 (internal) or 2 (external)
f_{22}	0.3	
lp'	1.743	average trip length without cruising (km)
N_p	[4750, ∞)	total number of on-street parking spots
n^p	1500	vehicles parked at t=0
l_1	$2\times L/N_p$	average distance traveled between two adjacent spots

Figure 4.5: Description of main parameters of the simulation

Traffic was simulated for a period of 4.5 hours (10000 time units) for $N_p \in [5000, \infty)$ using a discrete version of the dynamic equations (4.10)-(4.13). All the trips generated from the internal region start from the parking sub-reservoir R_p. Also, in all runs the system returns in a state of low traffic where cruising-for-parking is negligible. Thus, the total input and output of vehicles is the same for all runs and the results are comparable. Figure 4.6 presents the results of the simulations for different values of N_p. First, there is no cruising effect for $N_p > 10000$.

Figures 4.6a and 4.6b show time-series of accumulation and fraction of available parking spots. Note that for values of N_p smaller than 6000 the network becomes severely congested with states in regime III of the MFD, while for N_p>7500 remains in regime I. Note as well that, during the same time periods, the fraction of available parking spots is close to zero (<5%) for small N_p, as shown in figure 4.6c. This indicates that there are

cases where cruising-for-parking can lead uncongested systems, without very high traffic demand, to severely congested states.

Figures 4.6b and 4.6d show time-series of the number of vehicles cruising-for-parking and of cumulative output, i.e. total trip completion rate, expressed as an oblique plot. Note that when p is close to zero, the fraction of vehicles cruising-for-parking is significant (~30% for N_p=5000), which is a value observed in many cities around the world(Shoup, 2006), e.g. in Massachusetts, even 20 years ago (O' Malley, 1985). The area under the graph of figure 4.6b is the total delay for vehicles cruising for parking (family ***s***), d_s. A study on West Portal Avenue on the city of San Francisco, conducted by San Francisco State University, found that average cruising time to find a curb space was 3.2 minutes (Shoup, 2005b). This cruising time is the same as in the simulation for N_p=6000. But, this delay is not the only cost caused by limited availability in parking.

The total delay d_p due to limited parking for a value of N_p, is the area (see figure 4.6d) between the graphs (i) for the given N_p and (ii) for $N_p \to \infty$ (for N_p>10000 the delay is zero). There are cases where d_p is 2 or 3 times higher than d_s, which is something neglected in most of the existing studies. Figure 4.7 shows d_p, d_s and d_p-d_s for the whole period of the simulation for different values of N_p. Note that the effect for non-cruising vehicles, as this expressed by d_p-d_s, i.e. the total delay for all vehicles excluding the time cruising-for parking, becomes more intense than the effect for cruising vehicles (family ***s***) as N_p decreases--the elasticity of their ratio is about -3% per 100 parking spots for $N_p \in$[5500, 7000]. Note also that accumulation $n_i(t)$ is a convex function of N_p. For a given value of N_p <10000, $n_i(t)$ is higher than the sum of (i) $n_i\left(t\right)$ for $N_p \to \infty$ and (ii) $n_i^s\left(t\right)$ for the given value of N_p.

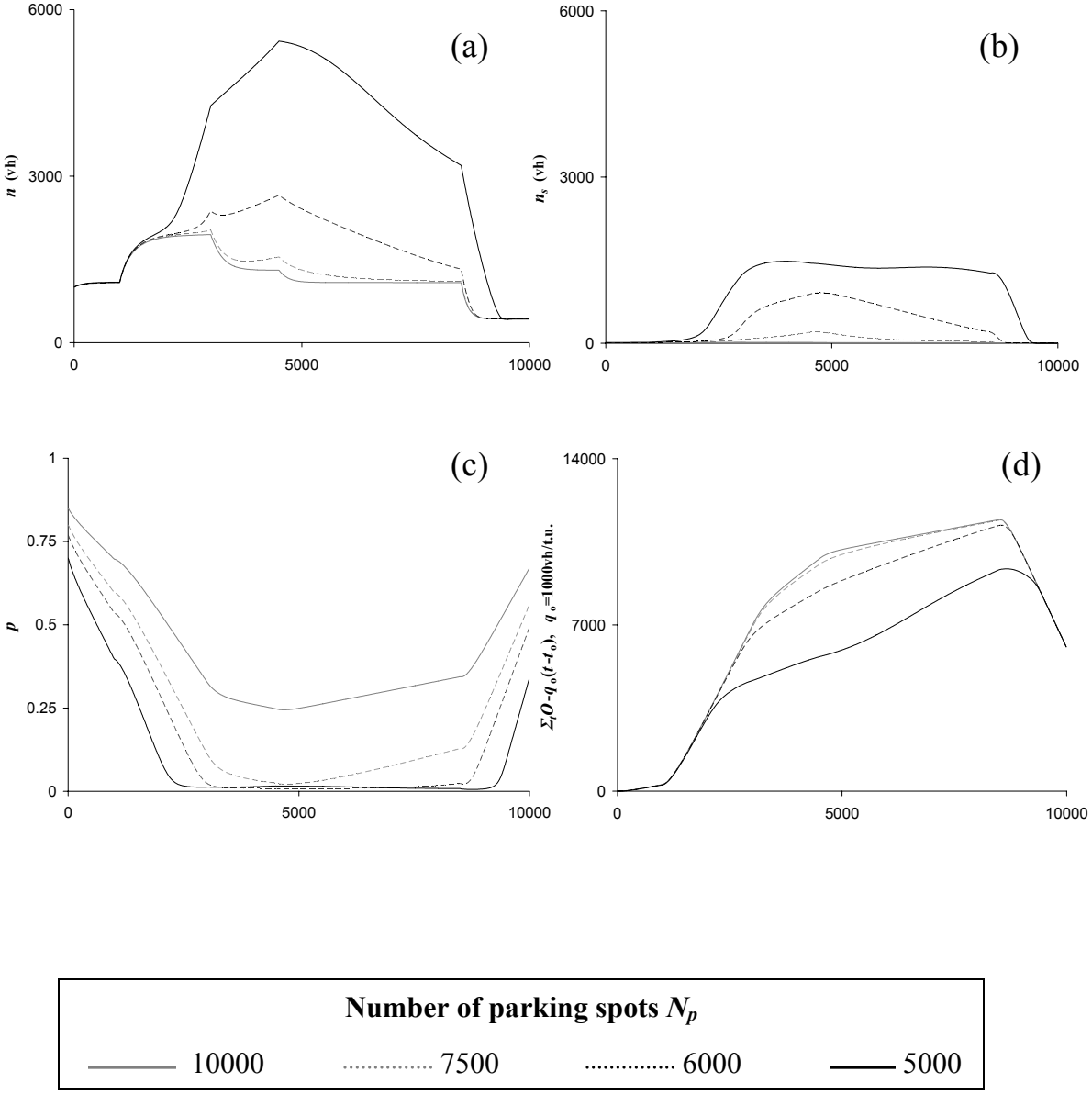


Figure 4.6: A simulation of searching-for-parking; time series of (a) accumulation; (b) vehicles searching for parking (c) % of available parking spots; (d) cumulative output

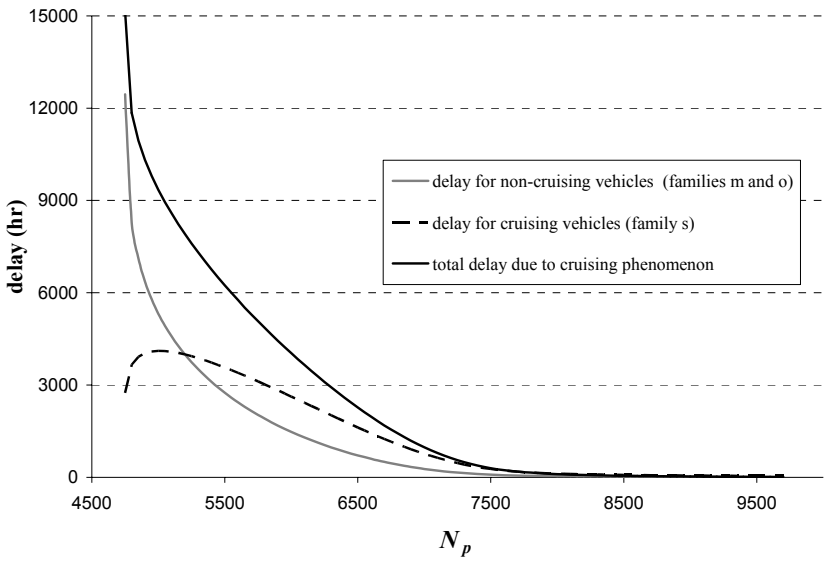


Figure 4.7: Delays d_p, d_s and d_p-d_s for different values of N_p

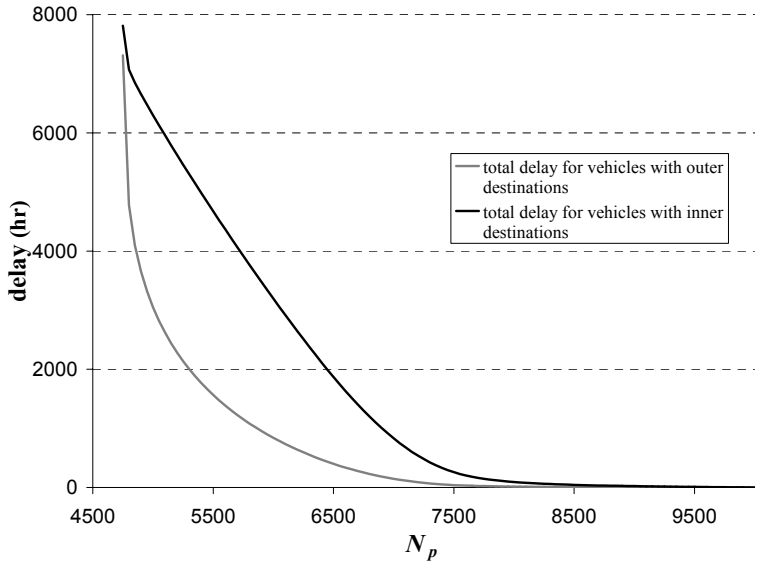


Figure 4.8: Total delay for vehicles with inner and outer destinations for different values of N_p

Figure 4.8 shows the total delay for vehicles with inner (including cruising) and outer (cruising is zero for these vehicles) destinations. Note that as the number of total spots decreases, the effect for vehicles with outer destinations becomes significant. Also, the marginal cost of an additional user with external destination is smaller than that of a

user who will cruise for parking; and that an additional user with internal destination, causes more delays in the non-cruising vehicles, than the cruising ones. This suggests that a pricing scheme (e.g. pricing on-street parking) to increase mobility in cities should consider all different types of costs for the users of the system.

Most of the existing studies evaluate the effect of the cruising-for-parking phenomenon by the average time cruising for parking. One should expect that for a given demand, average cruising time will increase as parking limitations become intense, i.e. N_p decreases. But, this is not necessarily true, especially when traffic is congested. A "suprising" observation is that (i) d_s becomes concave for values of N_p<5500, and (ii) d_s *decreases with* N_p for N_p<5000 (e.g. average cruising time is for 4.8min N_p=5200 and 4.3min for N_p=4800), as figure 4.7 shows. Thus, cruising delay d_s, is bounded but total delay is not. The reason for this is that the system reaches quickly regime III of the MFD and the output O_m, which is the input to R_s (sub-reservoir of cruising vehicles) is very small; at the same time, average speed is very low, too (<10km/hr). Thus, vehicles do not reach their internal destination *because* they are "stuck" in congestion and the number of vehicles cruising-for-parking is small.[16]

A thorough analysis of the simulated data sheds more light in the previous observation. Figure 4.9 shows the trajectory of (O_m, O_s) pairs, i.e. the input and output of the cruising sub-reservoir R_s, during two different runs; for (i) N_p=5200 (figure 4.9a) and (ii) N_p=4800 (figure 4.9b). In the beginning of the simulation R_s is almost empty, vehicles find parking almost immediately and $O_m \approx O_s$. After a while the parking sub-reservoir R_p loads with vehicles (p<0.3), vehicles spend more time in cruising and $O_m > O_s$. This results

[16] Results for values of N_p<4750 are not presenting because the system reaches gridlock and all the delays approach infinity.

an increase in accumulation, and the system reaches regime III of an MFD, with even smaller completion rates because of longer trips. This effect is more severe for smaller values of N_p and it can lead the system in states where $O_m<O_s$, i.e. number of vehicles cruising-for-parking decreases. Note that for N_p=4800 this occurs at time t=3650 and lasts until t=6800 (a period of ~1.5hr). For N_p=5200 this occurs later, at t=4800, and this effect is mild because (i) it lasts less; (ii) accumulation is smaller and (iii) demand decreases for t>5000. Thus, number of cruising vehicles and average cruising time are smaller in case of N_p=4800, but total delay d_p is higher, as expected.

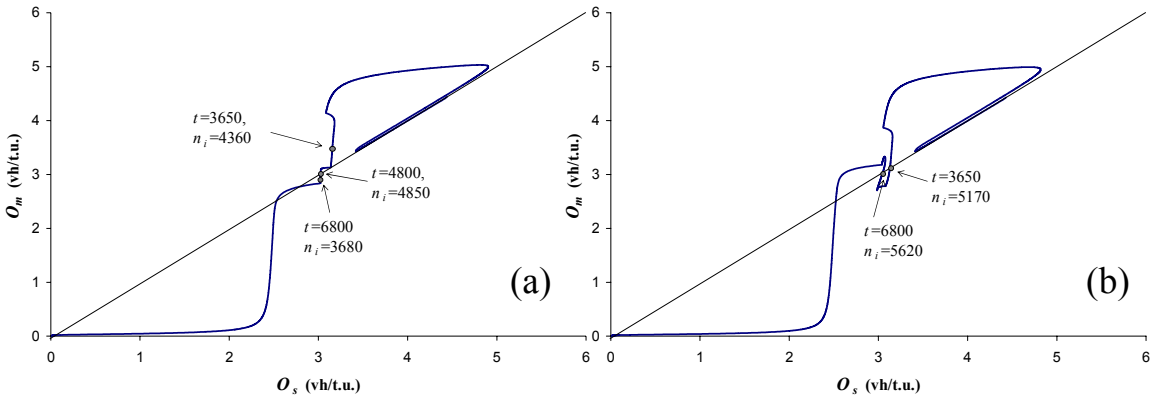


Figure 4.9: Trajectory of (O_m, O_s) for (a) N_p=5200; (b) N_p=4800

4.3 Final Remarks

This chapter has described the dynamics of the rush hour for multi-neighborhood cities, where congestion is not evenly distributed across a city. This analysis is an intermediate, but necessary, step in developing perimeter control strategies, which can increase the mobility of the system, without requiring the knowledge of detailed information about O-D tables. These are presented in Chapter 5.

Also, this chapter modeled in a parsimonious way the cruising-for-parking phenomenon and showed the multiple effects in all the drivers, even those with destinations outside the cruising region. These results can be misinterpreted. We are not advocating increasing parking to meet the demand for travel. This can be a devastating decision because of (i) the tremendous cost to keep pace with increase in travel demand and (ii) the phenomenon of induced demand. For example, the city of San Francisco has around 450,000 registered vehicles, and this increases by about 35,000 vehicles during each workday (Department of Traffic and Parking, 2007). But, dwellers or San Francisco in 2007 elections, voted against a proposition, which would prevent all reductions in parking spaces, allow developers to provide more parking spaces in downtown, and set minimum parking requirements for neighborhoods.

Only if we decide to limit parking (e.g. with pricing), we can limit the demand of individual cars and have positive consequences for more sustainable and less congested cities. In this chapter we have developed all the necessary tools to estimate the total costs for a city due to cruising phenomenon, given the total number of parking spots and the aggregated demand. We can then estimate the percentage of demand decrease needed, to eliminate this effect and improve the mobility of the system. In case of high demand for internal destinations, pricing only these types of users (e.g. by pricing curb parking) can help. But, in case of high demand for external (with regards to the cruising region) destinations, toll pricing can be more efficient.

Chapter 5

Perimeter Control for multi-neighborhood cities

THIS chapter provides all the necessary tools to apply perimeter control strategies, which do not rely on forecasting, for multi-neighborhood cities. Control decisions are based on monitoring of the system, which replaces prediction. The system is repeatedly modified based on observations. The logic is that the flow towards an overloaded area of a city is restricted while the flow towards an underutilized area is promoted. Prevention of overcrowding on a city centre is succeeded by metering of access to maintain the mobility of cars at a stabilized level. For example, longer red times can be applied in the periphery of the centre during peak hours for phases which direct vehicles to the centre. Section 5.1 shows mathematically how perimeter control can increase mobility and system efficiency, expressed by the rate vehicles reach their destinations. Section 5.2 presents some applications of perimeter control in simulated cities with different structures (single or multi reservoir, with or without cruising for parking phenomenon).

5.1 Perimeter control for two-reservoir cities

Daganzo (2007) proved that, for single reservoir systems, the policy that minimizes total delay and maximizes the output from the system, is that which does not allow the system to reach states in regime III of the Macroscopic Fundamental Diagram (MFD). This policy distributes benefits widely as all car trips end sooner. Analysis also revealed the there are fewer vehicle-hours and higher speeds at all times. A simple way to apply this strategy is to monitor the system and when accumulation passes a chosen critical value in regime II, to restrict input flow.

While this strategy is Pareto efficient for single reservoir systems, it ignores the delays for users the periphery that do not want to travel to the center. This is not a problem in case most of the trips have destinations inside the reservoir, because this policy still gives to every user the most advanced departure time possible (assuming a first-in first-out system). But, when there is a significant amount of destinations outside the reservoir, this policy may not be the most efficient (in terms of trip completion rate).

We now show under what conditions perimeter control can be beneficial for-two reservoir cities and increase the rate at which vehicles reach their destinations. Generally speaking, this is the case when some parts of a city are more congested than others and by controlling the intertransfers the overall mobility of the system can improve significantly.

Consider a two-reservoir system (i=1, 2 and i'=3-i), where (i) output O_i is function of accumulation n_i and (ii) the state of this system is expressed in the dynamic case by equations 4.5 and 4.6. We assume that the output function, $O_i(n_i)$, is non-negative and concave with $O_i(0)$=0 and has 3 different regimes: (regime I) $O_i(n_i)$ is monotonically

increasing for $n_i \le n_i^-$; (regime II) $O_i(n_i)$ exhibits constant maxima, at value γ_i, for $n_i^- \le n_i \le n_i^+$; (regime III) $O_i(n_i)$ is monotonically decreasing for $n_i^+ \le n_i$. We consider 6 different cases for values of accumulations in each of the 3 regimes for each of the reservoirs. Cases 1-3 deal with uncongested cities, while cases 4-6 with congested ones.

The dynamic equations for the system, after some manipulations in equations 4.5 and 4.6 and ignoring the term of inflow capacity, C_{ij}, (see section 4.1 for an explanation), can be expressed as:

$$\begin{aligned} n_i^{t+1} = n_i^t + q_i^t - \min\left(x_{i'}^t, \phi_{ii'}^t \cdot O_i\left(n_i^t\right)\right) - \phi_{ii}^t O_i\left(n_i^t\right) + \\ + \min\left(x_i^t, \phi_{i'i}^t \cdot O_{i'}\left(n_{i'}^t\right)\right) \quad , \quad i = 1,\ 2 \text{ and } i' = 3 - i \end{aligned} \tag{5.1}$$

$$\begin{aligned} \phi_{ii}^{t+1} n_i^{t+1} = \phi_{ii}^t n_i^t + f_{ii}^t q_i^t - \phi_{ii}^t O_i\left(n_i^t\right) + \\ + \min\left(x_i^t, \phi_{i'i}^t \cdot O_{i'}\left(n_{i'}^t\right)\right), \quad i = 1,\ 2 \text{ and } i' = 3 - i \end{aligned} \tag{5.2}$$

where n_i^t is the accumulation at reservoir i, at time t; ϕ_{ii}^t is the fraction of vehicles in i with destination in i, at time t; x_i^t is a control variable for movements directing to i, at time t; f_{ii}^t is the fraction of trips generated at i with destination i, at time t; q_i^t is the generated flow at i, at time t.

We are interested in maximizing the total output of the system at time t+1, given the state of the system at time t, $o_{t+1} = \phi_{11}^{t+1} \cdot O_1\left(n_1^{t+1}\right) + \phi_{22}^{t+1} \cdot O_2\left(n_2^{t+1}\right)$, by controlling movements in the boundary of the two reservoirs. Figure 5.1 summarizes the perimeter control strategy (values of x_1^t , x_2^t), where quantities A and B are given by equations (5.3) and (5.4). Appendix A provides a proof that this strategy will maximize output.

R_1 / R_2	**I**	**II**	**III**
I	$x_1^t \geq \phi_{21}^t \cdot O_2\left(n_2^t\right)$ $x_2^t \geq \phi_{12}^t \cdot O_1\left(n_1^t\right)$		
II	$x_1^t \geq \phi_{21}^t \cdot O_2\left(n_2^t\right)$ $x_2^t \geq \phi_{12}^t \cdot O_1\left(n_1^t\right)$	$x_1^t \geq \phi_{21}^t \cdot O_2\left(n_2^t\right)$ $x_2^t \geq \phi_{12}^t \cdot O_1\left(n_1^t\right)$	
III	$x_1^t \geq \phi_{21}^t \cdot O_2\left(n_2^t\right)$ $x_2^t = \begin{cases} 0 & \text{, if } B \geq 0 \\ \phi_{12}^t \cdot O_1\left(n_1^t\right) & \text{, o.w.} \end{cases}$	$x_1^t \geq \phi_{21}^t \cdot O_2\left(n_2^t\right)$ $x_2^t = \begin{cases} 0 & \text{, if } A \geq 0 \\ \phi_{12}^t \cdot O_1\left(n_1^t\right) & \text{, o.w.} \end{cases}$	$x_1^t = \begin{cases} 0 & \text{, if } B \leq 0 \\ \phi_{21}^t \cdot O_2\left(n_2^t\right) & \text{, o.w.} \end{cases}$ $x_2^t = \begin{cases} 0 & \text{, if } B > 0 \\ \phi_{12}^t \cdot O_1\left(n_1^t\right) & \text{, o.w.} \end{cases}$

Figure 5.1: Perimeter control strategy for two-reservoir systems

$$A = \phi_{22} \cdot \frac{dO_2\left(n_2\right)}{dn_2} - \left(1-\phi_{22}\right)\frac{O_2\left(n_2\right)}{n_2} - \frac{\phi_{11} \cdot \gamma_1}{n_1} \tag{5.3}$$

$$B = -\phi_{22}\frac{dO_2\left(n_2\right)}{dn_2} - \left(1-\phi_{22}\right)\frac{O_2\left(n_2\right)}{n_2} + \phi_{11}\frac{dO_1\left(n_1\right)}{dn_1} - \phi_{11}\frac{O_1\left(n_1\right)}{n_1} \tag{5.4}$$

Qualitatively, this is an explanation of the strategy: (i) if none of the reservoirs is in regime III, you do not restrict any of the movements in both directions; (ii) if only one reservoir is in regime III, then you restrict the input to that, only if many trips have destinations inside this reservoir and (iii) if both reservoirs are in regime III, you restrict the input to the one with higher density of destinations.

There is a caveat in the model of the two-reservoir system. If we control the input by restricting the perimeter capacity, we will change the spatial distribution of vehicles on the sending reservoir. This may invalidate the homogeneity assumption of traffic loads within a reservoir (if the reservoir is too big) and reduce the accuracy of the model. In

that case, accuracy can be improved by using smaller reservoirs, i.e. an "onion structure" with 3 or 4 nested reservoirs.

5.2 Applications

5.2.1 A single-reservoir city

A simulated test of "a single-reservoir" control strategy (Daganzo, 2007), is now presented for the San Francisco network, described in Chapter 2. We considered the whole network as a single reservoir system and we restricted the input from the boundary of the reservoir, by changing the signal settings in the periphery, to keep accumulation as close to the critical value. We applied perimeter control for a run which, without control, created congested conditions and led the system to gridlock. Our limitation was that the demand was changed a priori, without monitoring the system, as the simulation package is not interactive. A test under this constraint is strict, since feedback information to optimize control is not possible. Figure 5.2 presents the results of this test.

The cumulative demand with control was lower than the one without control for every time t. By trying to keep the accumulation in regime II the total output of the system increased by 34% (60347 trips instead of 45083 in a 4 hour period). The plot for the original uncontrolled case shows that the rate of trips ending increases at first as the accumulation rises and then drops about half way through the simulation, as accumulation surpasses the sweet-spot and traffic congestion prevents vehicles from reaching their destinations. In the signal-controlled case, the higher rate of trips ending is maintained by preventing some vehicles from entering the reservoir when its accumulation is at the sweet-spot. This allows more vehicles to reach their destination in

the same amount of time. So, as claimed, the restriction improves everyone's accessibility for the given demand. Note that this is done with fewer vehicles on the road.

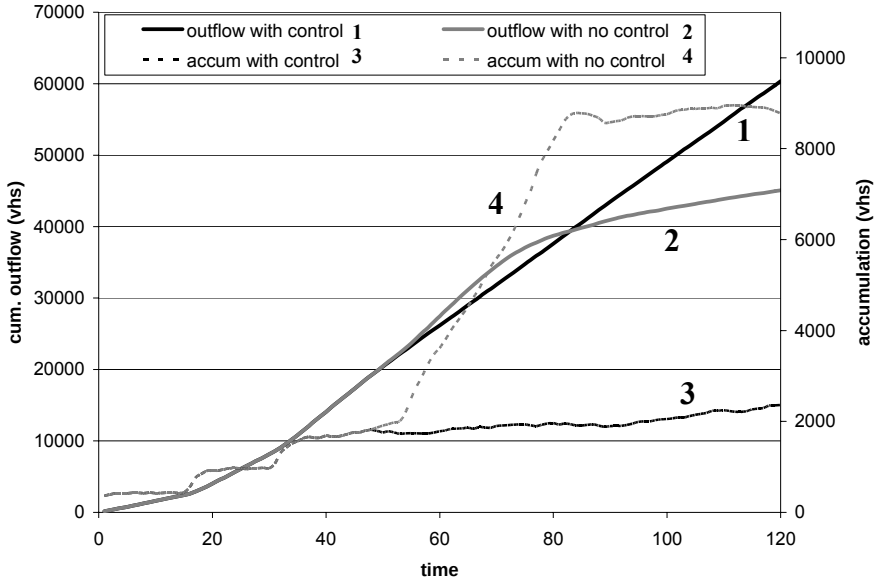


Figure 5.2: Time series of cumulative output (left axis) and accumulation (right axis) for the San Francisco network with and without perimeter control

Pre-timed signals controlled the reservoir to maintain a higher output, but signals work best when timed to a known demand. If the system is monitored, the control can be varied in response to real changes in accumulation, and the system could be designed more efficiently to maintain maximum output by keeping accumulation in the sweet-spot.

5.2.2 A two-reservoir city

We now simulate a city with two nested reservoirs during the morning commute where the majority of people are moving towards the centre of the city. We assume that the dynamic behavior of the system is governed by equations 5.1 and 5.2 and we apply

perimeter control strategies. We also assume that there are infinite available parking spots and vehicles do not cruise for parking.

The internal reservoir (R_1) is ten times smaller than the external reservoir (R_2). Vehicles are uniformly distributed in the city and at time zero a few vehicles are moving and the rest are parked. Initially, vehicles enter the system at a constant rate; demand decreases with time after 350 time units (total duration of the simulation is 1000 time units). A trip generated from the interval reservoir (R_1), will finish in R_1 with probability f_{11}=0.7 and in the external reservoir (R_2) with probability f_{12}=0.3, while a trip generated from R_2 has equal probability to finish in R_1 or R_2 (f_{21}=f_{22}=0.5). A three-regime output vs. accumulation function $O_i(n_i)$ is assumed for each of the reservoirs, as a scaled up version of the output-accumulation function $O(n)$, of the San Francisco simulated network (figure 2.10a). $O(n)$ was estimated for regime I and III as the best fit 4^{th} degree polynomial for the (output, accumulation) pairs and as the best fit zero degree polynomial for accumulation in the range [2500,3500]. Figure 5.3 illustrates this result. Critical values of the parameters are assumed to be proportional to the size of the reservoirs, i.e. $O_2(n_2)$=10×$O_1(10n_1)$= 25×$O(25n)$. Also, the same procedure was applied for the inflow capacity function $C_i(n_i)$, using data from Chapter 2.

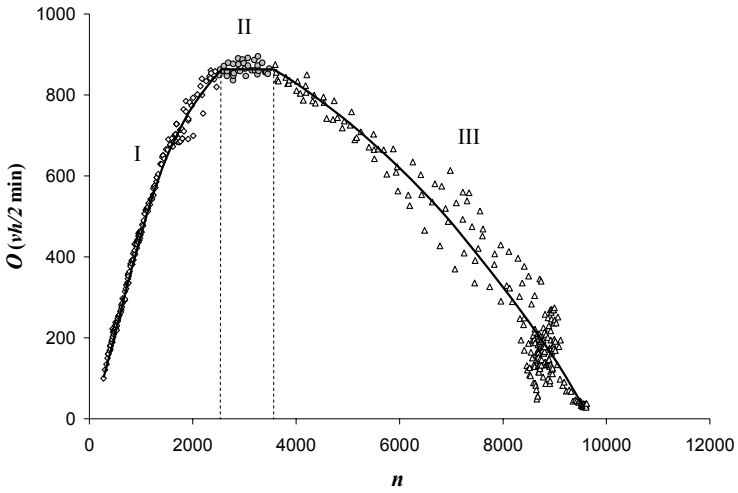


Figure 5.3: 3-regime MFD based on San Francisco network data

First, we let the simulation run without applying any control in the boundary of the two reservoirs. We observe (figures 5.3b and 5.3d) that after some time the internal reservoir R_1 is subject to gridlock and output decreases to zero.

We repeat the simulation for the same demand pattern, but we now apply the control strategy analyzed in section 5.1. The philosophy of the control is that we try to keep accumulation in the internal reservoir in a range that maximizes output. When accumulation is higher than the critical value then inflow capacity in the boundary is restricted. This in practice is feasible by increasing the red phases for movements directing in the centre.

The results shown in figures 5.3a and 5.3c are encouraging and show that perimeter control can provide significant benefits to all the users of a city. The number of vehicles finished their trips is higher, for both the internal and the external reservoir, and average speeds are much higher when control is applied.

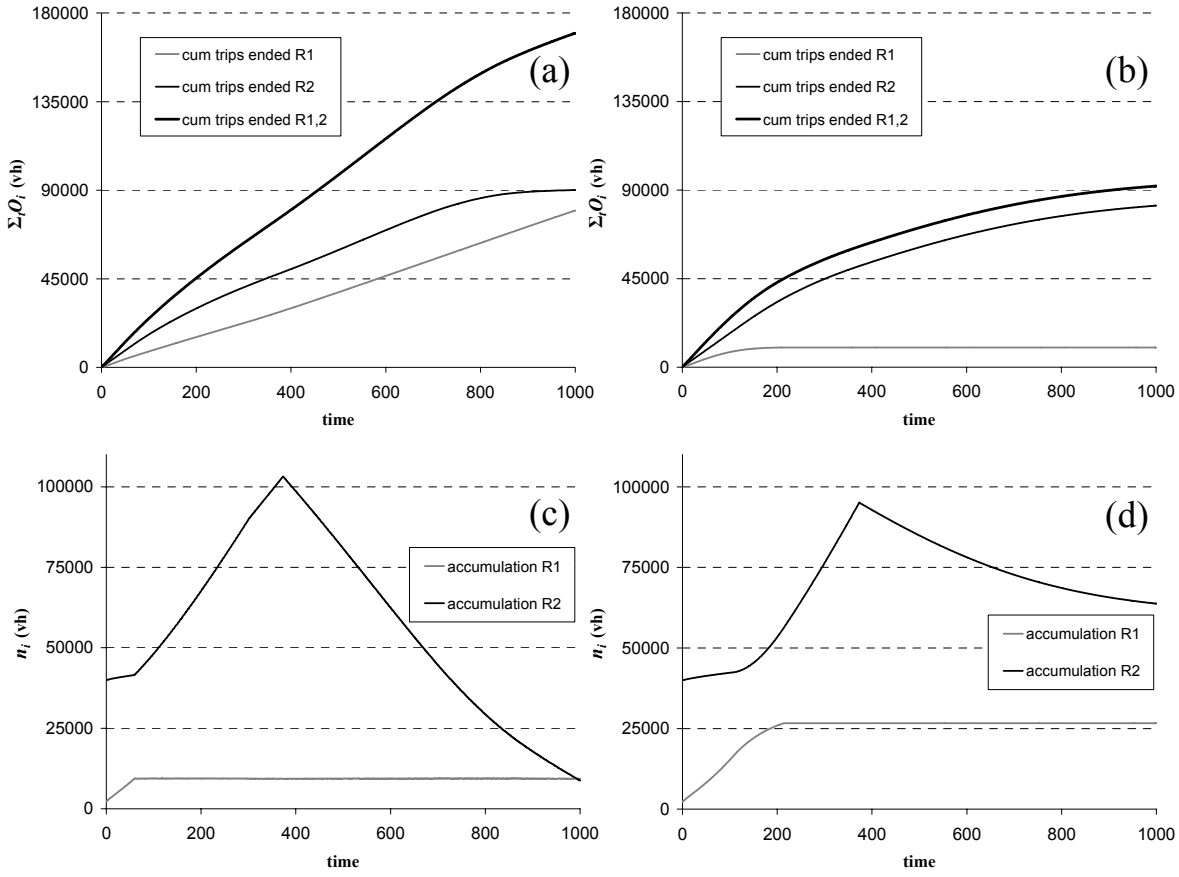


Figure 5.4: A two-reservoir system simulation: Time-series of (a) cumulative output with control; (b) cumulative output without control; (c) accumulation with control; (d) accumulation without control.

Furthermore, the proposed control strategy is quite parsimonious. Even if the relationship between production or output and accumulation is not accurately known, the critical values of accumulation (or average speed) are needed for the control. Even the critical value does not have to be precisely known as it can be estimated by tweaking the controls over days.

5.2.3 Cruising-for-parking

The previous sections showed that perimeter control can be beneficial for multi-neighborhood cities. A main characteristic of this strategy is that it treats equally all the users that are metered, independently of their destination. A fair question to ask is if perimeter control can improve mobility when "metered" users have different characteristics, e.g. when the cruising-for-parking phenomenon is intense. In that case, as shown in subsection 4.2.3, people with destinations inside the inner reservoir inconvenience the system more than users driving through the inner reservoir, with outer destinations.

To get a better understanding, we now test how traffic conditions change when perimeter control is applied while vehicles are cruising-for-parking. For comparison purposes, we simulate the same network used to analyze the cruising-for-parking phenomenon (section 4.2.3). For different values of the total number of parking spots N_p we run the simulation with exactly the same traffic and demand characteristics. The only difference is that we apply two different types of perimeter control policies: (Strategy I) a simple "single-reservoir strategy" (Daganzo, 2007), where when accumulation enters in regime III of the MFD, we meter the inputs from the periphery; and (Strategy II) a more strict strategy where we restrict input flows either when accumulation is in regime III or when accumulation is in regime II and parking availability is very limited (p<0.2). The second strategy aims to moderate the cruising-for-parking phenomenon. We also estimate the additional vehicle hours of delay experienced by the metered vehicles, for both strategies. For simplicity, we assume that the outer region is completely uncongested and

the only delay because of the perimeter control is the one while waiting for entering the inner region. Figure 5.5 summarizes the results.

Figure 5.5a shows the total delay due to cruising-for-parking, d_{tot}, (including the delay of the metered vehicles) and the delay while searching for a spot, d_s, for the two different strategies. We see that in all the cases the more strict "strategy II" moderates some of the searching time, but the penalty experienced by metered vehicles is much higher, i.e. strategy II is always worse that strategy I. The reason is that, while the perimeter control meters the input from the outer region, it cannot restrict the input from the inner region. Thus, the parking sub-reservoir still saturates quickly due to high demand of inner trips and perimeter control makes the overall situation worse. Let us now compare strategy I with the original case of no control, presented in subsection 4.2.3.

Figure 5.5b shows different types of delays for control strategy I and the original case for different values of N_p. First note that perimeter control has minor effect in the delay of vehicles cruising-for-parking, d_s. Note also that the delay for all the vehicles of the inner region, d_p, (excluding metered vehicles) decreases significantly (e.g. more than 30% for N_p<5200). But, there is major penalty for the metered vehicles, which is increasing when the original system becomes heavily congested for smaller values of N_p (accumulations in regime III). Thus, the overall delay decreases 10-15% and only when the inner region reaches states in regime III.[17]

[17] As mentioned in subsection 4.2.3, for the given demand when N_p is high and cruising negligible, the system remains uncongested at all times.

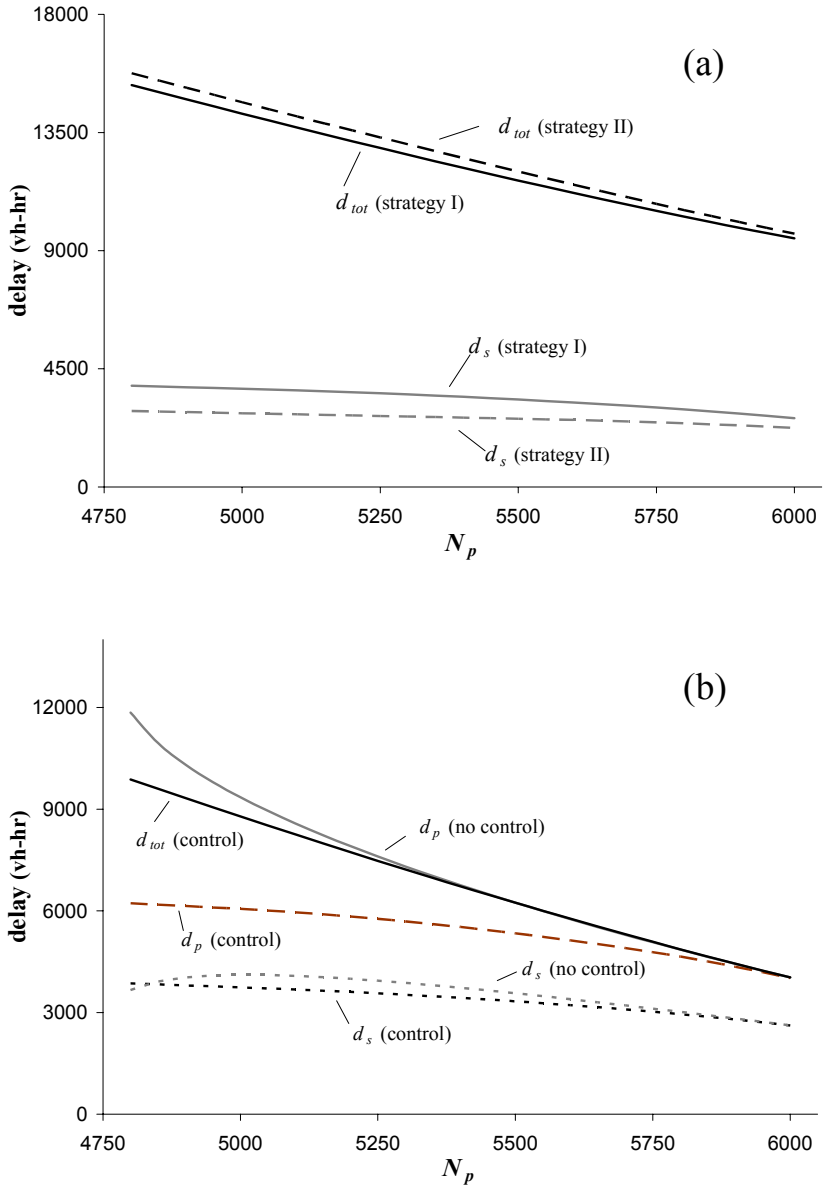


Figure 5.5: Simulation results for different values of N_p: (a) d_{tot} and d_s for two different strategies; (b) comparison of control strategy I and no control

These results show that perimeter control cannot have a significant positive effect (i) when many trips are originated from the inner zone and (ii) when the cruising-for-parking phenomenon is intense. In these cases, it is necessary to decrease the demand of vehicles

on the road. Congestion pricing or parking enforcement are alternative methods that can lead to better results, in terms of mobility and accessibility for the system.

5.3 Final Remarks

This chapter developed perimeter control policies which can improve significantly the mobility of two-reservoir systems in many cases. These ideas can be extended to multi-neighborhood cities (with $N>2$). The mathematical analysis for $N>2$ is tedious in the general case, but similar results to the two-reservoir case are expected, when regions in regime III are not adjacent. An effective policy would be to control the inter-transfers between the reservoirs and keep accumulation in levels, for which output is maximized, for areas *with high density of destinations*. Equity issues should also be considered because these strategies may not be Pareto efficient, e.g. when cruising-for-parking is intense. It is possible that this control favors people live in the centre of a city to people in the periphery. Thus, more opportunities should be given for these people, which usually belong to the poorer community groups. These opportunities could include attractable alternative modes of transport, congestion pricing for vehicles, bus lanes etc.

Chapter 6

Conclusions

This final chapter summarizes the results in this dissertation and proposes ideas for future research.

6.1 Summary

In this dissertation, we have analyzed how we can develop physically realistic models of urban congestion in cities and how, based on these models, we can develop control strategies to improve mobility.

We argued that the traditional disaggregated models provide only limited answers to these questions because: (i) they require extensive data that are unavailable, particularly in the dynamic case; and (ii) the laws of behavior cannot be verified, and these models cannot be expected to produce reliable outputs. We also argued that the existing aggregated models (i) fall short in describing the rush hour dynamically, and (ii) their behavior for different demand profiles has not been tested.

As a remedy, we proposed a general analysis framework that for the first time incorporates (i) realistic traffic behavior with variable inputs and outputs that could describe a rush hour dynamically and (ii) control decisions that are based on observations and not on forecasting. This framework builds on the gridlock model in Daganzo (2007).

Finding from chapters 2 and 3 (compare figures 2.6 and 3.3 with 2.7a and 3.5a) show that large systems observed over long times are predicted more reliably than small systems; "bigger is better". This suggests that large scale (the nemesis of traditional models) actually works in favor of the aggregate approach and that we can shift the modeling emphasis from microscopic predictions to macroscopic monitoring and control based on accumulation or average speed. The system performance in terms of vehicle hours, vehicle kilometers, trip completion rates or average speeds can be reliably modeled from the vehicle accumulations in a neighborhood of a city, of a size comparable with a trip length.

This dissertation work showed that a macroscopic fundamental diagram linking production (the product of average flow and network length), accumulation (the product of average density and network length) and speed exists for neighborhoods of cities in the order of 5-10km^2. This happens even as the origin-destination table changes. The tests also revealed an invariant relation between the space-mean flows on the whole network, which are easy to estimate given the existence of an MFD, and the trip completion rates, which dynamically measure accessibility and are difficult to observe. No evidence of hysteresis in the MFD was noticed. In other words, average flow and density pairs during growth and decay of demand followed the same MFD. In addition, the findings of section 3.3 are important because: (i) they establish that the MFD is a property of the network

and not of the detectors' locations; and (ii) they show how to estimate accumulation and speed for the whole network. This means that many cities with this widespread supporting infrastructure can readily monitor their state and benefit from control strategies, similar to these presented in chapter 5.

By exploiting the insights and the properties of the MFD, we were able to describe the rush hour dynamically in case of multi-region cities that are not uniformly congested everywhere (Chapter 4). Also, in section 4.2 we developed a cruising-for-parking model. This model filled three important gaps: (i) it described the physics of overcrowding caused by the phenomenon; (ii) it showed that cruising-for-parking affects all the users of the system, even those with destinations outside the "limited parking region" and (iii) it provided tools to estimate the direct costs of all users, as these expressed by additional vehicle-hours traveled. Also, it differed from the existing economic models because: (i) it can describe the rush hour dynamically while vehicles search for a spot (properly recognizing that delays are greatest when flows are lowest, unlike existing economic models) and (ii) it contains variables that are readily observable or can be easily estimated with field experiments (unlike the demand curve in most economic models).

From a methodological point of view, the models of chapter 4 provided us with all the necessary tools to incorporate perimeter control strategies in the general framework. Findings from chapter 5 are encouraging because they show that perimeter control can provide significant benefits to all the users of a city, and that detailed origin-destination tables are not needed for these strategies. Qualitatively, the perimeter control strategy restricts the inputs to congested regions of a city which also have high densities of destinations. Also, this strategy is quite parsimonious. Even if the MFD is not accurately

known, critical values of accumulation (or average speed) are needed for the control. Even the critical values do not have to be precisely known as they can be estimated by tweaking the controls over days. Also, results from section 5.5 show that perimeter control cannot have a significant positive effect (i) when many trips originate from the congested zone and (ii) when the cruising-for-parking phenomenon is intense. In these cases, it is necessary to decrease the demand of vehicles on the road using other methods (e.g. congestion pricing or parking enforcement).

6.2 Future work

This dissertation is not the final word on the analysis of macroscopic modeling of traffic in cities. Several research areas related to this thesis deserve further investigation.

(a) *Macroscopic modeling of multimodal, multi-reservoir systems*: The multi-reservoir, single-mode models, developed in chapter 4, can also be extended to multimodal, multi-reservoir transportation systems. This can be done by treating the system as an interconnected network of reservoirs, where each reservoir represents the streets in a neighborhood. The different modes can share the same road space, as when buses and cars on city streets or general purpose freeway lanes, or the modes can be separated by dedicating road space for some of modes. In this extension, different parts of a city can implement different management strategies. Perhaps bus-only streets are allocated only in the central business district while other parts of the city allow vehicles to operate in mixed traffic. The effect of changes in one reservoir on the behavior of adjacent reservoirs need also to be considered with this model. The basic building block of this theory is the multi-reservoir serving one mode (chapter 4).

A question that needs an answer is if an MFD exists for transit systems or mixed systems with transit and cars. The field experiment in chapter 3 sheds some light on car systems, but further analysis is necessary in understanding the physics of multimodal systems.

(b) *Congestion pricing control strategies revisited:* With reliable estimates of real-time accumulations, one could devise many different and effective policies to keep accumulations in their sweet-spots so as to enhance mobility and accessibility. Perimeter control by modifying traffic signals, as was done in chapter 5, is one method of achieving this without affecting the number of trips per mode. Strategies could also be used to change the number of trips per mode, by encouraging the use of more sustainable modes, such as shifting trips from cars to buses. Chapter 4 provided all the necessary tools to estimate different types of costs to the system and to separate users. Also, chapters 2 and 3 emphasized what is the number of cars that a city can afford without experiencing heavy congestion. Parking or peak hour tolls are pricing strategies that can do this. But, most of the existing congestion pricing models assume a supply curve where speed is a decreasing function of flow, which is a wrong assumption according to the findings of this dissertation. The correct physics must involve a non-monotonic supply curve, where effects manifest themselves in a time-dependent fashion. Now, by knowing (i) an experimentally verified supply curve (chapter 3) and (ii) estimates of the total delay to the system in a time-dependent model (chapter 4), congestion pricing strategies can be redefined and produce more effective results. But, it should be recognized that under these strategies, different

modes may be competing for scarce resources. Thus, there is another possible avenue of future research.

(c) *Allocation of urban space between different modes*: To avoid "lose-lose" effects, significant shifts in mode share should be accompanied by shifts in how road space is allocated between modes. Macroscopic modeling can do more than design policies in neighborhoods served by single modes. The models may also support the design of multimodal policies. For example, a fixed number of lanes can be reserved everywhere in a network for more sustainable modes (e.g. transit), or this amount can vary from place to place (e.g. suburbs vs. city center). The number of lanes reserved can be the same at all times or vary with the time of day. These spatiotemporal decisions allow for customized and targeted control strategies, e.g. a combination of congestion charging in the center of the city (e.g. the London example) with dedicated bus lanes (e.g. the Athens example) can create an efficient and equitable system. But if space is allocated incorrectly, system capacity could be wasted. Then, the problem becomes finding the careful balance that gives priority to more sustainable modes without creating the kind of waste that does damage. [18] Figure 6.1 summarizes the proposed future work in the development of control strategies, which do not rely on forecasting, but in monitoring of the system.

[18] As shown in Daganzo and Cassidy (2007), for freeway systems with high occupancy vehicle (HOV) lanes, it is possible to identify when waste is and is not damaging.

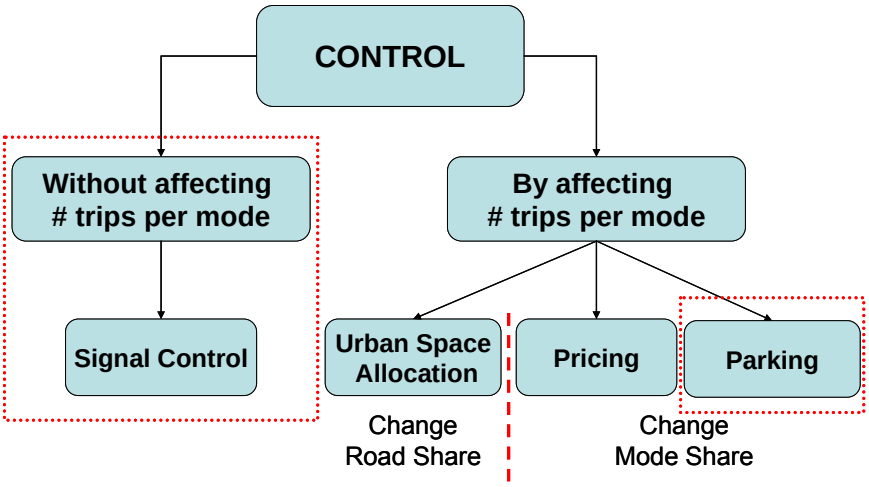


Figure 6.1: Different types of control strategies

(d) *Partitioning of a city to multiple reservoirs*: Prior to developing multi-reservoir, multimodal control strategies it is necessary to have a methodology on how to partition a city in multiple reservoirs, for which a reasonable static and dynamic system representation will still exist. It may be considered that each of the reservoir should be more or less homogeneous in traffic loads, especially when control is applied, e.g. during the peak hour. Also, it should be mentioned that a large number of reservoirs will contradict the main motivation of this dissertation research, as it may require knowledge of disaggregated demand; a value between 2 and 7 seems logical.

(e) *System monitoring*: Chapters 2, 3 and 4 described how we model a city in a macroscopic basis. Chapter 5 described how to apply system-wide control strategies to improve mobility. One important support tool both for modeling and control is a system for network monitoring. The success of the control methods heavily relies on

how accurately we can monitor the network and measure its state variables. Also, the partitioning of a city in multiple reservoirs and the estimation of an MFD requires knowledge, which can be obtained only from surveillance devices or other similar techniques. Section 3.3 provided some insight how to monitor the system using fixed and mobile sensors (detectors and taxis). Many crowded cities that could benefit from these controls do not have a supporting infrastructure to monitor their state. Nevertheless, these cities often have vehicles equipped with GPS that, like the Yokohama taxis, can serve as city-wide probes. Thus, more research is needed to better understand the quantity and character of the probes necessary for an accurate estimation of a neighborhood's traffic state. An effort is also needed to produce middleware that will support these types of control strategies on traffic signals of different types.

Bibliography

Anderson, S., De Palma, A. (2004) The economics of pricing parking, Journal of Urban Economics 55, 1–20.

Arnott, R., De Palma, A., Lindsey, R. (1992) A temporal and spatial equilibrium analysis of commuter parking, Journal of Public Economics 45, 301–335.

Arnott, R., Rowse, J. (1999) Modeling parking, Journal of Urban Economics 45, 97–124.

Calthrop, E., Proost, S., Van Dender, K. (2000) Parking policies and road pricing, Urban Studies 37, 63–76.

Daganzo, C. F. (1998) Queue spillovers in transportation networks with a route choice, Transportation Science 32 (1), 3–11.

Daganzo, C.F. (2002) A Behavioral Theory of Multi-Lane Traffic Flow Part I: Long Homogeneous Freeway Sections, Transportation Research Part B, 36, 131-158.

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

Douglas, R. (1975) A parking model—the effect of supply on demand, American Economist 19, 85–86.

Edie, L.C. (1963) Discussion of traffic stream measurements and definitions, Proc. 2nd Int. Symp. On the Theory of Traffic Flow, (J. Almond, editor), pp. 139-154, OECD, Paris, France.

Federal Highway Administration (2003) CORSIM Version 5.1 Users Manual. US Department of Transportation, Washington DC.

Federal Highway Administration (2004) Traffic Congestion and Reliability: Linking Solutions to Problems, 2004 FWHA Report prepared by Cambridge Systematics, Inc. with Texas Transportation Institute (http://ops.fhwa.dot.gov/congestion_report/) Accessed July 17, 2005.

Glazer A., Niskanen, E. (1992) Parking fees and congestion, Regional Science and Urban Economics 22,123–132.

Hellenic Institute of Transportation Engineers (2006) (http://www.ses.gr) Accessed April 23, 2006.

Herman, R. and Ardekani S. A. (1984) Characterizing Traffic Conditions in Urban Areas. Transportation Science, Vol. 18, No. 2.

Herman, R., Malakhoff, L., Ardekani, S. (1988) Trip time-stop time studies of extreme driver behaviors. Transportation research. Part A, Vol. 22A, no. 6, 427-433.

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

Joos, Ernst. Deputy Director of Zurich Transport Authority (2000) Economy and ecology are no contradictions, EcoPlan International. Paris, FR. (http://www.ecoplan.org/politics/general/zurich.htm) Accessed July 14, 2005.

Kuwahara, M. (2007) Private communication.

Little, J. D. (1961) A Proof of the Queueing Formula L = λ W. Operations Research, 9, 382-387.

Mahmassani, H., Williams, J. C., Herman, R. (1987) Performance of urban traffic networks. In: Gartner, N. H., Wilson, N. H. M. (Eds.), 10th Int. Symp. on Transportation and Traffic Theory. Elsevier, Amsterdam, The Netherlands.

Ressler, N. (1999) MIT conference on "Traffic Congestion: A Global Perspective." , 8-9 June, MIT (http://web.mit.edu/newsoffice/1999/traffic-0714.html) Accessed July 20, 2006.

Sarvi, M., Horiguchi, R., Kuwahara, M., Shimizu, Y., Sato A., Sugisaki Y. (2003) A methodology to identify traffic condition using intelligent probe vehicles, Proceedings of 10th ITS World Congress, Madrid.

Shoup, D. (2005a) The high cost of free parking, American Planning Association

Shoup, D. (2005b) The high cost of free parking, San Francisco Chronicle Article, June 3 (http://shoup.bol.ucla.edu/San%20Francisco%20article.htm) Last accessed November 21, 2007.

Small, K. (2004) Urban Transportation. Concise Encyclopedia of Economics, 2nd edition. (Indianapolis: Liberty Fund)

Smeed, R. J. (1966) *Road Capacity of City Centers.* Traffic Engineering and Control, Vol. 8, No. 7

Smeed, R. J. (1968) *Traffic Studies and Urban Congestion.* Journal of Transport Economics and Policy, Vol. 2, No. 1

Texas Transportation Institute (2005) 2005 Urban Mobility Study. (http://mobility.tamu.edu/ums/) Accessed August 12, 2005

Thomson, J. M. (1967) Speeds and Flows of Traffic in Central London: 1. Sunday Traffic Survey. Traffic Engineering and Control, Vol. 8, No. 11

Transport for London (2004) Congestion Charging: Impacts Monitoring, Second Annual Report, London (www.tfl.gov.uk/tfl/cclondon/cc_monitoring-2nd-report.shtml) Accessed August 14, 2005

Treiterer, J. and Myers, J.A. (1974) The Hysteresis Phenomenon in Traffic Flow. Proceedings of 6th International Symposium on Transportation and Traffic theory, (D.J. Buckley ed.), 13-38

Wardrop, J. G. (1952) Some Theoretical Aspects of Road Traffic Research. Proceedings of the Institution of Civil Engineers, Vol. 1, Part 2

Wardrop, J. G. (1968) Journey Speed and Flow in Central Urban Areas. Traffic Engineering and Control, Vol. 9, No. 11.

Williams, J. C., Mahmassani, H. S. and Herman R. (1987) Urban Traffic Network Flow Models, Transportation Research Record 1112, Transportation Research Board

Zahavi, Y. (1972) Traffic Performance Evaluation of Road Networks by the α-Relationship. Parts I and II, Traffic Engineering and Control, Vol. 14, No. 5 and 6, 228-231 and 292-293.

Appendix A

Perimeter control strategy for two-neighborhood cities: A proof

We now prove that given the state of a two-reservoir system at time t, output is maximized when the control strategy presented in section 5.1 is applied.

Proposition A: *Assume that a function* $f(x)$ *is concave and* $f(0)=0$. *Then* $x_1 f(x_2) \geq x_2 f(x_1) \quad \forall x_1, x_2$ *with* $x_1 > x_2 \geq 0$.

Proof: Using Jensen's inequality (Jensen, 1906) we get

$$f\left(\frac{x_2}{x_1}x_1 + \left(1-\frac{x_2}{x_1}\right)\cdot 0\right) \geq \frac{x_2}{x_1} f(x_1) + \left(1-\frac{x_2}{x_1}\right) f(0) \Leftrightarrow x_1 f(x_2) \geq x_2 f(x_1). \square$$

The next lemma states that it is not possible to increase the number of trips ending internally in *i*, by restricting movements in *i* with destinations outside *i*.

Lemma B: *Assume that a two-reservoir system is governed by equations* (5.1) *and* (5.2) *and* O_i *is concave for* $n_i^t \in \left[0, n_i^+\right]$ *with* $O_i(0) = 0$. *Then, it is impossible to increase* $\phi_{ii}^{t+1} \cdot O_i^{t+1}\left(n_i^{t+1}\right)$ *by decreasing* $x_{i'}^t$ *if* $n_i^t < n_i^+$.

Proof: Consider the case $x_{i'}^t \le \phi_{ii'}^t \cdot O_i\left(n_i^t\right)$, otherwise restriction in control has no effect for the system. By decreasing $x_{i'}^t$ by $\Delta n_i^t > 0$ and denoting by 0 and 1 states without and with restriction, accumulation in reservoir *i* is:

$$\left(n_i^{t+1}\right)_1 = \left(n_i^{t+1}\right)_0 + \Delta n_i^t, \tag{A.1}$$

This restriction does not affect the number of trips willing to finish in reservoir *i*:

$$\left(\phi_{ii}^{t+1} n_i^{t+1}\right)_1 = \left(\phi_{ii}^{t+1} n_i^{t+1}\right)_0. \tag{A.2}$$

After some manipulations in (A.3) and (A.4), we get:

$$\left(\phi_{ii}^{t+1}\right)_1 = \frac{\left(n_i^{t+1}\right)_0}{\left(n_i^{t+1}\right)_0 + \Delta n_i^t}\left(\phi_{ii}^{t+1}\right)_0. \tag{A.3}$$

By comparing $\phi_{ii}^{t+1} \cdot O_i\left(n_i^{t+1}\right)$ before and after the restriction we have:

$$\left(\left(n_i^{t+1}\right)_0 + \Delta n_i^t\right) \cdot O_i\left(\left(n_i^{t+1}\right)_0\right) \ge \left(n_i^{t+1}\right)_0 \cdot O_i\left(\left(n_i^{t+1}\right)_0 + \Delta n_i^t\right), \tag{A.4}$$

which is true, according to Proposition A, for $\Delta n_i^t > 0$. □[19]

[19] Note that if O_i is proportional to n_i, i.e. $O_i(n_i)$=a×n_i, control has no effect in the internal output of *i*.

Corollary: *If* $n_i^t < n_i^+$*, then* $\phi_{ii}^{t+1} \cdot O_i^{t+1}\left(n_i^{t+1}\right)$ *cannot decrease when* $x_{i'}^t$ *increases.*

By increasing $x_{i'}^t$, Δn_i^t in equations (A.2) and (A.4) is negative so according to Proposition A, the number of trips ending inside the reservoir does not decrease. Especially, for O_i strictly concave, $\phi_{ii}^{t+1} \cdot O_i^{t+1}\left(n_i^{t+1}\right)$ increases. □

We now analyze the 6 different cases and we show for what values of x_i^t the total output of the system o_{t+1} is maximized.

Case 1: $n_i^t \le n_i^- \ \forall i = 1,2$

When x_i^t decreases, both ϕ_{ii}^{t+1} and n_i^{t+1} decrease; $\phi_{ii}^{t+1} \cdot O_i\left(n_i^{t+1}\right)$ also decreases (O_i is increasing for $n_i^t < n_i^-$) . According to Lemma B, $\phi_{i'i'}^{t+1} \cdot O_{i'}\left(n_{i'}^{t+1}\right)$ decreases, and consequently o_{t+1} decreases. By increasing x_i^t, number of trips finishing in i and i' increases. So, all the times x_i^t should satisfy inequality (A.5) and intertransfers between the reservoirs should be unrestricted.

$$x_i^t \ge \phi_{i'i}^t \cdot O_{i'}\left(n_{i'}^t\right) \tag{A.5}$$

Case 2: $n_i^t \le n_i^-$ and $n_{i'}^- \le n_{i'}^t \le n_{i'}^+$

By decreasing $x_{i'}^t$, $\phi_{ii}^{t+1} \cdot O_i\left(n_i^{t+1}\right)$ decreases (Lemma B). Also, $\phi_{i'i'}^{t+1}$ decreases and $O_{i'}\left(n_{i'}^{t+1}\right)$ remains constant, so o_{t+1} decreases. By decreasing x_i^t, o_{t+1} decreases as ϕ_{ii}^{t+1}, n_i^{t+1} and $\phi_{i'i'}^{t+1}$ decrease. It is intuitive that the results are the opposite if any of x_i^t, $x_{i'}^t$

increase and equation (A.5) does not hold. Thus, similar to case 1, x_1^t, x_2^t should satisfy constraint (A.5) and intertransfers between the reservoirs should be unrestricted.

Case 3: $n_i^- < n_i^t < n_i^+ \quad \forall i = 1,2$

By decreasing any of x_i^t, $x_{i'}^t$ both ϕ_{ii}^{t+1}, $\phi_{i'i'}^{t+1}$ decrease and as O_i is constant, o_{t+1} decreases. By increasing any of x_i^t, $x_{i'}^t$, when inequality (A.5) is not satisfied, o_{t+1} increases. Thus, the result is the same as in case 2.

Summarizing cases 1 to 3, if $n_i^t < n_i^+ \ \forall i = 1,2$, o_{t+1} is maximized when inequality (A.5) is satisfied. This strategy is Pareto efficient as internal output increases for both reservoirs. Let us look at the most interesting cases, where at least one of the reservoirs is in the declining branch of $O_i(n_i)$, i.e. in regime III.

Case 4: $n_{i'}^t \geq n_{i'}^+$ and $n_i^- < n_i^t < n_i^+$

By decreasing x_i^t, the number of trips inside both reservoirs decreases because (i) for reservoir i': $\phi_{i'i'}^{t+1}$ decreases and $n_{i'}^{t+1}$ increases, ($O_{i'}$ is decreasing for $n_{i'} \geq n_{i'}^+$) and (ii) for reservoir i: ϕ_{ii}^{t+1} decreases (O_i is constant for $n_i^- < n_i < n_i^+$).By increasing x_i^t when equation (A.5) does not hold, this has the opposite effect for both reservoirs; and o_{t+1} increases. This means that we should not restrict any movements from i' to i. But, restrictions in the other direction (from i to i') can have positive effect.

By decreasing $x_{i'}^t$ by $\Delta x_{i'} > 0$ and denoting 0 and 1 states without and with restriction, equations (A.1) and (A.3) hold for i. For i', we have:

$$\left(n_{i'}^{t+1}\right)_1 = \left(n_{i'}^{t+1}\right)_0 - \Delta n_i^t \tag{A.6}$$

$$\left(\phi_{i'i'}^{t+1} n_{i'}^{t+1}\right)_1 = \left(\phi_{i'i'}^{t+1} n_{i'}^{t+1}\right)_0 - \Delta n_i^t \tag{A.7}$$

$$\left(\phi_{i'i'}^{t+1}\right)_1 = \frac{\left(\phi_{i'i'}^{t+1}\right)_0 \left(n_{i'}^{t+1}\right)_0 - \Delta n_i^t}{\left(n_{i'}^{t+1}\right)_0 - \Delta n_i^t} < \left(\phi_{i'i'}^{t+1}\right)_0 \tag{A.8}$$

Let us now look at the rate trips end inside each reservoir. For reservoir i, internal output decreases from $\left(\phi_{ii}^{t+1}\right)_0 \cdot \gamma_i$ to $\frac{\left(n_i^{t+1}\right)_0}{\left(n_i^{t+1}\right)_0 + \Delta n_i^t} \cdot \left(\phi_{ii}^{t+1}\right)_0 \cdot \gamma_i$. For reservoir i' the difference in the internal output $\Delta o_{i'}^{t+1}$ is (indexes of time and control (state 0 or 1) are omitted for simplicity):

$$\Delta o_{i'} = \Delta n_i \left(\phi_{i'i'} \cdot \frac{O_{i'}\left(n_{i'} - \Delta n_i\right) - O_{i'}\left(n_{i'}\right)}{\Delta n_i} - \left(1 - \phi_{i'i'}\right) \frac{O_{i'}\left(n_{i'} - \Delta n_i\right)}{n_{i'} - \Delta n_i} \right) \tag{A.9}$$

If (A.9) is smaller than zero then internal output decreases for both reservoirs and no restrictions should be applied. Also, note that $\Delta o_{i'}$ is monotonically increases with accumulation $n_{i'}$ for $O_{i'}$ decreasing and concave. Qualitatively, $\Delta o_{i'}$ is positive for high values of $\phi_{i'i'}$ and when reservoir i' is very congested. Change in the total system output Δo_{t+1} is (indexes are omitted):

$$\Delta o = \Delta o_{i'} - \frac{\Delta n_i}{n_i + \Delta n_i} \cdot \phi_{ii} \cdot \gamma_i \tag{A.10}$$

Thus, if Δo>0 perimeter control increases the output of the system. Let us give an arithmetic example: For a two-reservoir system with (i) the area of the uncongested reservoir i, 5 times bigger than the congested reservoir i'; (ii) $O_{i'}(n_{i'})$ function, the one

estimated for San Francisco[20] (figure 2.10) and (iii) $O_i(n_i)=5\ O_{i'}(5n_{i'})$, we have that, when the congested reservoir is in state C of figure2.7a (average speed 7km/hr), Δo is positive (perimeter control is effective) for $\phi_{i'i'} > 0.4$.

Case 5: $n_{i'}^t \geq n_{i'}^+$ and $n_i^t \leq n_i^-$

By increasing x_i^t when equation (A.5) does not hold, internal output for both reservoirs increases. As also stated in case 4, we should not restrict any movements from i' to i.

By decreasing $x_{i'}^t$ by $\Delta n_i^t > 0$ and denoting 0 and 1 states without and with restriction, equations (A.1) and (A.3) hold for reservoir i and (A.6) and (A.8) for reservoir i'. According to Lemma A, internal output decreases in reservoir i. Change in internal output for reservoir i' is determined by the sign of equation (A.10). Change in the system output, Δo_{t+1}, is (indexes of time and control (state 0 or 1) are omitted for simplicity):

$$\begin{aligned}\Delta o = \Delta n_i \Bigg(& \phi_{i'i'} \frac{O_{i'}(n_{i'} - \Delta n_i) - O_{i'}(n_{i'})}{\Delta n_i} - (1-\phi_{i'i'}) \frac{O_{i'}(n_{i'} - \Delta n_i)}{n_{i'} - \Delta n_i} + \\ & + \phi_{ii} \frac{O_i(n_i + \Delta n_i) - O_i(n_i)}{\Delta n_i} - \phi_{ii} \frac{O_i(n_i + \Delta n_i)}{n_i + \Delta n_i} \Bigg) \Rightarrow \quad . \qquad \text{(A.11)} \\ \Delta o = \Delta n_i & \left(\phi_{i'i'} s_{OA} - (1-\phi_{i'i'}) s_{AB} + \phi_{ii} s_{CD} - \phi_{ii} s_{OD} \right)\end{aligned}$$

where $s_{OA}, s_{AB}, s_{CD}, s_{OD}$ are the absolute values of slopes for lines shown in figure A.1 and perimeter control improves accessibility when Δo>0.

[20] The MFD for reservoir i, is assumed to be a scaled up version of the MFD for i'.

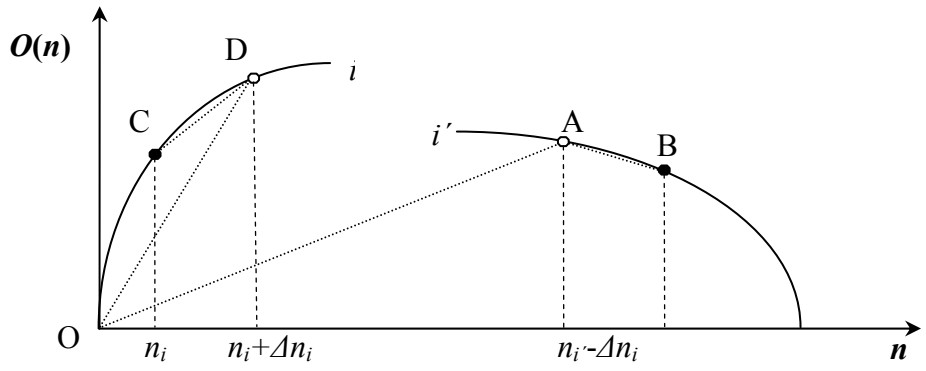


Figure A.1: MFDs: (a) regime I for reservoir *i*, (b) regime III for *i'*

Analysis in chapters 2 and 3 show that the Macroscopic Fundamental Diagram in regime I is approximately linear, for a range of accumulation, with $O_i(0)=0$. In that case, Δo_{t+1} is given by equation (A.10) and the arithmetic example described in case 4 is valid here, too. Note also that for linear $O_i(n_i)$ in regime I, the effect of perimeter control in the uncongested part of the city is almost negligible.

Case 6: $n_i^t > n_i^+ \quad \forall i = 1,2$

By decreasing $x_{i'}^t$ by $\Delta n_i^t > 0$ and denoting 0 and 1 states without and with restriction, equations (A.2) and (A.4) hold for reservoir *i* and (A.6) and (A.8) for reservoir *i´*. Change in system output Δo_{t+1} is given by equation (A.11), but now the third term inside the parenthesis is negative. Δo is positive when $\phi_{i'i'} > \phi_{ii}$ and reservoir *i´* is more congested than reservoir *i*. In this case, perimeter control cannot have a significant effect, as the whole city experiences heavy congestion. More drastic strategies are required (e.g. a combination of pricing and efficient transit), which will decrease the demand for cars and provide users alternative modes.

Appendix B

Glossary of symbols

Chapter 1

Two-fluid model

f_r	fraction of moving vehicles
k	parameter for the quality of traffic service in the network
T_r	inverse of v_r
T_m	inverse of v_m
T	average total trip time (including stopped time)
v_r	average speed of the moving vehicles
v_m	average maximum running speed

Gridlock model (Daganzo, 2007)

$f(t)$	input flow to the system at time t

l	average trip length
n_i	accumulation for individual link i
n	network accumulation
O	network output (trip completion rate)
P_i	travel production for individual link i
P	network travel production

Chapter 2

$\overline{q}$	average flow
$\overline{v}$	average speed
$\overline{k}$	average density
n_t	accumulation at time t
f_t	input flow at time t

Chapter 3

A	the set of lane segments in the study area
A'	subset of A with detectors
i	a road lane segment between intersections
l_i	length of i
q_i	flow measured in i by the detector
o_i	occupancy measured in i by the detector
k_i	density in i

s	space-mean effective vehicle length
S	state of a taxi
q^w	weighted average flow for all lane segments with detectors
q^u	unweighted average flow for all lane segments with detectors
o^w	weighted average occupancy for all lane segments with detectors
o^u	unweighted average occupancy for all lane segments with detectors
k^w	weighted average density for all lane segments with detectors
k^u	unweighted average density for all lane segments with detectors
A''	subset of A covered by detectors
O'	output for A'
O	output for A
D'	outbound perimeter flow
δ	total distance traveled in A by full taxis
τ	total time spent in A by full taxis
v_T	full taxis space-mean speed in A
n_T	number of full taxis in A
N_T	number of full taxis exited A along its perimeter and
M_T	number of full taxis finished a trip inside A
v	space-mean speed of all cars in A
n	accumulation of all cars in A
π	probability that a randomly selected car moving in A is a taxi
N'	numbers of vehicles exiting A' along streets with detectors
N'_T	numbers of full taxis exiting A' along streets with detectors

p	ratio of vehicles vs. full taxis exiting the zone from streets with detectors

Chapter 4

Multi-reservoir system

N	number of reservoirs
i	reservoir
l_i	average trip length for i
n_i	accumulation in i
P_i	production in i
V_i	average space-mean speed in i
O_i	output in i
O_{ij}	output from i for cars with final destination j
C_{ij}	inflow capacity to reservoir j from an adjacent reservoir i
a_{ij}	constant inversely proportional to l_i
n_{ij}	vehicles in i with final destination j
l_{ij}	average distance traveled in i, by all vehicles in i with destination j
$q_{ij}(t)$	exogenous flow generated in i with destination j at time t
$\delta_{i\to k}^{j}$	exogenous binary variable
$q_{i\to k}^{j}(t)$	transferring flow from reservoir i to k at time t, with final destination j
$x_{ij}^{*}(t)$	a control variable for the inflow capacity for movements from i to j

Cruising-for-parking

n_s	vehicles searching-for-parking
n_m	vehicles moving towards their destination internal to the reservoir, but not yet searching for parking
n_o	vehicles moving with external destinations
n_p	vehicles parked on street
N_p	total number of parking spots
p	percentage of available parking spots
d_1	average distance traveled between two adjacent spots
l_s	average distance traveled while searching for parking
O_m	output for n_m
O_o	output for n_o
O_s	output for n_s
l	average distance traveled per trip completion
$n^*(t)$	accumulation at time t when $N_p \to \infty$
d_p	total delay due to the cruising-for-parking
d_s	delay while cruising-for-parking

Chapter 5

n_i^t	accumulation at reservoir i, at time t
ϕ_{ij}^t	fraction of vehicles in i with destination in j, at time t
x_i^t	control variable for movements directing to i, at time t
f_{ij}^t	fraction of trips generated at i with destination j, at time t
q_i^t	generated flow at i, at time t.
o_t	total output of the system at time t